

Department of Justice
Office of Justice Programs
National Institute of Justice



Measurement and Analysis of Crime and Justice

**CRIMINAL
JUSTICE**

2000

VOLUME 4

182411

U.S. Department of Justice
Office of Justice Programs
810 Seventh Street N.W.
Washington, DC 20531

Janet Reno
Attorney General

Daniel Marcus
Acting Associate Attorney General

Mary Lou Leary
Acting Assistant Attorney General

Julie E. Samuels
Acting Director, National Institute of Justice

Office of Justice Programs
World Wide Web Site
<http://www.ojp.usdoj.gov>

National Institute of Justice
World Wide Web Site
<http://www.ojp.usdoj.gov/nij>

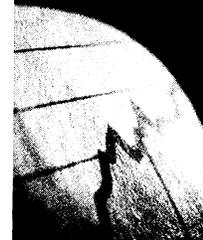
182411

Measurement and Analysis of Crime and Justice

PROPERTY OF
National Criminal Justice Reference Service (NCJRS)
Box 6000
Rockville, MD 20849-6000

July 2000

NCJ 182411





Julie E. Samuels
Acting Director

Eric Jefferis
Janice Munsterman
Criminal Justice 2000
Program Managers

Robert Kaminski
Nancy La Vigne
NIJ Coordinators
for Volume 4

Findings and conclusions of the research reported here are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The National Institute of Justice is a component of the Office of Justice Programs, which also includes the Bureau of Justice Assistance, the Bureau of Justice Statistics, the Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime.

From the Director

The celebration of the National Institute of Justice's 30th anniversary in the autumn of 1999 provided the Institute and the criminal justice community the perfect opportunity to reflect on three decades of criminal justice research accomplishments. A few months later, the dawn of the new millennium seemed the appropriate stage from which to look forward to what lies ahead for criminal justice in the United States.

As preparations were made to commemorate the Institute's anniversary, it became increasingly apparent to NIJ staff and the criminal justice research, policymaker, and practitioner communities that there needed to be one compilation comprising a comprehensive, scholarly examination and analysis of the current state of criminal justice in the United States. Consequently, NIJ conceived and launched a project to produce the four-volume *Criminal Justice 2000* series to examine how research has influenced current policy and practice and how future policies and practices can be built upon our current state of knowledge.

The themes developed for these volumes were purposefully broad in scope, to allow contributors the intellectual freedom to explore issues across criminal justice disciplines. In its competitive solicitation, NIJ asked the authors to explore and reflect on current and emerging trends in crime and criminal justice practice, based on scientific findings and analyses. An editorial board of eminent criminal justice researchers and practitioners then selected the proposals that displayed exceptional scholarly merit and contributed to the substantive themes of the volumes.

The result, the *Crime and Justice 2000* series, reflects the state of knowledge on a broad spectrum of crime and criminal justice issues. While the volumes do not comprehensively chronicle all topics vital to criminal justice in the United States at the year 2000, we hope the essays contained in these four volumes will stimulate thought and discussion among policymakers, practitioners, and scientists in the coming years and shape future research endeavors.

Julie E. Samuels

Acting Director

National Institute of Justice

Criminal Justice 2000

Editorial Board Members

Volume 1. The Nature of Crime: Continuity and Change

Editor: Gary LaFree

Editorial Board: James F. Short
Robert J. Bursik, Sr.
Ralph B. Taylor

Volume 2. Boundary Changes in Criminal Justice Organizations

Editor: Charles M. Friel

Editorial Board: Susan Keilitz
Charles Wellford
Chase Riveland
James Jacobs

Volume 3. Policies, Processes, and Decisions of the Criminal Justice System

Editor: Julie Horney

Editorial Board: John Martin
Doris L. MacKenzie
Ruth Peterson
Dennis Rosenbaum

Volume 4. Measurement and Analysis of Crime and Justice

Editor: David Duffee

Editorial Board: David McDowall
Brian Ostrom
Robert D. Crutchfield
Stephen D. Mastrofski
Lorraine Green Mazerolle

Table of Contents

From the Director	iii
<i>Criminal Justice 2000</i> Editorial Board Members	v
Introduction to Volume 4 Measurement and Analysis of Crime and Justice: An Introductory Essay	1
<i>by David Duffee, David McDowall, Lorraine Green Mazerolle, and Stephen D. Mastrofski</i>	
The Self-Report Method for Measuring Delinquency and Crime.	33
<i>by Terence P. Thornberry and Marvin D. Krohn</i>	
Self-Report Surveys as Measures of Crime and Criminal Victimization.	85
<i>by David Cantor and James P. Lynch</i>	
Theory, Method, and Data in Comparative Criminology.	139
<i>by Gregory J. Howard, Graeme Newman, and William Alex Pridemore</i>	
Spatial Analyses of Crime	213
<i>by Luc Anselin, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita</i>	
Measuring the Costs and Benefits of Crime and Justice	263
<i>by Mark A. Cohen</i>	
Measuring the Sexual Victimization of Women: Evolution, Current Controversies, and Future Research.	317
<i>by Bonnie S. Fisher and Francis T. Cullen</i>	
Measurement and Analysis of Drug Problems and Drug Control Efforts	391
<i>by Jonathan P. Caulkins</i>	
Fear of Crime in the United States: Avenues for Research and Policy.	451
<i>by Mark Warr</i>	
Measurement and Explanation in the Comparative Study of American Police Organizations	491
<i>by Edward R. Maguire and Craig D. Uchida</i>	
Standards and Measures of Court Performance	559
<i>by Ingo Keilitz</i>	
Appendix	595

Measurement and Analysis of Crime and Justice: An Introductory Essay

by David Duffee, David McDowall, Lorraine Green Mazerolle,
and Stephen D. Mastrofski

The turn of any century is traditionally seen as a milestone in human development. With the coming of the 21st century, the national media have reflected on the century past with lists of the most significant events and persons of the time. Political, social, and cultural events have welcomed the arrival of the new century—and the new millennium. The birth of a new century brings together celebrations of what the old century has accomplished with what we anxiously await in the new century.

The National Institute of Justice (NIJ) seized on the turn of the century as an opportunity to document 20th-century milestones in criminology and criminal justice and to highlight developments that will shape our justice system in the 21st century. NIJ commissioned a series of volumes that would span the breadth and depth of criminological and criminal justice thinking that has both created our current knowledge base and formed the foundation for our thinking in the 21st century. The fourth of these volumes is *Measurement and Analysis of Crime and Justice*, introduced in the following pages.

The original intent of volume 4 was to identify methodological debates that have shaped the evolution of crime and justice research. Under this broad umbrella, we sought to include papers in this volume that highlighted measurement dilemmas and solutions as well as analytic difficulties and applications that have contributed to what we now know and what we still do not know about crime and justice. In particular, the volume sought to explore the current knowledge, trends, and future directions in the measurement and analysis of crime and the criminal justice system,

David Duffee is Professor and David McDowall is Professor in the School of Criminal Justice, University at Albany, State University of New York. Lorraine Green Mazerolle is Associate Professor in the Division of Criminal Justice, University of Cincinnati. Stephen D. Mastrofski is Professor of Public and International Affairs and Director of the Administration of Justice Program at George Mason University.

the consequences of such measurement and analyses for justice processes and the research enterprise, and the context in which both crime and justice operate. When we developed the original solicitation for papers, we understood that the set of topics within this domain would be potentially vast: sampling, mathematical modeling, statistical analysis, data visualization, and research design, to list only a few possibilities.

This introduction identifies and interprets the common themes running through the 10 papers included in this volume and indicates other themes not included. We begin our introduction with a review of the substantive topic areas that make up this volume. We then examine four main themes that cut across the papers: theoretical framework, data and measurement, analytic problems, and use of research in decisionmaking. We conclude with a discussion of what these papers highlight as lessons for the future.

Topic Areas

When we began our search for papers to include in volume 4, we hoped to include papers that covered substantive criminological topics (such as sexual assault, drugs, homicide, and disorder) and substantive criminal justice topics (such as agency goal and policy setting and officials' decisionmaking). We also sought papers spanning cutting-edge analytic topics (such as spatial analysis and cost-benefit analysis), dominant methodologies (such as self-report surveys of victimization and offending), and themes that have dominated policy debates in local, State, and Federal arenas (such as fear of crime and victimization). With the limited number of papers that we could commission, we sought to include papers that covered at least one substantive area while also elaborating on at least one measurement or analytic development.

The volume begins with five chapters about general issues in the measurement and analysis of crime and crime control. The next three chapters examine similar issues, but they provide more detailed discussions of measurement and analysis about specific crimes or policy areas. Specifically, these three chapters cover the measurement of sexual victimization, the measurement and control of fear, and the measurement and control of drug abuse. The final two chapters examine measurement of the characteristics of agencies in the criminal justice system.

Chapters 1 and 2 highlight dilemmas and difficulties with self-report surveys: one covering the measurement of delinquency and crime (Terence P. Thornberry and Marvin D. Krohn) and the other covering the measurement of crime and victimization (David Cantor and James P. Lynch). Chapter 3 (Gregory J. Howard, Graeme Newman, and William Alex Pridemore) continues the discussion of the

measurement of crime on a larger scale by examining the problems and progress in cross-national comparisons. Chapter 4 (Luc Anselin, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita) maintains the focus on crime but shifts attention from measurement to the problems in analyzing crime data that are spatially and temporally clustered. Cohen's chapter on cost-benefit analysis applied to criminal justice rounds out this set of chapters by asking more generally how to assess policy impacts against the impacts of the social problem.

Chapters 6, 7, and 8 revisit many of the issues covered in earlier chapters but elaborate on measurement and analytical problems and solutions in specific areas. Chapter 6, by Bonnie S. Fisher and Francis T. Cullen, extends the discussion of self-report survey methods of crime and victimization by exploring the difficulties in measuring sexual victimization, one area in which the personal and political sensitivities of the problem complicates measurement. Chapter 7, by Jonathan P. Caulkins, is concerned both with the measurement of the social problem of drug abuse and with the measurement of the effects of interventions.

Like chapter 7, chapter 8, by Mark Warr, is also concerned both with gauging accurately the nature and scope of a problem (fear of crime) and with the dilemmas in policy attempts to control fear, in relation to controlling crime.

The final two chapters cover measurement and analytic issues that dominate key areas in the criminal justice system. Edward R. Maguire and Craig D. Uchida focus on police organizations, and Ingo Keilitz examines standards and measures of court performance. These two chapters, while focusing on measurement within specific domains of the criminal justice system, transcend a potentially narrow orientation and raise issues that are relevant to other aspects of the justice system.

Since all 10 chapters in this volume address substantive issues and methodological concerns, ordering the papers was problematic. Each possible order was bound to highlight some links (such as the survey methods used for measuring crime in Cantor and Lynch, Thornberry and Krohn, and Fisher and Cullen), while separating and perhaps downplaying other links (such as the discussion occurring much later of the police measurement of crime in Maguire and Uchida). In order to indicate the common themes that run across several chapters in this volume, exhibit 1 may be helpful to locate themes across the papers.

Much of the research reviewed throughout this collection of papers is concerned with the dependability or reliability of measures and their accuracy in representing the concepts used in theory.

Theoretical Framework

The inclusion of a “methods and analysis” volume in the National Institute of Justice Criminal Justice 2000 series might lead a reader to think that all discussions of theory and social context would be reserved for the other three volumes. However, volume 4 authors repeatedly emphasize the importance of theory to selecting and evaluating methods.

The editors and authors of volume 4 would appear to agree with Bernard and Ritti (1990, 1): “[T]heory alone is the distinguishing feature of the scientific

Exhibit 1. Topics by chapter in volume 4

Topic	Chapter number									
	1	2	3	4	5	6	7	8	9	10
Acceptance of measures										X
Community policing									X	
Cost of crime					X		X	X		
Cost of public response to crime					X		X	X		
Cost-benefit analysis					X		X			
Cost-effectiveness					X		X			
Crime hot spots				X						
Crime measurement	X	X	X	X	X	X	X		X	
Crime prevention				X	X		X		X	
Crime, violent	X	X	X			X				
Crime/delinquency, self-reports	X		X				X			
Cross-national data			X							
Data analysis				X						
Data collection, history	X	X				X			X	X
Data collection, police									X	
Data integration		X					X			
Data requirements/limits	X	X	X				X			X
Domestic violence		X	X			X				
Drug Use Forecasting/Arrestee Drug Abuse Monitoring	X						X			
Explanation of variation	X			X					X	X
Fear, behavioral indicators							X			
Fear, consequences							X			
Fear, measurement of							X			
Fear of crime							X			
Fear, promoting							X			

Exhibit 1 (continued)

Topic	Chapter number									
	1	2	3	4	5	6	7	8	9	10
Fear, reduction							X			
Fear, regulating							X			
Genocide			X							
Globalization			X							
Historical perspectives/context			X							
International data			X							
Interviewing	X	X				X				
Interviewing, computer-assisted	X	X				X				
Kernel estimation				X						
Law and economics					X					
Local indicators of spatial assoc.				X						
Longitudinal measures	X	X					X		X	
Measurement of police function									X	
Measurement of police structure									X	
Measurement procedures							X		X	X
Meta-theory			X							
Moran scatter plot				X						
National Crime Survey		X				X				
National Crime Victimization Survey		X				X				
National Violence Against Women Survey						X				
National College Women Sexual Victimization Study						X				
National Women's Study						X				
Offenses, seriousness	X						X			
Operational definitions	X	X	X			X				X
Organizational analysis									X	X
Panel effects	X	X								
Performance standards										X
Police discretion									X	
Police functions									X	
Police functions, explanations of									X	
Police organization									X	
Police org., explanations of									X	
Policy analysis			X		X		X	X		
Prediction	X			X						

continued

Exhibit 1 (continued)

Topic	Chapter number									
	1	2	3	4	5	6	7	8	9	10
Program accountability					X		X			
Program effectiveness					X		X			
Quadrant count method				X						
Rape, measurement of						X				
Reliability	X	X	X	X		X			X	
Reliability, internal consistency	X									
Reliability, test-retest	X									
Resistance to measurement										X
Response techniques	X	X				X				
Sample design		X								
Self-reports with adults	X	X								
Self-reports with children	X									
Sexual assault		X				X				
Sexual victimization						X				
Spatial autocorrelation				X						
Spatial data analysis				X						
Spatial modeling				X						
Stalking						X				
Survey methods	X	X				X	X		X	X
Survey of fear							X			
Survey of offending	X									
Survey of sexual victimization						X				
Survey, reference period	X	X				X				
Theory	X	X	X						X	
Theory and data	X	X	X	X	X	X	X	X	X	
Theory, contingency									X	
Theory, institutional									X	
Theory of crime	X	X		X						
Theory, routine activities				X						
Theory, social ecology				X						
Trial court outcomes										X
Use of research										X
Validity	X	X				X			X	X
Validity, construct	X									
Validity, construct continuity	X									
Validity, content	X									
Validity, criterion	X									

Exhibit 1 (continued)

Topic	Chapter number									
	1	2	3	4	5	6	7	8	9	10
Victimization, consequences of		X								
Victimization, measuring		X				X				
Victimization, responses to		X								
Victimization, screening question		X				X				
Victimization, self-reports		X	X			X				
Victimization of women						X				

enterprise. True, some of the [other features] are necessary and desirable, but regardless of measurement precision, quantification, or power of analytical tools, the activity is not science unless it involves an explicit theory.” Theory drives measurement and analysis because it identifies research questions, suggests acceptable methodological approaches, rationalizes the selection of variables and measures, and provides the framework for interpreting results (Bernard and Ritti 1990, 1–2).

This affirmation of the importance of theory in guiding and evaluating measurement and analysis does not contradict the importance of the interplay between methodological developments and the development and revision of theory. As Howard, Newman, and Pridemore point out in chapter 3, a number of cross-national theories are presently untestable because of the absence of relevant data. Anselin and colleagues review some of the problems in analyzing spatially and temporally clustered data. Assessing the impact of such problems on conclusions drawn from prior research is important in evaluating the soundness of our theoretical knowledge and in influencing the future direction of theoretical inquiry.

Much of the research reviewed throughout this collection of papers is concerned with the dependability or reliability of measures and their accuracy in representing the concepts used in theory. Howard, Newman, and Pridemore warn of the haphazard adoption of variables in cross-national research, based in part on the difficulty in obtaining data that have been collected for administrative and political, rather than theoretical, purposes. They also point out that the meaning of a datum or a trend is not self-evident but is connected to the relationship between an observable and its deployment as an indicator of a concept. Different cross-national investigations may use the same data in measuring different variables. This is not necessarily wrong. The same fact may have different meanings in different theories and, indeed, has scientific meaning only in such context.

Similarly, Maguire and Uchida (chapter 9) distinguish between the more common data collection activities that are assumed to “describe” police departments and the less common activities that measure theoretically meaningful aspects of police department function and structure. They indicate that the mountain of data that is now produced by and about many criminal justice agencies contains only some ore. Greater care conceptually might reduce the size of the data mountain, and it would also guide us to additional, sorely needed, data that are not routinely collected.

The importance of theory in guiding measurement and analysis is highlighted in the discussions of many of the important advances in measurement made in the 20th century. Thornberry and Krohn (chapter 1) and Cantor and Lynch (chapter 2) review two of the undeniable measurement advances that have shaped criminological research: survey methods for measuring crime from offender and victim sources. Both chapters stress the critical connection between theory and measurement. We do not have standards for assessing measures without reference to the concepts and connections among concepts for which we employ measures. Fisher and Cullen (chapter 6) then provide more detail on the process of developing measures for one type of crime/victimization. They demonstrate the interplay between anomalies in research findings and the reconceptualization of the meanings of victimization in the development of their national sexual victimization survey. In these discussions of measurement, we witness the process of theory guiding measurement and research revising theory. It is only through this process that researchers have avoided reinventing the wheel.

Many of the challenges for research in the 21st century identified by our authors are, in contrast, connected to the failures of research in the 20th century to take theory seriously enough. Authors have pointed out missed opportunities and poor measurement decisions that have hampered our research abilities. For example, Keilitz (chapter 10) reports that trial court outcome measures have been developed but that no attempt has been made to specify the court processes that might influence these outcomes. If future work begins to propose theoretical linkages between how these courts operate and what they achieve, researchers may determine that the most plausible connections require outcomes to be measured differently or that other likely outcomes have been overlooked in the original development of outcome variables.

One of the broad strategies that might be employed to identify opportunities more quickly and efficiently is greater attention to theory development as a criterion in justifying expenditures on measurement and analytic projects. Part of that strategy would include correcting an imbalance in current research that is more implied than explicit, but visible nonetheless, in the chapters contained in this volume.

There would appear to be greater collective investment in, and therefore more systematic accumulation of knowledge in, criminology than in criminal justice. The terms *criminology* and *criminal justice* are often both vague and synonymous. Snipes and Maguire (1995) suggested distinguishing them based on the dependent variables of interest. They use *criminological theory* to refer to theories that seek to explain variation in crime and *criminal justice theory* to identify theories that seek to explain variations of or in criminal justice systems. Distinguishing theoretical domains in this way, criminological theories are relatively more explicitly stated and pursued than criminal justice theories.

This disparity may be an artifact of different patterns of growth in the various disciplines that concern themselves with these issues. The study of crime seems to be increasingly concentrated among disciplinary specialists who identify themselves as criminologists, while the study of criminal justice remains more widely distributed across disciplines other than criminal justice, such as political science and sociology, and to a lesser extent, psychology and economics. Although criminology can scarcely be characterized at the turn of the century as showing consensus about the nature and causes of crime, it has achieved a lingua franca that clarifies theoretical debates and facilitates measurement and analysis. This is largely due, we speculate, to the tighter patterns of social interaction among scholars and researchers who study crime and its causes. Scholars and researchers who identify with criminal justice as a discipline would seem to place greater emphasis on applied matters, but—as some of our authors demonstrate—theory relevant to the explanation of criminal justice phenomena is available across a wide range of disciplines.

One indicator of this imbalance is the frequent mention in this volume of specific criminological theories and the connections of each to particular measurement and analytic problems. For example, similar theories of delinquency have been used to guide the improvement of reliability and validity both in the cross-national research discussed by Howard, Newman, and Pridemore and in the self-reported victimization and offending discussed by Fisher and Cullen, Cantor and Lynch, and Thornberry and Krohn. In contrast, Maguire and Uchida indicate that almost all the research on police organizations in this century has been guided by a single strain of organizational theory (contingency theory) and that even the deployment of this theory has been more implicit than explicit. Readily identifiable schools of thought, indicating specific theoretical traditions, are rare in criminal justice research. As a result, data gathering and analysis in criminal justice are often guided by implicit program assumptions or current policy concerns. (We will return to the complex relationship between scientific research and policy use of research later.) We will need to rectify this imbalance if knowledge is to be built systematically.

Criminal justice theory building is important in its own right, but it becomes more attractive when we consider the possible connections between criminal justice and criminological theory. These theoretical domains are connected in two important ways: in the potential impact of criminal justice behavior on crime and in the potential impact of crime on criminal justice. The first of these connections, in which criminal justice behavior is the independent variable, includes two quite different approaches. Only the first of these has received much attention.

This first approach is visible in the frequent concern for criminal justice program evaluation. There have been many studies of the effectiveness of criminal justice practices and programs in influencing crime. Most of these are some form of treatment or specific deterrence theory, in which we are interested in the impact of an intervention on repeated crime by someone who has offended at least once. These are most often correctional or treatment evaluations. However, this same approach to the impact of criminal justice programs on crime is also visible in research examining the role of criminal justice in prevention. This research tends to focus on criminal justice impacts on crime rates rather than criminal justice impacts on the future behavior of individuals. Both types of research have consumed the lion's share of criminal justice research dollars, probably because these studies address the primary question of policymakers: Do our criminal justice programs achieve the goals we intend for them?

The attention devoted to this form of criminal justice influence on crime has generally found, with some notable and important exceptions, that criminal justice actions are rarely the most powerful explanatory variables in the fluctuation of crime. As the century turns, one of the more significant data trends is the substantial drop in crime, measured through either official or victim sources. As this decrease has continued, a large number of claims have identified the positive effects of crime control policy as the independent variable. If such claims can be substantiated by more exacting causal analysis than is now available, this discovery of criminal justice impacts on crime would stand in stark contrast to the "nothing works" cynicism of a decade ago. Certainly this is one more example of the need for greater theoretical preparation in research, so that the potential causal influences of policing strategy, drug treatment, community crime prevention, and so on could be compared with the influences of the growing economy and demographic shifts in the population.

The second approach to criminal justice as an independent variable in criminological theory has received less attention than has the program evaluation approach. The relative lack of attention given this approach is ironic, since in this research tradition, criminal justice appears to be the predominant variable in explaining fluctuations in crime. This approach is visible in theories that seek to determine what is or is not to be treated as a crime and in theories that

seek to explain fluctuations in decisions about how much attention different types of crimes might get. Possibly, the potential payoffs in these theories have been ignored in the past because they have often been advanced only in critical or radical theories. These theories generally argue that the criminal justice system operates to deflect attention away from structural inequalities and to concentrate attention on individual-level explanations for street crime.

More recent developments using this approach are more politically neutral. The political science of crime policy and the emerging field of cultural criminology serve as examples. These approaches to criminal justice examine the complex and poorly understood connections among popular culture, media portrayals of crime and justice, the structure and dynamics of crime in political campaigns, and the effect of these on which social harms will receive attention as crimes. Determining what crime is, is perhaps the most fundamental criminal justice decision. But our study of this impact of criminal justice on crime is poorly understood, in comparison to our understanding of specific criminal justice actions on targeted crimes or targeted criminals.

The second connection between criminological and criminal justice theories would be in the examination of the impact of crime on criminal justice behavior. This impact appears to be strong but indirect, as it is interpreted and moderated by fear (see, for example, Warr, chapter 8) and by media and political responses (Scheingold 1984, 1991; Wilkins 1991). We need greater attention to how criminal justice adjusts to crime and to whether the adjustments it makes are to crime or to other forces (such as extremism in political campaigns, as Scheingold and Wilkins suggest, or “moral panics” as Chiricos [1998] describes).

It will be particularly valuable if we begin to think more broadly about the range of criminal justice concepts that are worthy of explanation. For example, one might attempt to test whether the drug wars of the 1980s and 1990s were a response to actual increases in illicit drug use of a particular sort or whether other explanations are more powerful, such as the moral panic perspective. But one might also conceptualize justice in terms of the distribution of punishment, protection, and other benefits, exploring who benefited and who did not by the response of the justice system to the drug problem. One might also measure the practices of the justice system according to a variety of normative standards. Doing this first

In order to pay greater attention to measurement and analysis in the criminal justice domain, we need a research policy that recognizes that fluctuations in criminal justice variables are not trivial, even if they do not have visible connections to crime.

requires an explication of normative theories of justice, which must then be applied to the issue at hand so that proper measurement and analysis can follow.

These issues cannot be addressed systematically until the same care is given to conceptualization and relationships among concepts in criminal justice as is given in criminology. Only with that attention to theory will the development of measurement and analysis in criminal justice advance systematically. In order to pay greater attention to measurement and analysis in the criminal justice domain, we need a research policy that recognizes that fluctuations in criminal justice variables are not trivial, even if they do not have visible connections to crime. Accessibility to the courts, proper treatment of citizens, job satisfaction and turnover of employees, and punishment rates—to name a few criminal justice variables—all merit explanation, whether or not their fluctuation has, in turn, some effect on crime. They are key elements of normative theories of justice that guide, might guide, or should guide the actions of those involved in the administration of justice. But, as Maguire and Uchida point out in chapter 9, the theories that would lead to the measurement and analysis of such fluctuations lag behind the theoretical development, and therefore the measurement and analysis standards, in criminology.

Data and Measurement

The papers in this volume are united by their shared emphasis on data. The authors stress issues in finding, interpreting, and understanding data on crime and criminal justice. They discuss the strengths and limitations of existing data sources and describe how researchers might improve ongoing data collection efforts. They consider basic questions about the types of data useful for criminal justice, and they suggest how researchers might more creatively exploit the data that they gather.

We believe that the authors' emphasis on data and measurement, as opposed to analytic techniques, did not occur by chance and that the focus on data and measurement reflects the general priorities of criminal justice as a field. This is so for two reasons.

First, criminologists and criminal justice researchers depend on analytic methods that they import from other disciplines. As in several other fields of social research (for example, political science and sociology), the analytic methods of criminology and criminal justice originated in statistics, econometrics, epidemiology, and psychometrics. Developments in these areas are occurring at a rapid pace, and a set of papers that concentrated on promising analytical techniques would soon be amusingly obsolete.

We can reasonably predict that the range of analytic methods in criminology and criminal justice will continue to expand during the 21st century. We also can reasonably predict that most of these methods will continue to come from other areas. Because methodological development will respond to issues outside criminal justice, additional predictions would be hazardous.

We believe that the second, and more important, reason for the authors' focus on data and measurement issues is the central role of these topics in understanding crime and justice. More than two decades ago, Hubert Blalock (1979) suggested that the most important challenge to empirical research in sociology was not to develop more sophisticated analytical methods. Instead, according to Blalock, the key element in advancing knowledge about society was a better understanding of data and measurement.

This observation is equally true for inquiry in crime and justice today (e.g., Maltz 1992). Unlike statistical techniques, criminologists and criminal justice researchers play a major role in controlling and shaping the data they use. The form and content of data collection can greatly expand or limit the range of questions that scholars might address. Progress in the field then heavily depends on measurement decisions.

Fifty years ago, data and measurement also posed daunting problems for criminology and criminal justice scholarship. The chapters in this volume show that researchers have made considerable progress in addressing measurement and conceptualization issues. Yet they also show that old problems still vex the field and that other problems, not obvious in the past, now require solution. Innovations in computing and quantitative methods have created new opportunities to explain crime and the justice process. In turn, these new opportunities demand new forms of data.

Most criminal justice textbooks distinguish three major sources of data on crime: the Uniform Crime Reports (UCR), the National Crime Victimization Survey (NCVS), and self-report surveys of criminal offending. In this volume, Cantor and Lynch discuss self-report victimization surveys (primarily NCVS), while Thornberry and Krohn consider self-report surveys of offending. Both sets of authors show how these data collection methods evolved from modest beginnings to reach their current status as standard measures of crime. The authors also emphasize remaining questions about the reliability and validity of

Innovations in computing and quantitative methods have created new opportunities to explain crime and the justice process. In turn, these new opportunities demand new forms of data.

these methods, and they discuss efforts to improve the methods and extend them beyond their original uses.

The volume does not include a separate chapter on the Uniform Crime Reports. We would have preferred to include such a discussion, and the lack partly reflects our inability to agree on a satisfactory contribution. The Federal Bureau of Investigation is currently transforming UCR from an aggregate count of recorded crimes to a structure that provides information on individual incidents. Future uses of UCR data will depend on this transformation to the National Incident-Based Reporting System (NIBRS). NIBRS is still far from a complete system, however, and predictions about its uses must be largely speculative (see, e.g., the chapters in Maxfield and Maltz 1999).

Still, despite the lack of its own chapter, our authors did not neglect the current version of UCR. Caulkins, for example, examines UCR as a source of data on drug crime. Fisher and Cullen explore the differences between reported and unreported sexual victimization in both UCR and victim surveys. Maguire and Uchida explore UCR as a measure of police function.

One of the most notable characteristics of crime is that it clusters. Criminal acts do not extend evenly over space, and they are not constant over time. The first criminologists noticed these variations, and patterns in space and time were a major concern of the discipline from its beginning.

Beyond the three usual measures of crime and victimization, criminologists and criminal justice researchers, of course, use many other types of data as well. The chapters in this volume cover a wide variety of data forms and sources and discuss problems in defining concepts, obtaining measurements, and assessing reliability and validity. Maguire and Uchida consider data useful for evaluating police performance, while Keilitz describes a program for collecting data to evaluate court systems. Caulkins reviews measures of drug use, and Howard and associates assess problems and progress in collecting data on international crime rates. Cohen and Warr both consider issues in measuring outcomes and costs of crime, pointing out both the difficulties and advantages of these efforts. In a chapter that touches on many of the topics in Cantor and Lynch, Fisher and Cullen review efforts to use survey data to measure violence against women. Finally, Anselin and colleagues consider the data requirements for spatial analysis.

Data Clustering and Analytical Issues

One of the most notable characteristics of crime is that it clusters. Criminal acts do not extend evenly over space, and they are not constant over time. The first criminologists noticed these variations, and patterns in space and time were a major concern of the discipline from its beginning. As Anselin and colleagues note in chapter 4, Quetelet and other 19th-century statisticians closely studied differences in crime across communities. In 1837, Poisson derived his famous count distribution in a time-series study of criminal convictions (Stigler 1986).

At the end of the 20th century, clustering and its implications still play a central role in the study of crime. In this volume, Anselin and his coauthors address these issues in their chapter on geographical data analysis. More generally, however, clustering occurs in both temporal and cross-sectional data and in both individual and aggregate analyses. The two basic forms of analysis and two basic data structures create four possible combinations: individual temporal, aggregate cross-sectional, individual cross-sectional, and aggregate temporal. Although each combination poses special problems of its own, all four generate similar clustering issues. Cluster effects will likely continue to challenge and fascinate criminal justice scholars well into the future.

Currently, the best understood clustering issues involve aggregate temporal analyses, such as trends in drug use or the fear of crime. Here, clustering arises because observations that are close in time tend to be more similar to each other than to observations in the distant past or future. The autocorrelation that this clustering generates is the subject of a large and ever-growing statistical literature (for example, Enders 1995), from which criminologists often draw.

Researchers are also becoming increasingly sophisticated in their understanding of clustering in individual temporal analyses. Data structures of this type include, for example, the developmental studies that Thornberry and Krohn discuss. In these studies, clustering appears as persistent individual differences across study periods. Research on growth and change in criminal careers is profiting from developments in panel data analysis (for example, Diggle, Liang, and Zeger 1994; Hsiao 1986), and it seems likely that important progress will continue into the next century.

Clustering occurs in aggregate cross-sectional studies in “contextual” or “multi-level” analyses. Examples include studies of the effects of neighborhood conditions on victimization risks and sentencing outcomes for defendants in different court systems. Here, researchers examine both individual effects and the aggregate effects of the clusters. Statistical models for this situation were the subject of much attention in the 1990s (for example, Bryk and Raudenbush 1992).

These models have a great deal in common with panel data analysis, and developments in each area will likely enrich the other.

Finally, clustering in individual cross-sectional analyses appears from similarities in nearby units. Anselin and his coauthors provide a comprehensive and detailed review of work in this area. Although Anselin and colleagues are the only authors in the volume who explicitly consider analytic matters, the issues that they raise apply to other clustering situations as well. For example, what types of data arrangements are most helpful in visualizing clusters? How might one estimate the strength of relationships between clusters? When should one regard clustering as a statistical nuisance to avoid, and when should one regard it as a substantive opportunity?

The last question has especially important implications. As Anselin and associates note, researchers who study spatial data usually work with units such as counties, census tracts, or neighborhoods, whose boundaries are defined by others. Crime and victimization tend to overlap these units, creating correlations between crime patterns in adjacent areas.

Researchers studying spatial data must decide whether the geographic correlations are a nuisance that requires correction or a substantive phenomenon that is interesting in its own right. In the first case, the correlations violate the assumptions of conventional methods, and one should adjust the results to remove their impact. In the second case, the correlations are evidence of social interactions between the units, and one should incorporate these effects directly into the analysis.

As Anselin and colleagues point out, each possibility produces a similar data pattern, making it difficult to choose between them on statistical grounds alone. Yet the two possibilities require very different models, and an incorrect model will inevitably harm the conclusions.

In various guises, similar issues arise in the other types of analyses. Researchers studying individual temporal data face problems in distinguishing between heterogeneity and state dependence. That is, do offenders persist in crime because of their personal characteristics or because earlier offenses changed the circumstances of their lives? Similarly, aggregate individual studies must separate compositional effects from contextual effects. Are persons more likely to be victimized in “bad” neighborhoods because of neighborhood characteristics or because more potential victims (and offenders) live in these areas? Finally, aggregate temporal studies present problems in choosing between dynamics and autocorrelated errors. Are the residuals of a time-series regression correlated because of omitted variables or because the effects of the included variables are distributed over time?

Although future statistical developments will likely make it easier to detect the underlying form of clustering, theoretical and data collection efforts also could play major roles in resolving the problems. Stronger theory would be helpful in deciding when to expect one type of clustering instead of the other. Creative data collection could make it possible to eliminate one pattern, leaving the other as the only possibility.

As we suggested earlier, statistical techniques for understanding clustering effects are most likely to come from other areas of study. Yet given criminology's long interest in these effects, the field stands in a strong position to take advantage of this work.

Besides the issues that we have discussed here, other approaches to analyzing clusters might also be useful for research in criminal justice. Procedures for nonlinear modeling that originated in the physical sciences, for example, may yield important insights into crime. Methods such as state-space reconstruction and neural network forecasting (see, e.g., Weigend and Gershenfeld 1994) may be helpful in understanding crime rate changes. Nonlinear models for "small world" networks might be useful in explaining the transmission of criminal behavior (Watts 1999).

These possibilities are entirely speculative, of course. With a few notable exceptions, nonlinear methods have yet to prove their utility in the social sciences (Granger and Teräsvirta 1993). Still, the possibilities illustrate the larger point that criminal justice research will benefit from exciting developments in many fields of study. Rather than being an embarrassment, the eclectic nature of criminal justice methods frees the field to select whatever techniques are most useful.

Uses of Research in Decisionmaking

We examine data against the framework of meaning provided by the concepts we define and connect to each other. Research is careful with processes of collecting data, from developing measurement models that relate observables to concepts, to sampling so that population values can be estimated, to assessing the validity in conceptual and causal models. The quality of data is built into the processes of collecting it. It is also true that the processes for the use of data are also, to a large degree, designed in (or designed out) of our data collection. Data can be collected, or more precisely research can be done, with the intent to use it to produce change in a social system (Hornstein et al. 1971, 257). Characteristics of research with the intent of practical use may differ somewhat from the characteristics of research conducted without this intent. Any discussion of measurement

and analysis in criminal justice would be incomplete without attending to the design-for-use as well as to the design-for-research.

Many chapters in this volume are concerned with the processes that permit use of data by researchers in building knowledge. But another theme in these chapters is the use of data in shaping the everyday affairs of criminal justice: in policymaking, in management, and in individual decisionmaking. Several of the chapters in this volume consider some of the research-use issues directly. For example, Anselin and colleagues discuss spatially organized data about crime in police strategic and tactical choices. Caulkins examines the particular kind of drug-related information most relevant to particular drug control policies. Maguire and Uchida review police agency efforts to describe what they are and what they do (often to permit police managers to make meaningful comparisons of agencies). And Keilitz describes a long-term, ongoing effort to develop outcome measures of trial court performance that practitioners would use on a regular basis to monitor the quality of court services.

Our authors' description of the utilization of research suggests an uneasy alliance between research and practice. In this section, we will review briefly some of the more typical conflicts between research and practice. We will then argue that these tensions, although cause for concern, highlight the interdependence of research and practice about crime and criminal justice. Finally, we explore some recent steps taken to learn more about connecting research and action to reduce the conflicts and increase the cooperation in this interdependent relationship.

The crime and criminal justice research agenda is undoubtedly driven more by political and ideological shifts than by the incremental and developmental process of theory building and theory testing.

Research-practice conflicts

The inherent conflicts between criminal justice research and criminal justice policymaking are probably more often recognized than their possible complementarity. For example, researchers often argue that policy is made without reference to research. Policymakers may select an approach to a problem that implicitly or explicitly contradicts research knowledge about the causes of the problem (e.g., Wilkins 1991). Or policymakers may claim policy results that research findings contradict (e.g., Lerman 1975 on the cost savings of the California Probation Subsidy; more recently, see the political responses to D.A.R.E.[®] evaluations and to evaluations of three-strikes legislation). It is often argued that policymakers use research selectively to marshal support

for a program rather than use research dispassionately to draw balanced conclusions (Ilchman and Uphoff 1983; Fisher and Cullen in this volume).

Policymakers have equally apparent complaints about research and researchers. Research can be too expensive and too slow. Despite a great deal of research, there is never research relevant to the issue at hand. Researchers design research programs to gain understanding of a problem but report that none of the critical independent variables are amenable to intervention. Researchers obfuscate, complicate, and ignore the obvious (Klockars 1999). Many of these policy and practitioner complaints appear to motivate the current "common sense movement" in criminal justice (e.g., Lapp 1998), which appears to reject the use of research in crime policy and criminal justice decisions.

These commonly voiced complaints indicate several types of conflicts, some of which are quite serious. The vexation may be increasing. Nevertheless, these conflicts also serve as indicators that researchers and policymakers are enmeshed in a system on which both are dependent. The majority of criminal justice research dollars are not grants for pure or basic research, competitively awarded on the argument of how and how well the study advances theoretical knowledge. Instead, the available research dollars tend to shift as the policy agenda changes.

Research on officials' exercise of discretion and the role of nonlegal variables in the implementation of the law blossomed in the 1960s and 1970s. This spawned and was driven by outrage that in fact the law, far from determining the behavior of criminal justice authorities, was only one of many influences, some of which seemed insidious to reformers (such as race or wealth). Despite evidence that some decisions and policies are inequitable (e.g., Tonry 1994), research on decisionmaking and decision control has waned, to be replaced in large part by concerns for control of crime, especially in the control of illicit drugs. Earlier interest in therapeutic, mediating, and conciliatory approaches to matters once considered noncriminal (domestic disputes) has been overwhelmed by victims' rights groups demanding that spouse abuse be responded to as a crime. Research on victimization of women has followed the policy change and focuses now on measures of spouse abuse and sexual victimization and the effects of arrest and prosecution. Research on providing service to or meeting the needs of offenders in community correctional programs, which was common in the Reintegration Era from 1965 to 1974, virtually disappeared when correctional policy shifted toward retribution and deterrence.

The crime and criminal justice research agenda is undoubtedly driven more by political and ideological shifts than by the incremental and developmental process of theory building and theory testing. Indeed, the differences in style

and direction with which policymakers and researchers use research appear to be one of the main criticisms behind the academic backlash to the “nothing works” conclusion and the dumping of treatment for control-oriented programs. A scientific, theoretically driven approach to correctional treatment would have examined connections between processes and outcomes to improve, incrementally, the achievement of outcomes. The policy approach was not to improve treatment incrementally but, rather, to shift suddenly to another set of strategies for which there was little evidence of effectiveness and little attention to explicit connections between process and outcome. That is, the policy response to research was “all or nothing,” rejecting scientific incrementalism in the collection and use of treatment data and accepting another set of policies for which there were no data.

Many years ago, when considering this problem, a senior researcher who had been quite successful in using policy research to advance basic theory building explained his approach as “learning to hide a theoretical design in a proposal for policy research.” What he meant was that the researcher’s obligation to advance scientific knowledge often had to be added on to a project or hidden in a proposal that would be accepted on the basis of its policy pertinence. At least some portion of the research-policy conflict from the side of researchers is that the research should also have some scientific or theory relevance. Researchers may try to achieve this goal with research funds that have been allocated in a political process. They may often feel they are rushing to find scientific answers before the topic loses policy salience.

The needs of policy-makers and practitioners for research are visible in a number of ways. Perhaps the most important of these is that the rationalization of public policy has become an institutionalized value in the public policy environment.

At the same time, it is a fact that most crime and justice research is funded with public dollars on the presumption that their expenditure will achieve practical benefit. How well have the measures, data, and analysis served practitioners and policymakers? What is the researchers’ obligation in the area of crime and criminal justice to meet these goals? How well has research benefited policy, and can it do more?

Research-practice interdependence

There are, of course, examples that, on their face, suggest that research affects policy choices eventually, if not always with the immediacy that researchers might desire or in the direction that the researchers might have predicted. Among the best known of these examples, currently, is the collective impact of the Kansas City Preventive Patrol Experiment (Kelling et al.

1974) and the response time (Tien, Simon, and Larson 1978) and expert investigation studies (Greenwood, Petersilia, and Chaiken 1977) on the professional policing paradigm. *Delinquency in a Birth Cohort* (Wolfgang, Figlio, and Sellin 1972) has had significant effects on investigative, prosecutorial, and correctional policies with the intention of intercepting high-rate criminal careers. The development of victimization reporting (Cantor and Lynch in this volume) appears to have had significant effects on crime prevention policies. The seminal research in parole decisionmaking has spawned guideline approaches to all kinds of criminal process decisions (Wilkins et al. 1976; Knapp 1984; Petersilia and Turner 1987).

Examination of these apparent effects suggests that the linkages between the research and the policy influence are complex and difficult to establish. The policies may have changed anyway, and for reasons other than the research findings. The policies that explicitly allege the paternity of these research findings are, no doubt, using the research findings in a justificatory, rather than a generative, manner.

Nonetheless, just as researchers are uncomfortably dependent on a policy-research system to support research, policymakers and practitioners at all levels are also visibly dependent on the same policy-research system to support policy. The needs of policymakers and practitioners for research are visible in a number of ways. Perhaps the most important of these is that the rationalization of public policy has become an institutionalized value in the public policy environment (Meyer and Rowan 1977). Public agencies of all sorts—and criminal justice agencies are no exception—are expected to justify choices on the basis of evidence (Meyer 1994). One of the public organization responses to the pressure for rationalizing practice is structural: the creation of planning, research, and evaluation units staffed with persons trained in criminology, criminal justice, and policy analysis. The addition of such units can lead to symbolic, rather than technical, uses of research (e.g., Simon 1993; Feeley and Simon 1993). But there is little doubt that criminal justice agencies are better able to understand and perhaps to use research than has previously been the case. Moreover, the appreciation and use of research and research products has not been ghettoized in special units. Line executives in criminal justice are better educated than ever before (e.g., Carter and Sapp 1990) and are more intelligent and demanding consumers of research than previously (Langworthy 1999). Whatever their shortcomings, the various Federal progeny of the President's Commission on Law Enforcement and Administration of Justice and the Omnibus Crime Control and Safe Streets Act of 1968 have had a significant impact on these trends.

In addition to these changes in criminal justice, there is increased pressure on the government generally to be managed by demonstrable results (Osborne and Gaebler 1992). Total quality management, continuous quality improvement, performance-based management, and similar strategies can be hollow exercises (Zbaracki 1998), especially in such areas as criminal justice, where goal ambiguity is as much a political strategy as it is poor management (Lipsky 1980). But done correctly, connecting data interactively to policy and practice decisions has the potential for major benefits to both how well policy is implemented and how much we understand about the connections between outcomes and policy and practice (Senge 1990).

Within criminal justice, there have been some attempts to conceptualize the outcomes that agencies might seek to achieve (U.S. Department of Justice, Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures 1993; Boone and Fulton 1995) and models for connecting practice to outcome measures (Sherman 1998; Maxfield and Przybylski 1999). The Violent Crime Control and Law Enforcement Act of 1994 has had some influence on attempts to document criminal justice effects. Examples are emerging from the field in which agencies or collections of agencies attempt to institute such measures and to manage by them, the largest of which is probably the trial court standards project reviewed by Keilitz in this volume.

Research on the use of research

The mood is right to transform the interdependence of researchers and practitioners into a new collaboration. But how does one avoid the shadow, shallow, or symbolic invocation of research and develop a new system in which research is actually used in practice rather than merely having “implications” for practice? Such a system would have to integrate research into everyday practice, requiring considerable change on the part of practitioners. But such a system also has to integrate practice into the conduct of research, requiring equally striking changes on the part of researchers. That kind of integration implies that the needs of both groups have to be met simultaneously. What practitioners appear to need is not so much information on what works, but a workable strategy for how to make it work. What researchers appear to need is not so much access to data or practitioner cooperation, but a workable strategy for systematically building knowledge while participating in practice decisions. These needs both imply the importance of developing theory and are therefore not so far apart as they may appear on occasion (Bernard and Ritti 1990).

There is a lengthy literature on research about the use of research. Unfortunately, it is an eclectic literature, with roots as dispersed as research and development in agriculture and education (Chin and Benne 1969). But these roots also extend

into mental health practice (e.g., Fairweather, Sanders, and Tornatzky 1974), social work innovation (Rothman, Erlich, and Teresa 1981), and criminal justice (Johnson 1980).

Although it is risky to generalize across this diverse inquiry, a consistent finding in these studies is that new practices are not adopted because researchers can show them to be more effective (measured in some way) than other practices. Adoption is not often contingent on (some researchers' estimate of) the technical or rational need for the data. Instead, policy and practice adoption of research findings come through and are mediated by human networks. Research is used when the proximal agent of change is credible to potential users. The most credible agent is usually a member of the system (Lewin 1947; Hornstein et al. 1971; Fairweather, Sanders, and Tornatzky 1974; Rothman, Erlich, and Teresa 1981).

The importance of the relational network of the agent to the adoption of research does not necessarily indicate that practitioners do not respond rationally to the content of research in determining what to do. But it does suggest that joint participation of the user and researcher in the design, conduct, and interpretation of research is critical to capturing all the elements in the practitioners' criteria for decisionmaking. Commitment to using data is most often obtained when the users meaningfully participate in framing and addressing both the research problem and the practical problem. The research problem has to be related to the practical problem. Lewin (1947) coined the term "action research" to refer to the new information and decision system that can emerge from such collaboration.

In Lewin's view, without connecting explicit goals, practice options, and feedback on their effects, decisions in human groups would be based on "sentiment" (pp. 342–344). In criminal justice, the dominant sentiments are usually of two types (Lipsky 1980). They are either deeply held convictions about what should be done (e.g., probation supervision should involve treatment processes) or deeply ingrained habits that solve significant day-to-day work problems (e.g., what is the best way to process a large number of cases with limited resources). Without feedback about the results of choices, either set of sentiments can be protected with a variety of stock-in-trade rationalizations that may bear no resemblance to actual practice outcomes (e.g., Stageberg 1990).

The field could use more systematic research on the forces that give rise to various social control policy selections and on the political and social contexts in which policies are adopted and rejected.

The National Institute of Justice has begun to work on the development of action research in criminal justice through a variety of partnership programs and has also supported research on the characteristics of more and less viable action research systems (McEwen 1999). As Hornstein and his colleagues (1971) indicated some time ago, there are some real tradeoffs that need to be carefully assessed in building such systems. Among the more problematic are researchers' fears about losing precision and practitioners' fears that the research process cannot match the demands of funding and policy cycles and that research will lead to punishment of poor performance (Keilitz in this volume).

Although there are significant obstacles in initiating such systems, these are probably less severe than the obstacles in maintaining them. There are numerous examples of pilot or demonstration action research systems, and much of the research on the difficulties in implementation has concentrated on getting these up and running. There are far fewer examples of action research programs that have been maintained long enough and have sufficiently diffused through the practitioner-researcher system that they have affected the nature of policymaking rather than of specific programs. In other words, a key long-term question is: Can action research move from front-line innovation to affect policymaking?

We know significantly less about this part of criminal justice decisionmaking than we do about decisions and practices by front-line officials. This deficiency would call for research that is infrequently supported, perhaps because the criminal justice policymaking system has a significant role in determining what to research, as previously discussed. The research that does exist is highly speculative, inductive, and case specific (e.g., Griset 1996; Chiricos 1998; Wilkins 1991; Scheingold 1991). Critical components of the policy system that need examination appear to be the complex interplay of political campaigning, the media, and the shaping and interpretation of public opinion. There is certainly no consensus on how these pieces work separately, let alone how they work together.

As weak as this research is, it suggests that much policymaking is politically opportunistic and heavily affected by extreme positions of various interest groups, especially in more centralized political systems. While policymakers are influenced by the institutionalized value of rationalized practice—of showing results—it also appears that the urge for demonstrable results is largely contained to evaluating isolated programs (and often to those that are falling out of favor). The field could use more systematic research on the forces that give rise to various social control policy selections and on the political and social contexts in which policies are adopted and rejected. This research is probably more threatening to policymakers than research on whether particular programs work or not, or how they work. Nevertheless, determining whether an action research

system can be maintained will require more attention to this level of decision-making to determine if control of sentiment at the line level may be supported by controlling the use of sentiment at the policy level. A recent survey of probation officers in New York about the use of outcome measures in guiding probation practice indicates that officers are doubtful about the ability of county and State policymakers to support research-driven changes in practice (Duffee et al. 1999). The officers expressed cynicism that policymakers would adhere to findings about probation effectiveness in future funding and program decisions.

Lipsky (1980) and Braithwaite and Mugford (1994) argue that one potentially powerful force for maintaining action research, once it is initiated, is the natural curiosity of front-line staff to determine the effects of their actions and their commitment to be effective. If this speculation is correct, one might ask whether sufficient proliferation of action research at the bottom could tip the balance toward more rational action at the top. Will front-line staff champion a theory-driven action research system as a means of doing business? For example, would they object to the cancellation of programs that they knew to be efficacious?

The available research on the policymaking system might suggest that commitments from front-line practitioners to action research may be insufficient. Greater involvement of citizens in criminal justice and criminal justice research may also be necessary. Can the current means of assessing and shaping public opinion about what to do be altered through greater involvement of citizens in action research? The greatest potential for such involvement probably is located in the current practice of problem solving in the various forms of community justice (Karp 1997). However, the involvement of citizens in action research, other than as subjects or respondents, appears limited. Very few resources go into training citizens to participate actively in this process (Friedman 1994; Duffee 1998). Promoting such citizen involvement may be a more effective way to engage them in criminal justice than urging them to support police efforts or to mobilize for crime control (Buerger 1994). If citizens and practitioners were engaged in the same action research system, this coupling could provide the reconnection of street-level bureaucrats and their "clients," which Lipsky (1980) placed at the centerpiece of public organization reform.

The trick in linking researchers more closely to the development of crime policy and the administration of justice is in maintaining the degree of independence necessary to produce a disinterested analysis, one not biased by the need to advance a particular political ideology or goal. This is not the place for a disquisition on this classic dilemma, but the risks and costs of researchers selling out or becoming invested in a policy or practice, rather than in what the best evidence shows, are both significant. We also note that research needs to

lead as much as it needs to follow the agenda of policymakers and practitioners. When researchers conduct projects on topics that are not in vogue and seem irrelevant to the issues confronting those who must deal with crime and administer justice, they nonetheless may illuminate new areas worthy of attention and refresh debate about what matters.

Conclusions

The chapters in this volume cover a wide range of issues under the umbrella of data and methods. The authors define the current state of methodology in studies of crime and the justice system, and they provide a foundation for improving research in the 21st century. In this introductory essay, we have sought to identify a few themes common to the papers and to consider in a more general context the issues that they raise. We conclude by summarizing these major themes and by noting the lessons that they provide for future progress in data collection and analysis efforts.

The first message from the authors is that theory plays a central role in methodology. Collectively, the authors argue that measurement and analysis decisions should not merely be data driven or opportunity driven. Instead, they must be guided by theory, and they should contribute to theoretical knowledge. Our authors point to missed opportunities for collecting theoretically important data in the past, and they describe how data collection efforts in the next century may enhance theory. Implicit in their arguments is a desire to use theory development as a criterion for justifying expenditures on research. In practical terms, this means that granting agencies should ensure that project proposals, the review process, and funding decisions heavily weigh possible theoretical advances.

Second, the authors stress the need for more careful attention to measurement and data collection. Improvements in analytical methods and computing power are certain to continue into the 21st century, perhaps at an exponentially increasing rate. Yet the authors raise a sobering point: Without advances in data collection and measurement, we will be ill equipped to take full advantage of the new opportunities. While researchers in criminology and criminal justice can profit from methodological work in diverse fields of study, ultimately they alone are responsible for the quality of their data.

The third theme that runs through the papers is the challenge of analyzing clustered data. Although few authors consider the issues directly, the opportunities and difficulties that clusters present are implicit in many of their discussions of measurement efforts. Clustering patterns are important in understanding offending, victimization, sentencing outcomes, and spatial variations in crime, to name

but a few topics. Even when researchers are not explicitly concerned with clusters themselves, clustering issues arise naturally from the data that they analyze. Clustering is therefore tied to theoretical concerns through attempts to understand clustering phenomena and to measurement concerns through efforts to gather data that exploit the information that clusters provide.

The fourth general theme of the papers is the use, abuse, and frequent disregard for research in criminal justice decisionmaking. Many of our authors provide examples of how research has helped resolve difficult policy decisions. At the same time, they also point to opportunistic abuses of criminal justice research that have hindered, rather than helped, the pursuit of justice. We examined several factors that influence the use of research in policymaking, and we noted that the adoption of research findings most often comes through human networks. Designing networks that simultaneously serve the needs of both researchers and practitioners, although difficult, is possible.

The themes of theory, measurement, data clustering, and uses of research connect all of the chapters in this volume. Yet no introduction of this size can do justice to the diversity of issues raised by the authors. Readers will doubtless find other commonalities among the papers as they proceed through the volume. We encourage them to do so, and we believe that they will find the trip useful.

References

- Bernard, Thomas, and R. Richard Ritti. 1990. The role of theory in scientific research. In *Measurement issues in criminology*, edited by K. Kempf. New York: Springer-Verlag.
- Blalock, H.M., Jr. 1979. The Presidential address—Measurement and conceptualization problems: The major obstacle to integrating theory and research. *American Sociological Review* 44:881–894.
- Boone, Harry N., and Betsy A. Fulton. 1995. *Results-driven management: Implementing performance-based measures in community corrections*. Lexington, Kentucky: American Probation and Parole Association.
- Braithwaite, John, and Stephen Mugford. 1994. Conditions of successful reintegration ceremonies: Dealing with juvenile offenders. *British Journal of Criminology* 34:139–171.
- Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, California: Sage Publications.
- Buerger, Michael E. 1994. A tale of two targets: Limitations of community anticrime actions. *Crime & Delinquency* 40:411–436.

- Carter, David L., and Allen D. Sapp. 1990. The evolution of higher education in law enforcement: Preliminary findings from a national survey. *Journal of Criminal Justice Education* 1:59–86.
- Chin, Robert, and Kenneth D. Benne. 1969. General strategies for effecting change in human systems. In *The planning of change*. 2d ed. Edited by W.G. Bennis, K.D. Benne, and R. Chin. New York: Holt, Rinehart and Winston.
- Chiricos, Ted. 1998. The media, moral panics, and the politics of crime control. In *The criminal justice system: Politics and policies*. 7th ed. Edited by G.F. Cole and M.G. Gertz. Belmont, California: West/Wadsworth Publishing Company.
- Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of longitudinal data*. New York: Oxford University Press.
- Duffee, David E. 1999. *Executive summary: New York State probation outcomes project phase one*. Albany: New York State Division of Probation and Correctional Alternatives.
- . 1998. Democratizing research. Paper presented at the 1998 Annual Meeting of the American Society of Criminology, 14 November, Washington, D.C.
- Duffee, David E., Jack Kryzak, Thomas Sarchioto, and Thomas Slater. 1999. New York State probation outcomes project. Paper presented at the Annual Training Institute of the American Probation and Parole Association, 23 August, New York City.
- Enders, Walter. 1995. *Applied econometric time series*. New York: John Wiley & Sons.
- Fairweather, George W., D.H. Sanders, and Louis G. Tornatzky. 1974. *Creating change in mental health organizations*. New York: Pergamon.
- Feeley, Malcolm, and Jonathan Simon. 1993. The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology* 30:449–471.
- Friedman, Warren. 1994. The community role in community policing. In *The challenge of community policing: Testing the promises*, edited by D. Rosenbaum. Thousand Oaks, California: Sage Publications.
- Granger, Clive W.J., and Timo Teräsvirta. 1993. *Modeling nonlinear economic relationships*. New York: Oxford University Press.
- Greenwood, Peter W., Joan Petersilia, and Jan Chaiken. 1977. *The criminal investigation process*. Lexington, Massachusetts: Lexington Books.
- Griset, Pamala. 1996. The politics and economics of increased correctional discretion over time served: A New York case study. *Justice Quarterly* 12:307–323.

Hornstein, Harvey A., Barbara Benedict Bunker, W. Warner Burke, Marion Gindes, and Roy J. Lewicki. 1971. Introduction to part three: Data-based strategies of social intervention. In *Social intervention*, edited by H. Hornstein, B. Bunker, W. Burke, M. Gindes, and R. Lewicki. New York: Free Press.

Hsiao, Cheng. 1986. *Analysis of panel data*. New York: Cambridge University Press.

Ilchman, Warren F., and Norman T. Uphoff. 1983. Public policy and organizational theory. In *Organizational theory and public policy*, edited by R.H. Hall and R.E. Quinn. Beverly Hills: Sage Publications.

Johnson, Knowlton. 1980. Stimulating evaluation use by integrating academia and practice. *Knowledge: Creation, Diffusion, Utilization* 2:237–262.

Karp, David, ed. 1997. *Community justice*. Lanham, Maryland: Rowman & Littlefield.

Kelling, George, Tony Pate, D. Bieckman, and C.E. Brown. 1974. *The Kansas City Preventive Patrol Experiment: Executive summary*. Washington, D.C.: Police Foundation.

Klockars, Carl B. 1999. Some really cheap ways of measuring what really matters. In *Measuring what matters: Proceedings from the Policing Research Institute meetings*, edited by R. Langworthy. Research Report, NCJ 170610. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and Office of Community Oriented Policing Services.

Knapp, Kay A. 1984. What sentencing reform in Minnesota has and has not accomplished. *Judicature* 68:181–189.

Langworthy, Robert. 1999. Measuring what matters: A Policing Research Institute. In *Measuring what matters: Proceedings from the Policing Research Institute meetings*, edited by R. Langworthy. Research Report, NCJ 170610. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and Office of Community Oriented Policing Services.

Lapp, Katherine. 1998. Jenna's Law a must for public safety. *Albany Times Union*, 9 April.

Lerman, Paul. 1975. *Community treatment and social control: A critical analysis of juvenile correctional policy*. Chicago: University of Chicago Press.

Lewin, Kurt. 1947. Group decision and social change. In *Readings in social psychology*, edited by T.M. Newcomb and E.L. Hartley. New York: Holt, Rinehart and Winston.

Lipsky, Michael. 1980. *Street-level bureaucracy: Dilemmas of the individual in public service*. New York: Russell Sage Foundation.

Maltz, Michael. 1992. Deviating from the mean: The declining significance of significance. *Journal of Research in Crime and Delinquency* 33:434–463.

- Maxfield, Michael G., and Michael D. Maltz. 1999. Special issue on the National Incident-Based Reporting System. *Journal of Quantitative Criminology* 15.
- Maxfield, Michael G., and Roger Przybylski. 1999. Frugal evaluation for justice policy decisions. Paper presented at the Annual Conference on Criminal Justice Research and Evaluation, 19 July, Washington, D.C.
- McEwen, Thomas. 1999. Evaluation of the locally initiated research partnership program. Paper presented at the annual Conference on Criminal Justice Research and Evaluation, 19 July, Washington, D.C.
- Meyer, John W. 1994. Rationalized environments. In *Institutional environments and organizations*, edited by W.R. Scott and J.W. Meyer. Thousand Oaks, California: Sage Publications.
- Meyer, John W., and Brian Rowan. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology* 83:340–363.
- Osborne, David, and Ted Gaebler. 1992. *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. Reading, Massachusetts: Addison-Wesley.
- Petersilia, Joan, and Susan Turner. 1987. Guideline-based justice: Prediction and racial minorities. In *Prediction and classification: Criminal justice decision making*, edited by D.M. Gottfredson and M. Tonry. Vol. 9 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.
- Rothman, Jack, John L. Erlich, and Joseph G. Teresa. 1981. *Changing organizations and community programs*. Newbury Park, California: Sage Publications.
- Scheingold, Stuart A. 1991. *The politics of street crime: Criminal process and cultural obsession*. Philadelphia: Temple University Press.
- . 1984. *The politics of law and order: Street crime and public policy*. New York: Longman.
- Senge, Peter M. 1990. *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday/Currency.
- Sherman, Lawrence W. 1998. *Evidence-based policing*. Washington, D.C.: Police Foundation.
- Simon, Jonathan. 1993. *Poor discipline: Parole and the social control of the underclass, 1890–1990*. Chicago: University of Chicago Press.
- Snipes, Jeffrey B., and Edward R. Maguire. 1995. Understanding criminal justice theory: First steps in forging its future. Paper presented at the 1995 Annual Meeting of the Academy of Criminal Justice Sciences, 10 March, Boston.

Stageberg, Paul. 1990. An examination of officials' perceptions of the operation of the Polk County Department of Correctional Services. In *Community corrections: A community field approach*, edited by D.E. Duffee and E.F. McGarrell. Cincinnati: Anderson Publishing Company.

Stigler, Stephen M. 1986. *The history of statistics: The measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

Tien, James M., James W. Simon, and Richard C. Larson. 1978. *An alternative approach in police patrol: The Wilmington split-force experiment*. Research Report, NCJ 43735. Washington, D.C.: U.S. Department of Justice, Law Enforcement Assistance Administration, and Wilmington (Delaware) Bureau of Police.

Tonry, Michael. 1994. Racial politics, racial disparities, and the war on crime. *Crime & Delinquency* 40:475–497.

U.S. Department of Justice, Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. 1993. *Performance measures for the criminal justice system*. NCJ 143505. Washington, D.C.

Watts, Duncan J. 1999. *Small worlds: The dynamics of networks between order and randomness*. Princeton, New Jersey: Princeton University Press.

Weigend, Andreas S., and Neil A. Gershenfeld. 1994. *Time series prediction: Forecasting the future and understanding the past*. Reading, Massachusetts: Addison-Wesley.

Wilkins, Leslie T. 1991. *Punishment, crime, and market forces*. Brookfield, Vermont: Dartmouth Publishing Company.

Wilkins, Leslie T., Jack M. Kress, Don M. Gottfredson, Joseph C. Calpin, and Arthur M. Gelman. 1976. *Sentencing guidelines: Structuring judicial discretion*. Albany, New York: University at Albany, Hindelang Criminal Justice Research Center.

Wolfgang, Marvin, Robert Figlio, and Thorsten Sellin. 1972. *Delinquency in a birth cohort*. Chicago: University of Chicago Press.

Zbaracki, Mark J. 1998. The rhetoric and reality of total quality management. *Administrative Science Quarterly* 43:602–636.

The Self-Report Method for Measuring Delinquency and Crime

by Terence P. Thornberry and Marvin D. Krohn

The self-report technique is one of three major ways of measuring involvement in delinquent and criminal behavior. The basic approach of the self-report method is to ask individuals if they have engaged in delinquent or criminal behavior, and if so, how often they have done so. In this chapter, we review the origins of the self-report method in the 1950s, the growth and refinement of this measurement technique since then, and its role in criminological research, especially longitudinal research on the etiology of delinquent and criminal behavior. Particular attention is paid to assessing the reliability and validity of self-reported measures of delinquency. We also discuss specialized data collection methods, such as random response techniques and audio assisted computer-based interviewing, that have the potential to increase the accuracy of responses. Overall, we conclude that the psychometric quality of the self-report method has increased considerably since its inception in the 1950s. Although there is much room for continued improvement, self-report data appear acceptably valid and reliable for most research purposes.

Terence P. Thornberry, Ph.D., is a Professor with, and former Dean of, the School of Criminal Justice, University at Albany, State University of New York. Marvin D. Krohn, Ph.D., is a Professor in the Department of Sociology, University of Albany, State University of New York.

A
B
S
T
R
A
C
T

33

The development and widespread use of the self-report method of collecting data on delinquent and criminal behavior is one of the most important innovations in criminological research in the 20th century. Currently, this method of data collection is used extensively both within the United States and abroad (Klein 1989). Because of its common use, we often lose sight of the major impact that self-report studies have had on the research concerning the distribution and patterns of crime and delinquency, the etiology of juvenile delinquency, and the juvenile justice system, including the police and courts.

Reliance on official sources introduces layers of potential bias between the actual behavior and the data.

Thorsten Sellin made the simple but critically important observation that “the value of a crime rate for index purposes decreases as the distance from the crime itself in terms of procedure increases” (1931, 337). Thus, prison data are less useful than court or police data as a measure of actual delinquent or criminal behavior because they are generated not only by the behavior of the perpetrators of offenses, but also by the behavior of police and court officials. Moreover, the reactions of the juvenile and criminal justice systems often rely on information from victims or wit-

nesses of crime. A substantial amount of crime is not reported, but even many crimes reported or brought to the attention of law enforcement agents are not officially recorded. Thus, reliance on official sources, such as the Uniform Crime Reports (UCR) or the National Prison Statistics, introduces layers of potential bias between the actual behavior and the data. Yet, throughout the first half of this century, our understanding of the behavior of criminals and those who reacted to crime was based almost entirely on official data.

Although researchers were aware of many of these limitations, the dilemma they faced was how to obtain information closer to the source of criminal and delinquent behavior. Observing the behavior taking place would be one method, but given the illegal nature of the behavior and the potential consequences if caught, participants in crime and delinquency are reluctant to have their behavior observed. Even when observational studies were conducted, for example, in studies of gangs (e.g., Thrasher 1927), researchers could only observe a very small portion of crime that took place. Hence, although these studies generated theoretical ideas about why and how crimes took place, they had limited utility in describing the distribution and patterns of criminal behavior.

If one could not observe the behavior taking place, self-reports of delinquent and criminal behavior would be the nearest data source to the actual behavior. There was great skepticism about whether respondents would agree to tell researchers about their participation in illegal behaviors. However, early studies

(Porterfield 1943; Wallerstein and Wylie 1947) found that not only were respondents willing to self-report their delinquency and criminal behavior, they did so in surprising numbers.

Since those early studies, the self-report methodology has become much more sophisticated in design, making it more reliable and valid and extending its applicability to a myriad of issues. These developments include the use of inventories with a wide array of delinquency items incorporating serious offenses; the use of open-ended frequency response sets instead of a relatively small number of categories; and the use of followup questions to eliminate trivial, and perhaps not criminal, acts. Much work has been done on improving the reliability and validity of self-reports, including specialized techniques to enhance the quality of self-report data. The use of self-report surveys within the context of longitudinal designs has given rise to other concerns that are not as problematic in cross-sectional research, such as construct continuity and testing or panel effects.

These developments have made self-report studies an integral part of the way crime and delinquency is studied. In this chapter, we review the history of the self-report methodology, assess the psychometric properties of self-report instruments, discuss the innovative ways in which the technique has been improved, examine the particular problems in using the technique within longitudinal designs, and suggest some future directions for the application of self-reports.

Historical Perspective

Early studies on delinquency and crime in America relied on official sources of data, such as police, court, and prison records. With these data, criminologists mapped the geography of crime (Park, Burgess, and McKenzie 1928; Shaw and McKay 1942) and, to the extent possible, identified the sociodemographic characteristics of delinquents and criminals. The data indicated that crime was disproportionately located in disadvantaged areas of the city and that those convicted of crime were more likely to be of lower class status and to be minority group members.

Although relying on official sources of data to make such generalizations, many scholars recognized that these data were not ideal for the task (Merton 1938; Sutherland 1939) because they did not tap “hidden delinquency” that

The self-report methodology has become much more sophisticated in design, making it more reliable and valid and extending its applicability to a myriad of issues.

constituted the “dark figure of crime” (Gibbons 1979). An early study by Robison (1936) found that estimates of the number of delinquents doubled when they included those referred to unofficial agencies rather than sent through the Children’s Court. Moreover, she reported that social status characteristics, including race and religion, seemed to be related to where children were referred. Robison concluded that “court figures alone are not only insufficient, but also misleading” (p. 76). Similar conclusions were reached by Murphy, Shirley, and Witmer (1946), after analyzing caseworker records of boys brought to the juvenile court. They found that less than 1.5 percent of law violations in the caseworker reports had resulted in official complaints.

Gibbons (1979) credits Edwin Sutherland for providing the impetus for self-report studies. Sutherland’s (1949) landmark work on white-collar crime provided what Gibbons (p. 81) characterizes as the first important challenge to the prevailing wisdom that individuals from favored social backgrounds were unlikely to break the law. The apparent discrepancy between reports relying on official data about “street crimes” and Sutherland’s observations about crime among the upper classes led criminologists to seek alternative means of measuring crime.

Austin Porterfield (1943, 1946) provided the first published results from a self-report survey on crime. Porterfield analyzed the juvenile court records of 2,049 delinquents from the Fort Worth, Texas, area and identified 55 offenses for which they had been adjudicated delinquent. He then surveyed 200 men and 137 women from three colleges in northern Texas to determine if and how frequently they had committed any of the 55 offenses. He found that every one of the college students had committed at least one of these offenses. The offenses committed by the college students were as serious as those committed by the adjudicated delinquents (although not as frequent), yet few of the college students had come into contact with legal authorities.

Inspired by Porterfield’s findings, Wallerstein and Wylie (1947) sampled a group of 1,698 adult men and women and examined self-reports of their delinquent behavior committed before the age of 16. They mailed questionnaires containing 49 offenses to their sample. Almost all reported committing at least one delinquent act, and 64 percent of the men and 29 percent of the women had committed at least 1 of the 14 felonies included on their checklist.

The Porterfield and the Wallerstein and Wylie studies are methodologically unsophisticated. Evaluated on criteria used today, they are problematic in terms of sample representivity, selection of delinquency items, failure to examine the reliability and validity of these items, and reliance on descriptive analysis to examine poorly stated hypotheses. They are still landmark studies in the history

of self-report methodology, however, because they not only alerted criminologists to the existence of extensive hidden delinquency, but demonstrated a methodology for measuring such behavior.

Although the contributions of Porterfield and Wallerstein and Wylie are significant developments in the self-report methodology, the work of James Short and F. Ivan Nye (1957, 1958) “revolutionized ideas about the feasibility of using survey procedures with a hitherto taboo topic” and changed the thinking about delinquent behavior itself (Hindelang, Hirschi, and Weis, 1981, 23). What distinguishes Short and Nye’s research from previous self-report methods is their attention to methodological issues—such as scale construction, reliability and validity, and sampling—and their explicit focus on the substantive relationship between social class and delinquent behavior.

Short and Nye collected self-report data from high school students in three Western communities varying in population from 10,000 to 40,000; from three Midwestern communities varying across rural, rural-urban fringe, and suburban areas; and from a training school for delinquents in a Western State. A 21-item list of criminal and antisocial behaviors was used to measure delinquency although most of their analyses employed a scale composed of a subset of only 7 items. Focusing on the relationship between delinquent behavior and the socioeconomic status (SES) of the adolescents’ parents, Nye, Short, and Olson (1958) found that, among the different SES groups, relatively few differences in delinquent behavior were statistically significant.

Short and Nye’s work stimulated much interest in both the use of the self-report methodology and the substantive issue concerning the relationship between some measure of social status (socioeconomic status, ethnicity, race) and delinquent behavior. The failure to find a relationship between social status and delinquency challenged prevailing theories built on the assumption that an inverse relationship did in fact exist, and suggested that the juvenile justice system might be using extralegal factors in making decisions concerning juveniles who misbehave. A number of studies in the late 1950s and early 1960s used self-reports to examine the relationship between social status and delinquent behavior (Akers 1964; Clark and Wenninger 1962; Dentler and Monroe 1961; Empey and Erickson 1966; Erickson and Empey 1963; Gold 1966; Slocum and Stone 1963; Vaz 1966; Voss 1966). These studies advanced the use of the self-report method by applying it to different, more ethnically diverse populations (Gold 1966; Clark and Wenninger 1962; Voss 1966), attending to issues concerning validity and reliability (Gold 1966; Clark and Tifft 1966; Dentler and Monroe 1961), and constructing measures of delinquency that specifically addressed issues regarding offense seriousness and frequency (Gold 1966). These studies found that although most juveniles engaged in some delinquency,

relatively few committed serious delinquency repeatedly. For example, Gold (1966) found that 88 percent of his sample committed one or more delinquent acts, but only 6 percent of the boys and none of the girls committed armed robbery. With few exceptions, these studies supported Short and Nye's general conclusion that if there were any statistically significant relationship between measures of social status and self-reported delinquent behavior, it was weak and did not mirror the findings of studies using official data sources.

During the 1960s, researchers began to recognize the true potential of the self-report methodology. By including questions about other aspects of adolescent life with a delinquency scale in the same questionnaire, researchers could explore etiological issues. Theoretically interesting issues concerning the family (Nye and Olson 1958; Dentler and Monroe 1961; Voss 1964; Stanfield 1966; Gold 1970), peers (Short 1957; Voss 1964; Reiss and Rhodes 1964; Matthews 1968; Erickson and Empey 1963; Gold 1970), and school (Reiss and Rhodes 1963; Elliott 1966; Kelly 1974; Polk 1969; Gold 1970) emerged as the central focus of self-report studies. The potential of the self-report methodology in examining etiological theories of delinquency was perhaps best displayed in Travis Hirschi's (1969) *Causes of Delinquency*.

The use of self-report studies to examine theoretical issues continued throughout the 1970s. In addition to several partial replications of Hirschi's arguments (Conger 1976; Hepburn 1976; Hindelang 1973; Jensen and Eve 1976), other theoretical perspectives such as social learning theory (Akers et al. 1979), self-concept theory (Jensen 1973; Kaplan 1972), strain theory (Elliott and Voss 1974; Johnson 1979), and deterrence theory (Waldo and Chiricos 1972; Silberman 1976; Jensen, Erickson, and Gibbs 1978; Anderson, Chiricos, and Waldo 1977) were evaluated using data from self-report surveys.

Another development during this period of time was the introduction of national surveys of delinquency and drug use. Williams and Gold (1972) conducted the first nationwide survey, with a probability sample of 847 boys and girls who were from 13 to 16 years of age. Among the issues examined was the relationship between social status characteristics and delinquent behavior, for which they found little support.

One of the larger undertakings on a national level is the National Youth Survey (NYS), conducted by Elliott and colleagues (Elliott, Huizinga, and Ageton 1985). NYS began in 1976 surveying a national probability sample of 1,725 youths ages 11 through 17. The survey design corrected a number of methodological deficiencies of prior self-report studies and has been greatly instrumental in improving measurement of self-reported delinquent behavior. NYS is also noteworthy because it is a panel design, having followed the original respondents into their thirties.

Monitoring the Future (Johnston, O'Malley, and Bachman 1996) is a national survey on drug use that has been conducted since 1975. It began as an in-school survey of a nationally representative sample of high school seniors and has since expanded to include 8th and 10th grade students. It also conducts follow-up surveys by mail on representative subsamples of respondents from the previous 12th grade sample. Its findings have been the primary source of information on the trends in drug use among youths in this country.

Despite the expanding applications of this methodology, questions remained about just what self-report instruments measure. The discrepancy in findings regarding the social status-delinquency relationship, based on self-report versus official (and victim) data, continued to perplex scholars. Self-reports have come under increasing criticism on a number of counts, including sample selection and the selection of delinquency items. Gwynn Nettler (1978, 98) stated that "an evaluation of these unofficial ways of counting crime does not fulfill the promise that they would provide a better enumeration of offensive activity." Gibbons (1979, 84) was even more critical in his summary evaluation, stating: "The burst of energy devoted to self-report studies of delinquency has apparently been exhausted. This work constituted a criminological fad that has waned, probably because such studies have not fulfilled their early promise."

Two studies were particularly instrumental in pointing to the flaws in self-report measures. Hindelang, Hirschi, and Weis (1979) illustrated the problems encountered when comparing results from studies using self-reports with those using official data. They employed a third source of data on crime—victimization data—and compared characteristics of offenders from the three data sources. They concluded that there is more similarity in those characteristics when comparing victimization data with UCR data than between self-report data and the other two sources. They argued that self-report instruments do not include many of the more serious crimes for which people are arrested, which are included in victimization surveys. Thus, self-reports tap a different domain of behaviors than either of the other two sources, and discrepancies in observed relationships when using self-reports should not be surprising. The differential domain of crime tapped by early self-report measures could also explain the discrepancy in findings regarding the association between social status and delinquency.

Elliott and Ageton (1980) also explored the methodological shortcomings of self-reports. They observed that a relatively small number of youths commit a disproportionate number of serious offenses. However, most early self-report instruments truncate the response categories for the frequency of offenses and do not include serious offenses in the inventory at all. In addition, many of the samples did not include enough high-rate offenders to clearly distinguish them

The development of instruments to better measure serious and very frequent offenses and the suggestion to acquire data from high-risk samples coincided with a substantive change in the 1980s in the focus of much criminological work on the etiology of offenders. The identification of a relatively small group of offenders who commit a disproportionate amount of crime and delinquency led to a call to focus research efforts on the "chronic" or "career" criminals.

from other delinquents. By allowing respondents to report the number of delinquent acts they committed rather than specifying an upper limit (e.g., 10 or more), and by focusing on high-rate offenders, Elliott and Ageton found relationships between engaging in serious delinquent behavior and measures of social status that are more consistent with results from studies using official data.

The Hindelang, Hirschi, and Weis (1979) and the Elliott and Ageton (1980) studies both suggested designing self-report studies so that they would acquire sufficient data from those high-rate, serious offenders most likely to come to the attention of authorities. They also suggested a number of changes in the way in which we measure self-report data to reflect the fact that some offenders contribute disproportionately to the rate of serious and violent delinquent acts.

The development of instruments to better measure serious and very frequent offenses and the suggestion to acquire data from high-risk samples coincided with a substantive change in the 1980s in the focus of much criminological work on the etiology of offenders. The identification of a relatively small group of offenders who commit a disproportionate amount of crime and delinquency led to a call to focus research efforts on the "chronic" or "career" criminals (Wolfgang, Figlio, and Sellin 1972; Blumstein et al. 1986). Blumstein and his colleagues' observation that we need to study the career of criminals—including early precursors of delinquency, maintenance through the adolescent years, and later consequences during the adult years—was particularly important in recognizing the need for examining the lifecourse development of high-risk offenders with self-report methodology.

The self-report methodology continues to advance, both in terms of its application to new substantive areas and the improvement of its design. Gibbons' (1979) suggestion that self-reports were just a fad whose use was likely to disappear is clearly wrong. Rather, with improvements in question design, administration

technique, reliability and validity, and sample selection, this technique is being used in the most innovative research on crime and delinquency. The sections that follow describe the key methodological developments that have made such applications possible.

Development of the Self-Report Method

Self-report measures of delinquent behavior have advanced remarkably in the 30-odd years since their introduction (see Thornberry 1989, 347–350). The prototypical “early” self-reported delinquency scale was developed by Short and Nye (1957; Nye and Short 1957). The inventory included 21 items, but most analyses were limited to 9, and in many cases 7, items that formed a Guttman scale of delinquency. The scale items refer to trivial forms of delinquent behavior—for example, there is no item measuring violent behavior, and the most serious theft item concerns stealing things worth less than \$2. Moreover, subjects were only afforded a four-category response set (“no,” “once or twice,” “several times,” and “often”), and the reference period for the instrument (“since you began grade school”) was both long and somewhat varied for these high school respondents.

Since its introduction by Short and Nye, considerable attention has been paid to the development and improvement of the psychometric properties of the self-report method. The most sophisticated and influential work was done by Elliott and his colleagues (Elliott and Ageton 1980; Elliott, Huizinga, and Ageton 1985; Huizinga and Elliott 1986) and by Hindelang, Hirschi, and Weis (1979, 1981). From their work, a set of characteristics for acceptable (i.e., reasonably valid and reliable) self-report scales has emerged. Four of the most salient characteristics are the inclusion of a wide array of delinquency items, serious offenses, frequency response sets, and followup questions.

Inclusion of a wide array of delinquency items

The domain of delinquency and crime covers a wide range of behaviors, from truancy and running away from home to aggravated assault and homicide. If the general domain of delinquent and criminal behavior is to be represented in a self-report scale, it is necessary for the scale to cover that same wide array of human activity. Simply asking about a handful of these behaviors does not accurately represent the theoretical construct of crime. In addition, empirical evidence suggests that crime does not have a clear unidimensional structure that would facilitate the sampling of a few items from a theoretically large pool to represent adequately the entire domain.

These considerations suggest that an adequate self-report scale for delinquency will be relatively lengthy. A large number of individual items are required to represent the entire domain of delinquent behavior, to represent each of its subdomains, and to ensure that each subdomain—e.g., violence, drug use—is itself adequately represented.

It is essential that a general self-reported delinquency scale tap serious as well as less serious behaviors. Failure to do so misrepresents the domain of delinquency and contaminates comparisons with other data sources.

Inclusion of serious offenses

Early self-report scales tended to ignore serious criminal and delinquent events and concentrated almost exclusively on minor forms of delinquency. As a result, only certain subdomains of delinquency, such as petty theft and status offenses, were measured, even though theoretical interest and conclusory statements focused on juvenile delinquency broadly construed.

It is essential that a general self-reported delinquency scale tap serious as well as less serious behaviors. Failure to do so misrepresents the domain of delinquency and contaminates comparisons with other data sources. In addition, it misrepresents the dependent variable of many delinquency theories that set out to explain serious, repetitive delinquency (e.g., Elliott, Huizinga, and Ageton 1985; Thornberry 1987).

Inclusion of frequency response sets

Many self-report studies rely on response sets with a relatively small number of categories, which tend to censor high-frequency responses. For example, Short and Nye (1957) used a four-point response, with the most extreme category being “often.” As a result, a respondent who committed a theft 5 times would be treated the same as a respondent who committed the act 50 times. Aggregated over many items, the use of limited response sets has the consequence of lumping together occasional and high-rate delinquents rather than discriminating between these behaviorally different groups.

When frequency responses are used, a number of specific indicators can be constructed from the basic inventory. The three most common are prevalence, incidence, and variety. Prevalence refers to the proportion or percentage of *people* who report involvement in delinquency—the percentage of the sample who answer “yes.” Incidence (also called frequency) refers to the number of delinquent *acts* reported—the total number of times the person reports committing different acts. Variety refers to the number of different *types* of delinquency

reported by the person. For example, if the index has six items (offense types), the variety score can vary from 0 to 6. Each of these basic measures can also be created for different time periods. One of the most common is a “lifetime” or “ever” measure; for example, an “ever-prevalence” measure of marijuana use would indicate the percentage of the people who had ever used marijuana. Most measures are time-limited; for example, referring to offenses committed during the past year or past 6 months.

Inclusion of followup questions

Self-report questions seem to have an inherent tendency to elicit reports of trivial acts that are very unlikely to elicit official reactions, or even acts that are not violations of the law. This occurs more frequently with less serious offenses but also affects responses to serious offenses. For example, respondents have listed such pranks as hiding a classmate’s books in the respondent’s locker between classes as “theft,” or roughhousing between siblings as “serious assault.”

Some effort must be made to adjust or censor the data to remove these events if the delinquency of respondents is to be reflected properly and if the rank order of respondents with respect to delinquency is to be portrayed properly. Two strategies are generally available. First, one can ask a series of followup questions designed to elicit more information about the event, such as the value of property stolen, the extent of injury to the victim, and the like. Second, one can use an open-ended question asking the respondent to describe the event, and then probe to obtain information necessary to classify the act. Both strategies have been used with some success.

Summary

Recent examinations of the self-report method have identified a number of shortcomings in earlier scales and suggested ways of improving the technique’s psychometric properties. The more salient suggestions include the following:

- Self-report scales should include a wide range of delinquent acts so that the general domain of delinquency, as well as its various subdomains, is adequately represented.
- The scale should include serious as well as minor acts.
- A frequency scale should be used to record responses so that high-rate offenders can be isolated from low-rate offenders.
- Extremely trivial, nonactionable acts that are reported should be identified and eliminated from the data.

These procedures improve our ability to identify delinquents and discriminate among different types of delinquents. Thus, they are likely to improve the validity, and to some extent the reliability, of self-report scales. These are clearly desirable qualities.

To gain these desirable qualities, however, requires a considerable expansion of the self-report schedule. This can be illustrated by describing the major components of the index currently being used in the Rochester Youth Development Study (Smith and Thornberry 1995), as well as in the other two projects of the Program of Research on the Causes and Correlates on Delinquency (see Browning et al. 1999). The inventory includes 32 items tapping general delinquency and 12 tapping drug use, for a total of 44 items. For each of these items, the subjects are asked if they ever committed the act, and if so, if they had committed the act in the past 6 months. For the most serious of each type of delinquency reported in the past 6 months, subjects are asked to describe the event by responding to the question: "Could you tell me what you did?" If that open-ended question does not elicit the information needed to describe the event adequately, a series of probe questions, which vary from 2 to 14 probes depending on the offense, are asked.

Although most of these specific questions are skipped for most subjects, since delinquency remains a rare event, this approach to measuring self-reported delinquency is a far cry from the initial method of using a few categories to respond to a small number of trivial delinquencies, with no followup items. In the remaining pages, we evaluate this approach for measuring delinquent and criminal behavior.

Reliability and Validity

For any measure to be scientifically worthwhile, it must possess both reliability and validity. Reliability is the extent to which a measuring procedure yields the same result on repeated trials. For example, if a bathroom scale were reliable, it would yield the same reading of your weight if you got on and off that scale 10 times in a row. If it were unreliable, the reading of your weight would vary somewhat, even though your true weight would not change in the space of time it would take you to get on and off the scale 10 times.

No measure is absolutely, perfectly reliable. Repeated use of a measuring instrument will always produce some variation from one application to another. That variation can be very slight to quite large. So the central question in assessing the reliability of self-reported delinquency measures is not whether the measure is reliable but *how* reliable it is; reliability is always a matter of degree.

Validity is a much more abstract notion than is reliability. The best definition of validity is something like as follows: A measure is valid to the extent to which it measures the concept you set out to measure, and nothing else. Whereas reliability focuses on a particular property of the measure—namely, its stability over repeated uses—validity concerns the crucial relationship between the theoretical concept you are attempting to measure and what you actually measure.

For example, let us say we are interested in measuring an individual's actual involvement in criminal behavior over the past year. During that time period, there are some people who never commit a crime, while others do. Our measure would be completely valid if it accurately identified as criminals all of the people who did commit crimes and also accurately identified as noncriminals all of the people who did not commit crimes. That is, the measure would accurately reflect our theoretical concept—involvement in crime during the past year. As with reliability, the assessment of validity is not an either/or proposition. There are no perfectly valid measures, but some are more valid than others.

Even though both validity and reliability are always a matter of degree, we often see statements that assert that a particular measure “is valid” or that another measure “is unreliable.” That is a shorthand way of saying that the first measure possesses sufficient validity for the analytic purpose at hand, but the second measure does not.

The relationship between validity and reliability is asymmetrical. A particular measure can be highly reliable but have little or no validity. For example, a bathroom scale could be very consistent but the calibration could be off by 50 pounds. In this case, we have very reliable measures, but every one of them would be wrong—too low or too high by 50 pounds.

In contrast, if a measure is valid, it is also reliable. Since a valid measure is one that accurately measures what it sets out to measure, by definition it must be consistent, yielding the same estimates time after time.

All scientifically adequate measures must possess high levels of both validity and reliability. We now turn to an assessment of whether self-reported measures of delinquency are psychometrically acceptable.

Assessing reliability

There are two classic ways of assessing the reliability of social science measures: “test-retest” reliability and internal consistency. Huizinga and Elliott (1986) make a convincing case that the test-retest approach is fundamentally more appropriate for assessing self-reported measures of delinquency.

Internal consistency

Internal consistency simply means that multiple items measuring the same underlying concept should be highly intercorrelated. This can be illustrated by returning to our example of weight and bathroom scales. If you weighed yourself on 15 different bathroom scales, you would get slightly different readings on each, but the answers should be highly correlated; that is, your weight should be the same. If one scale differed substantially from the others, you would likely throw it out as being inaccurate. The same approach is used in assessing attitudes and opinions. Many questions tapping the same concept are asked, and the expectation is that the answers on these items will be highly intercorrelated. For example, in assessing attachment to parents, one could ask an adolescent to respond to statements such as “I think my mother is really terrific” and “I have a great deal of respect for my mother.” It is reasonable to expect that an adolescent strongly attached to his or her mother would respond positively to both statements, and an adolescent who is very alienated from his or her mother would respond negatively. That is, across these and similar items, responses would be highly correlated.

This expectation is much less reasonable for behavioral inventories such as self-report measures of delinquency, however. Current self-report measures typically include 30 or 40 items measuring a wide array of delinquent acts. Just because someone reports being truant is no reason to expect he or she would be involved in theft or vandalism. Similarly, if someone reports being involved in assaultive behavior, there is no reason to assume that he or she has been involved in drug sales or loitering. Indeed, given the relative rarity of involvement in delinquent acts, it is very likely that most people will respond negatively to most items and affirmatively to only a few. This is especially the case if we are asking about the past year or the past 6 months. Because of this, there is no strong underlying expectation that the responses will be highly intercorrelated. Therefore, an internal consistency approach to assessing reliability is not particularly appropriate.

Test-retest reliability

Thus, we will focus on the test-retest method of assessing reliability. This approach is quite straightforward. A sample of respondents is administered a self-reported delinquency inventory (the test); then, after a short interval, the same inventory is readministered (the retest). In doing this, the same questions and the same reference period should be used at both times.

It is also important to pay attention to the time lag between the test and the retest. If it is too short, answers to the retest likely will be a function of memory;

respondents are likely to remember what they said the first time and simply repeat it. If so, estimates of reliability would be inflated. On the other hand, if the time period between the test and the retest is too long, responses to the retest would probably be less accurate than those to the test simply because of memory decay. In this case, the reliability of the scale would be underestimated. There is no hard and fast rule for assessing the appropriateness of this lag, but the optimal time lag appears to be in the range of 1 to 4 weeks.

The simplest way of deriving a reliability coefficient for the test-retest method is to correlate the first and second sets of responses. The correlations should be reasonably high, preferably in the range of 0.70 or greater.

A number of studies have assessed the test-retest reliability of self-reported delinquency measures. In general, the results of these studies indicate that these measures are acceptably reliable. The reliability coefficients vary somewhat, depending on the number and types of delinquent acts included in the index and the scoring procedures used (e.g., simple frequencies or ever-variety scores). But scores well above 0.80 are common. In summarizing previous literature in this area, Huizinga and Elliott (1986, 300) stated:

Test-retest reliabilities in the 0.85–0.99 range were reported by several studies employing various scoring schemes and numbers of items and using test-retest intervals of from less than 1 hour to over 2 months (Kulik et al., 1968; Belson, 1968; Hindelang et al., 1981; [Braukmann] et al., 1979; Patterson and Loeber, 1982; [Skolnick] et al., 1981; Clark and [Tifft], 1966; Broder and Zimmerman, 1978).

Perhaps the most comprehensive assessment of the psychometric properties of the self-report method was conducted by Hindelang, Hirschi, and Weis (1981). Their self-report inventory was quite extensive, consisting of 69 items divided into the following major subindexes: official contacts, serious crimes, delinquency, drugs, and school and family offenses. To see whether the method of administration matters, some subjects were interviewed and others responded on a questionnaire. For both types of administration, some subjects responded anonymously and others were asked to provide their names.

To maximize variation in the level of delinquency, the study sample was selected from three different populations in Seattle, Washington. The first consisted of students without an official record of delinquency attending Seattle schools. The second consisted of adolescents with a police record but no court record, and the third group consisted of adolescents with a juvenile court record. Within these three major strata, subjects were further stratified by gender, race, and, among the whites, socioeconomic status.

Several self-reported measures of delinquency were created. The major ones include an ever-variety score (the number of delinquent acts the respondents report ever having committed), a last year variety score (the same type of measure for the past year), and a last year frequency score (the total number of times respondents report committing each of the delinquent acts).

As indicated earlier, internal consistency methods can be used to assess the reliability of self-reported responses. The classic way of doing so is with Cronbach's alpha. Although mindful of the limitations of internal consistency approaches, Hindelang, Hirschi, and Weis (1981) report alpha coefficients for a variety of demographic subgroups and for the ever-variety, last year variety, and last year frequency scores. The coefficients range from 0.76 to 0.93. Most of the coefficients are above 0.8, and 8 of the 18 coefficients are above 0.9.

Hindelang, Hirschi, and Weis (1981) also estimated test-retest reliabilities for these three self-report measures for each of the demographic subgroups. Unfortunately, only 45 minutes elapsed between the test and the retest, so it is quite possible that the retest responses are strongly influenced by memory effects. Nevertheless, they report substantial degrees of reliability for the self-report measures. Indeed, most of the test-retest correlations are above 0.9.

Thus, whether an internal consistency or test-retest approach is used, the Seattle data indicate a substantial degree of reliability for a basic self-reported delinquency measure. Hindelang, Hirschi, and Weis point out that reliability scores of this magnitude are higher than those typically associated with many attitudinal measures and conclude that "the overall implication is that in many of the relations examined by researchers, the delinquency dimension is more reliably measured than are many of the attitudinal dimensions studied in the research" (1981, 82).

The other major assessment of the psychometric properties of the self-report method was conducted by Huizinga and Elliott, using data taken from the well-known National Youth Survey. NYS began in 1976 with a nationally representative sample of 1,725 American youths between the ages of 11 and 17. At the fifth interview, 177 respondents were randomly selected and reinterviewed approximately 4 weeks after their initial assessment. Based on these data, Huizinga and Elliott (1986) estimated test-retest reliability scores for the general delinquency index and for several subindexes. They also estimated reliability coefficients for frequency scores and for variety scores.

The general delinquency index appears to have an acceptable level of reliability. The test-retest correlations are 0.75 for the frequency score and 0.84 for the variety score. For the various subindexes—ranging from public disorder

offenses to the much more serious index offenses—the reliabilities vary from a low of 0.52 (for the frequency measure of felony theft) to a high of 0.93 (for the frequency measure of illegal services). In total, Huizinga and Elliott (1986) report 22 estimates of test-retest reliability—across indexes and across frequency and variety scores—and the mean reliability coefficient is 0.74.

Another way of assessing the level of test-retest reliability is by estimating the percentage of the sample who changed their frequency responses by two or less. If the measure is highly reliable, one would expect few such changes. For most subindexes, there appears to be acceptable precision and reliability based on this measure. For example, for index offenses, 97 percent of the respondents changed their answers by two delinquent acts or less. Huizinga and Elliott (1986, 303) summarize these results as follows:

Scales representing more serious, less frequently occurring offenses (index offenses, felony assault, felony theft, robbery), have the highest precision, with 96 to 100 percent agreement, followed by the less serious offenses (minor assault, minor theft, property damage), with 80 to 95 percent agreement. The public disorder and status scales have lower reliabilities (in the 40 to 70 percent agreement range), followed finally by the general SRD [self-reported delinquency] scale, which, being a composite of the other scales, not surprisingly has the lowest test-retest agreement.

Huizinga and Elliott also report little evidence of differential reliability across various subgroups. They found no consistent differences across sex, race, class, place of residence, or delinquency level in terms of test-retest reliabilities (see also Huizinga and Elliott 1983).

Summary

Overall, these studies suggest that the self-report method possesses acceptable reliability for most analytic purposes. Test-retest correlations are often 0.80 or higher, and self-reported delinquency responses are no less reliable than other social science measures. That is particularly impressive considering the sensitive nature of the topic: unreported criminal activity. Although this assessment is generally positive, it does not mean that there are no reliability problems for self-reported responses. Some subindexes have low reliabilities, and more research is needed to identify which indexes are most reliable across different samples and which are least reliable. Despite these concerns, it appears that self-reports of delinquent acts are fairly stable over time. As Hindelang, Hirschi, and Weis conclude: “If self-report measurement is flawed, it is not here, but in the area of validity” (1981, 84).

Assessing validity

Recall that validity refers to the accuracy of a measure. A measure is valid to the extent to which it accurately measures the concept that you set out to measure. There are several ways to assess validity. We will concentrate on three: content validity, construct validity, and criterion validity.

Content validity

Content validity is a subjective or logical assessment of the extent to which the measure adequately reflects the full domain, or the full content of the concept being measured. For example, if one were interested in assessing arithmetic ability among grade school children and had a test that only included questions on addition, the test would lack content validity. That is, by not containing questions to assess subtraction, multiplication, and division, the test would not measure the full domain, or the full content, of the concept of arithmetic ability.

Note that our assessment implied that we have a clear definition of what is contained in the concept of arithmetic ability. Only by knowing that arithmetic includes these four basic functions can we draw the conclusion that a test that measures only one of them is inadequate in terms of its content validity. As in all assessments of validity, content validity requires a clear theoretical definition of the concept.

To argue that a measure has content validity, we must meet the following three criteria. First, we must define the domain of the concept clearly and fully. Second, we must create questions or items to cover the whole range of the concept under investigation. And third, we must sample items or questions from that range so that the ones that appear on the test are representative of the underlying concept.

In our case, we are interested in measuring involvement in delinquency and crime. A reasonable definition of delinquency and crime is the commission of behaviors that violate criminal law and that place the individual at some risk of arrest if such a behavior were known to the police. Can we make a logical case that self-report measures of delinquency are valid in this respect?

As noted before, the earlier self-report inventories contained relatively few items to measure the full range of delinquent behavior. For example, the Short and Nye (1957) inventory only contains 21 items and most of their analysis was conducted with a 7-item index. Similarly, Hirschi's self-report measure (1969) is based on only 6 items. More importantly, the items included in these scales are clearly biased toward the minor or trivial end of the continuum. For example, Hirschi's inventory includes only one item measuring a violent crime:

“Not counting fights you may have had with a brother or sister, have you ever beaten up on anyone or hurt anyone on purpose?”

More recent self-report measures appear much better in this regard. For example, the Hindelang, Hirschi, and Weis (1981) index includes 69 items that range from status offenses, such as skipping class, to violent crimes, like serious assault and armed robbery. The NYS index (Elliott, Huizinga, and Ageton 1985) has 47 items designed to measure all but 1 (homicide) of the 8 UCR Part I offenses and 60 percent of the 21 Part II offenses, as well as offenses that juveniles are likely to commit. The self-report inventory used by the three projects of the Program of Research on the Causes and Correlates of Delinquency has 32 items measuring delinquent behavior and 12 measuring substance use. These more recent measures, although not perfect, tap into a much broader range of delinquent and criminal behavior. As a result, they appear to have reasonable content validity.

Construct validity

Construct validity refers to the extent to which the measure being validated is related in theoretically expected ways to other concepts or constructs. In our case, the key question is: Are measures of delinquency based on the self-report method correlated in expected ways with other variables?

In general, self-report measures of delinquency and crime, especially the more recent longer inventories, appear to have a high degree of construct validity. They are generally related in theoretically expected ways to basic demographic characteristics and to a host of theoretical variables drawn from various domains such as individual attributes, family structure and processes, school performance, peer relationships, and neighborhood characteristics. Hindelang, Hirschi, and Weis offer one of the clearer assessments of construct validity (1981, 127ff). They correlate a number of etiological variables with different self-report measures, collected under different conditions (e.g., interviews or questionnaires). With a few nonsystematic exceptions, the correlations are in the expected direction and of the expected magnitude.

Overall, construct validity may offer the strongest evidence for the validity of self-reported measures of delinquency and crime. Indeed, if one examines the general literature on delinquent and criminal behavior, virtually all theoretically expected relationships are actually observed for self-report measures of delinquency and crime. It is unfortunate that this approach is not used to assess validity more formally and systematically.

Criterion validity

Criterion validity “refers to the relationship between test scores and some known external criterion that adequately indicates the quantity being measured” (Huizinga and Elliott 1986, 308). There is a fundamental difficulty in assessing the criterion validity of self-reported measures of delinquency and crime and, for that matter, all measures of delinquency and crime. Namely, there is no “gold standard” against which to judge the self-report measure. That is, there is no fully accurate assessment to use as a benchmark. In contrast, to test the validity of self-reports of weight, one could ask people to self-report their weight and then weigh them on a scale, an external criterion. Given the secretiveness of criminal behavior, however, there is nothing comparable to a scale in the world of crime. As a result, the best approach is to compare different flawed measures of criminal involvement to find similar responses. The similarity of results from different measurement strategies heightens the probability that the various measures are tapping into the underlying concept of interest. Although not ideal, this is the best possible approach in this area of inquiry.

There are several ways of assessing criterion validity. One of the simplest is called “known group validity.” In this approach, one compares scores for groups of people who are likely to differ in terms of their underlying involvement in delinquency. For example, one would expect the delinquency scores of seminarians to be lower than the delinquency scores of street gang members.

Over the years, a variety of group comparisons have been made to assess the validity of self-report measures. They include comparisons between individuals with and without official arrest records, between individuals convicted and not convicted of criminal offenses, and between institutionalized adolescents and high school students. In all cases, these types of comparisons indicate that the group officially involved with the juvenile justice system self-reported substantially more delinquents act than the other group. (See, for example, the work by Hirschi, 1969; Hardt and Petersen-Hardt 1977; Erickson and Empey 1963; Farrington 1973; Hindelang, Hirschi, and Weis 1981; Short and Nye 1957; Voss 1963; Kulik, Stein, and Sarbin 1968).

Although comparisons across known groups are helpful, they offer a very minimal test of criterion validity. The real issue is not whether groups differ but whether *individuals* have similar scores on the self-report measure and on other measures of criminal behavior. As mentioned previously, the basic problem in this area is there is no perfect benchmark against which to judge the self-report measures. Thus, a variety of external criteria have been used (see the discussion in Hindelang, Hirschi, and Weis 1981, 97–101). The two most common approaches are to compare self-reported delinquency scores with official arrest records and with self-reports of official arrest records.

The premise behind these comparisons is quite simple. If the measures are valid, they should produce similar scores for both the prevalence and frequency of delinquent and criminal involvement. That is, if the self-report measure identifies certain individuals as essentially nondelinquent, we should not expect to find them in official records. In contrast, if the self-report measures identify individuals as highly delinquent, we should expect both to find them in official records and to have extensive criminal histories. If this is the case, the two measures would be positively correlated and the correlation would suggest that the measures have some degree of validity. As with reliability assessment, the most sophisticated examinations of this topic have been conducted by Hindelang, Hirschi, and Weis (1981) and Huizinga and Elliott (1986).

We can begin by examining the correlation between self-reported official contacts and official measures of delinquency as presented by Hindelang, Hirschi, and Weis (1981). In this case, the correlations are quite high, ranging from 0.70 to 0.83. Correlations of this magnitude are reasonably large for this type of data.¹ Adolescents seem quite willing to self-report their involvement with the juvenile justice system.

The generally high level of concordance between self-reports of being arrested or having a police contact and having an official record has been observed in other studies as well. For example, Hardt and Petersen-Hardt (1977) found that 78 percent of the juveniles with police records self-report that they have been arrested. Similar results are reported by Hathaway, Monachesi, and Young (1960) and, for status offenses, by Rojek (1983). When convictions are examined, even higher concordance rates are reported by Blackmore (1974) and Farrington (1977).

The most important comparison presented by Hindelang, Hirschi, and Weis (1981) is between self-reported delinquent behavior and official measures of delinquency. It is important because these are *independent* measures of an individual's involvement in delinquent behavior. One is based on self-reports and one is based on official police records. Hindelang, Hirschi, and Weis present correlations using a number of different techniques for scoring the self-report measures. However, we will focus on the average correlation across these different measures and on the correlation based on the ever-variety scores, as presented in their figure 2 (1981, 113).

Overall, these correlations are reasonably high, somewhere around 0.60 for all subjects. The most important data, though, are presented for race-by-gender groups. For white and African-American females and for white males, the correlations range from 0.58 to 0.65 when the ever-variety score is used. For correlations averaged across the different self-report measures, the magnitudes

range from 0.50 to 0.60. For African-American males, however, the correlation is at best moderate. For the ever-variety self-reported delinquency score, the correlation is 0.35, and the average across the other self-reported measures is 0.30.

Putting this together leads to a somewhat mixed assessment of the validity of self-report measures based on the Seattle data. On the one hand, the overall validity of self-report data seems to be in the moderate to strong range. For the link between self-reported delinquent behavior and official measures of delinquency (the only link based on independent sources of data), the overall correlations are smaller but still acceptable. On the other hand, if we look at the issue of differential validity, there appears to be a substantial difference between African-American males and other respondents. Official measures of delinquency and self-report measures of delinquency are not correlated very highly for African-American male adolescents. It is hard to determine whether this is a problem with the self-report measures, the official measures, or both. We will return to a discussion of this issue after additional data are presented.

The majority of individuals who have been arrested self-report their delinquent behavior, and the majority of offenses they commit are also reported.

Huizinga and Elliott (1986), using data from NYS, also examine the correspondence between self-reports of delinquent behavior and official criminal histories. They recognize that there can be considerable slippage between these two sources of data, even when the same event is recorded in both datasets. For example, an adolescent can self-report a gang fight, but it may be recorded in the arrest file as disturbing the peace; an arrest for armed robbery can be self-categorized as a mugging or theft by the individual. Because of this, Huizinga and Elliott provide two levels of matching. In one, there is “a very tight match of the self-report behavior to the arrest behavior,”

and in the second, there is a broad match “in which any self-reported offense that could conceivably have resulted in the recorded arrest was allowed” (1986, 317). The analysis provides information on both the percentages of youths who provide tight and broad matches to their arrest records and the percentage of arrests that are matched by self-reported behavior.

As expected, there are substantial differences in results, depending on whether tight or broad matches are used. For the tight matches, almost half of the respondents (48 percent) concealed or forgot at least some of their offensive behavior, and about a third (32 percent) of all of the offenses were not reported. When the broad matches are used, however, the percentage of respondents concealing or forgetting some of their offenses drops to 36 percent, and the percent of arrest

offenses not self-reported drops to 22 percent. Although the rates of underreporting are substantial, it should also be noted that the majority of individuals who have been arrested self-report their delinquent behavior, and the majority of offenses they commit are also reported.

The reporting rates for gender, race, and social class groupings are quite comparable to the overall rates, with one exception. As with the Seattle data, African-American males substantially underreport their involvement in delinquency.

The most recent major study assessing the criterion validity of self-reported measures was conducted by Farrington and colleagues (1996), using data from the middle and oldest cohorts of the Pittsburgh Youth Study. The Pittsburgh study, one of three projects in the Program of Research on the Causes and Correlates of Delinquency, uses the same self-reported delinquency index as described earlier for the Rochester Youth Development Study. In this analysis, Farrington and colleagues classified each of the boys in the Pittsburgh study into one of four categories based on their self-reports: no delinquency, minor delinquency only, moderate delinquency only, and, finally, serious delinquency. They then used juvenile court petitions as an external criterion to assess the validity of the self-reported responses. Both concurrent and predictive validity were assessed; the former used court petitions prior to the first self-report assessment, and the latter used court petitions after the first self-report assessment.

Overall, this analysis suggests that there is a substantial degree of criterion validity for the self-report inventory used in the Program of Research on the Causes and Correlates of Delinquency. Respondents in the most serious category based on their self-report responses are significantly more likely to have juvenile court petitions, both concurrently and predictively. For example, the odds ratio of having a court petition for delinquency is about 3.0 for the respondents in the most serious category versus the other three.

Farrington and colleagues (1996) also present information on the issue of differential validity. Their results indicate that African-American males are no more or less likely to self-report delinquent behavior than are white males. With few exceptions, the odds ratios comparing self-reported measures and official court petitions are significant for both African-Americans and whites. In some cases, the odds ratios are higher for whites, and in other cases, they are higher for African-Americans.

These researchers also compared the extent to which boys with official court petitions self-reported being apprehended by the police. Overall, about two-thirds of the boys with court petitions answered in the affirmative. Moreover, there was no evidence of differential validity. Indeed, the African-American

respondents were more likely to admit being apprehended by the police than were the white respondents. Farrington and colleagues conclude that “concurrent validity for admitting offenses was higher for Caucasians but concurrent validity for admitting arrests were higher for African-Americans. There were no consistent ethnic differences in predictive validity” (1996, 509).

Other studies have also examined the concordance between self-reports of delinquent behavior and official records. For example, Elliott and Voss (1974) examined this issue in a high school sample drawn from southern California. Overall, they found that 83 percent of the arrest offenses were self-reported by the respondents, but the rate varied by offense type. In general, more serious offenses were more likely to be underreported than were minor offenses. Based on a school sample from Honolulu, Voss (1963) found that 95 percent of arrest offenses were reported in the self-report inventories.

Rather than relying on police records as the external criterion, Gold (1970) relied on reports by friends and classmates. He found that, of the respondents whose friends had said they engaged in delinquent acts, 72 percent self-reported delinquencies, 17 percent concealed their delinquent acts, and, in 11 percent of cases, the outcome was uncertain.

The previous studies have all focused on types of delinquent or criminal behavior that have no true external criterion for evaluating validity. There is an external criterion for one class of criminal behavior; namely, substance use. Physiological data—for example, from saliva or urine—can be used to independently assess recent use of various substances. The physiological data can then be compared with self-reports of substance use to assess the validity of the self-report instruments. A few examples of this approach can be offered.

We begin with a study of a minor form of deviant behavior, adolescent tobacco use. Akers and colleagues (1983) examined tobacco use among a sample of junior and senior high school students in Muscatine, Iowa. The respondents provided saliva samples, which were used to detect nicotine use by the level of salivary thiocyanate. The students also self-reported whether they smoked and how often they smoked. The self-report data had very low levels of either underreporting or overreporting tobacco use. Overall, Akers and colleagues estimated that 95 to 96 percent of the self-reported responses were accurate and valid.

The Arrestee Drug Abuse Monitoring (ADAM) program, formerly the Drug Use Forecasting (DUF) program, sponsored by the National Institute of Justice, is an ongoing assessment of the extensiveness of drug use for samples of arrestees in cities throughout the country. Individuals who have been arrested and brought to central booking stations are interviewed and asked to provide

urine specimens. Both the urine samples and the interviews are provided voluntarily; there is an 80-percent cooperation rate for urine samples and a 90-percent cooperation rate for interviews. The urine specimens are tested for 10 different drugs: cocaine, opium, marijuana, PCP, methadone, benzodiazepines, methaqualone, propoxyphene, barbiturates, and amphetamines. The arrestees also are interviewed, and some interviews include a self-reported drug use inventory. Assuming that urine samples provide a reasonably accurate estimate of actual drug use, they can be used to validate self-reported information.

DUF compares 1988 urinalysis test results for male arrestees with self-reported drug use (U.S. Department of Justice 1990, 12); the results vary considerably by type of drug. There generally is a fairly high concordance for marijuana use. For example, in New York City, 28 percent of the arrestees self-report marijuana use, and 30 percent test positive for marijuana use. Similarly, in Philadelphia, 28 percent self-report marijuana use, and 32 percent test positive. The worst comparison in this particular examination of DUF data came from Houston, where 15 percent of arrestees self-report marijuana use and 43 percent test positive.

For more serious drugs, however, underreporting is much more common. For cocaine, for example, 47 percent of New York City arrestees self-reported use, while 74 percent tested positive. Similar numbers were generated in Philadelphia, where 41 percent self-reported cocaine use, but 72 percent tested positive. Similar levels of underreporting have been observed in other cities for other hard drugs, such as heroin.

The data collected in DUF differ considerably from those collected in typical self-report surveys. The sample is limited to people just arrested, who then are asked to provide self-incriminating evidence to a research team while in a central booking station. How this setting affects the results is not entirely clear. On the one hand, individuals are likely to be reluctant to provide additional self-incriminating evidence after having just been arrested. On the other hand, if one has just been arrested for a serious crime like robbery, auto theft, or burglary, admitting to recent drug use may not be considered a big deal. In any event, one has to be cautious in using these data to generalize to the validity of typical self-report inventories.

Summary

We have examined three different approaches to assessing the validity of self-reported measures of delinquency and crime: content validity, construct validity, and criterion validity. Several conclusions appear warranted, especially for the more recent self-report inventories.

On the one hand, the self-report method for measuring this rather sensitive topic—undetected criminal behavior—appears to be reasonably valid. The content validity of the recent inventories is acceptable, the construct validity is quite high, and the criterion validity appears to be in the moderate-to-strong range. Putting this all together, one could conclude that for most analytic purposes, self-reported measures are acceptably accurate and valid.

On the other hand, despite this general conclusion, there are still several substantial issues concerning the validity of self-report measures. First, the validity of the earlier self-report scales, and the results based on them, are at best questionable. Second, based on the results of the tests of criterion validity, there appears to be a substantial degree of either concealing or forgetting past criminal behavior. Although the majority of respondents report their offenses and the majority of all offenses are reported, there is still considerable underreporting.

Third, there is an unresolved issue of differential validity. Compared with other race-gender groups, the responses provided by African-American males appear to have lower levels of validity. Specifically, Hindelang, Hirschi, and Weis (1981) and Huizinga and Elliott (1986) report that African-American males self-report fewer of the offenses found in their official criminal histories. More recently, however, Farrington and colleagues (1996), using data from the Pittsburgh Youth Study, find no evidence of differential validity. It seems that the level of difference in the validity of self-reports for African-American males versus other groups has yet to be determined. If it is less, the processes that bring it about are frankly not understood. This is perhaps the most important methodological issue concerning the self-report method and should be a high priority for future research efforts.

Fourth, based on studies of self-reported substance use, there is some evidence that validity may be less for more serious types of offenses. In the substance use studies, the concordance between the self-report and physiological measures was strongest for adolescent tobacco use, and then for marijuana use; it was weakest for hard drugs, such as cocaine and heroin. A similar pattern is also seen for several studies of self-reported delinquency and crime (e.g., Elliott and Voss 1974; Huizinga and Elliott 1986).

What then are the psychometric properties of self-reported measures of delinquency and crime? With respect to reliability, this approach to measuring involvement in delinquency and crime appears to be acceptable. Most estimates of reliability are quite high, and there is no evidence of differential reliability. With respect to validity, the conclusion is a little murkier; we find a considerable amount of underreporting and a potential problem of differential reporting for African-American males. Nevertheless, content and construct validity

appear to be quite high, and criterion validity would be in the moderate to strong range overall. Perhaps the conclusion Hindelang, Hirschi, and Weis reached in 1981 (p. 114) is still the most reasonable:

[T]he self-report method appears to behave reasonably well when judged by standard criteria available to social scientists. By these criteria, the difficulties in self-report instruments currently in use would appear to be surmountable; the method of self-reports does not appear from these studies to be fundamentally flawed. Reliability measures are impressive and the majority of studies produce validity coefficients in the moderate to strong range.

Specialized Response Techniques

Because of the sensitive nature of asking people to report undetected criminal behavior, there has always been concern about how best to ask these questions to maximize accurate responses. Some early self-report researchers favored self-administered questionnaires, and others favored personal, face-to-face interviews. Similarly, some argued that anonymous responses were inherently better than nonanonymous responses. In their Seattle study, Hindelang, Hirschi, and Weis (1981) directly tested these concerns by randomly assigning respondents to one of four conditions: nonanonymous questionnaire, anonymous questionnaire, nonanonymous interview, and anonymous interview. Their results indicate that there is no strong method effect in producing self-report responses, and that no one approach is consistently better than the others. Similar results are reported by Krohn, Waldo, and Chiricos (1974). Some research, especially in the alcohol and drug use area, has found a methods effect. For example, Aquilino (1994) finds that admission of alcohol and drug use was lowest in telephone interviews, somewhat higher in face-to-face interviews, and highest in self-administered questionnaires (see also Aquilino and LoSciuto 1990; Turner, Lessler, and Devore 1992). Although evident, the effect size typically is not very great.

Although basic method effects do not appear to be very strong, there is still concern that in all of these approaches to the collection of survey data, respondents will feel vulnerable about reporting sensitive information. Because of that, a variety of more specialized techniques have been developed to protect respondents' confidentiality, hopefully increasing the level of reporting.

Randomized response technique

The randomized response technique assumes that the basic problem with the validity of self-reporting responses is that respondents are trying to conceal

sensitive information; that is, they are unwilling to report undetected criminal behavior if others, including the researchers, might link the behavior to them. Randomized response techniques allow respondents to conceal what they really did, while at the same time providing useful data to the researchers. There are a variety of ways of accomplishing this. We can illustrate how the basic process works with a simple example of measuring the prevalence of marijuana use.

Imagine an interview setting in which there is a screen between the interviewer and respondent so that the interviewer cannot see what the respondent is doing. The interviewer asks the sensitive question: "Have you ever smoked marijuana?" The interviewer gives the following special instructions: "Before answering, please flip a coin. If the coin lands on heads, please answer "yes" regardless of whether or not you smoked marijuana. If the coin lands on tails, please tell me the truth." Thus, the interviewer cannot know whether a "yes" response is truthful or is produced by the coin landing on heads. In this way, the respondent can admit to sensitive behavior but other people, including the interviewer, cannot know whether the admission is truthful.

From the resulting data, though, we can estimate the prevalence of marijuana use. Say we receive 70 "yes" responses from a sample of 100 respondents. Fifty of those would be produced by the coin landing on heads and can simply be ignored. Of the remaining 50 respondents, however, 20 said "yes" because they have smoked marijuana, so the prevalence of marijuana use is 20 out of 50, or 40 percent.

This technique is not limited to "yes" or "no" questions or to flipping coins. Any random process can be used as long as we know the probability distribution of bogus versus truthful responses. From these data, we can estimate prevalence, variety, and frequency scores and means and variances, and we can correlate the information with other variables, just as we do with regular self-report data.

Weis and Van Alstyne (1979) tested a randomized response procedure in the Seattle study. Based on their data, they concluded that the randomized response approach is no more efficient in eliciting positive responses to sensitive items than are traditional methods of data collection. This finding is consistent with the overall conclusion in the Hindelang et al. (1981) Seattle study that the method of administration does not significantly affect the validity of self-report responses.

The other major assessment of the randomized response technique was conducted by Tracy and Fox (1981). They sampled people who had been arrested in Philadelphia and sent interviewers to their homes. The interviewers did not

know that the sample consisted only of people with official arrest records. Respondents were asked if they had been arrested and, if so, how many times. Since this information was already known from the arrest records, the validity of the self-reported responses could be assessed. (This is much like the “reverse record check” technique used in victimization surveys; see Cantor and Lynch in this volume.) The Tracy and Fox study employed two methods of data collection, a randomized response procedure and a regular self-report interview.

The results indicate that the randomized response approach does make a difference. For all respondents, there was about 10 percent less error in the randomized response technique. For respondents who had been arrested only once, the randomized response approach actually increased the level of error. But for recidivists, the randomized response technique reduced the level of error by about 74 percent.

The randomized response technique also generated random errors (errors not correlated with other important variables). The regular self-reported interview, however, generated systematic error or bias. In this approach, underreporting was higher for females, African-American females, respondents with high need for approval, lower income respondents, and those with a larger number of arrests.

Overall, it is not clear to what extent a randomized response approach generates more complete and accurate reporting. The two major studies of this topic produce different results: Weis and Van Alstyne (1979) report no effect, and Tracy and Fox (1981) report sizable and positive effects. It should be noted, however, that Tracy and Fox’s results only generalize to self-reports of being arrested, and may or may not apply to self-reports of undetected delinquent behavior. The value of the randomized response approach requires additional research, which should be conducted within the context of audio and computer-assisted interviewing, the topic to which we now turn.

Computer-assisted interviewing

Advances in both computer hardware and software have made the introduction of computers in the actual data collection process not only a possibility but, according to Tourangeau and Smith (1996, 276), “perhaps the most commonly used method of face-to-face data collection today.” The use of computers in the data collection process began in the 1970s with computer-assisted telephone surveys (Saris 1991). This technique is used by the National Crime Survey and described in Cantor and Lynch in this volume. The technology was soon adapted to the personal interview setting, either with the interviewer administering the schedule, the Computer-Assisted Personal Interview (CAPI), or with the respondent self-administering the schedule by reading the questions on the

computer screen and entering his or her responses, the Computer-Assisted Self-Administered Interview (CASI). It is also possible to have an audio version in which the questions are recorded and the respondent listens to them, rather than having them read by the interviewer or having the respondent read them. This is called an Audio Computer-Assisted Self-Administered Interview (ACASI).

The use of computerized tools is one of two trends that have transformed survey research in the United States; the other trend is the collection of increasingly sensitive information concerning illegal and embarrassing behaviors.

Tourangeau and Smith (1996) suggest that the use of computerized tools is one of two trends that have transformed survey research in the United States; the other trend is the collection of increasingly sensitive information concerning illegal and embarrassing behaviors. One reason for the use of computer-assisted data collection that is particularly relevant for this chapter, is its potential for collecting sensitive information in a manner that increases the confidentiality of responses. By not having the interviewer read the questions or be involved in the recording of answers, the respondent does not have to reveal potentially embarrassing behavior directly to another person. In addition, the responses cannot be overheard by other people (e.g., family members or teachers) who might be nearby. Of course, the same advantage could be acquired by administering a paper-and-pencil self-administered questionnaire. However, computer-assisted techniques have other potential advantages.

A key advantage of computer-assisted administration of interview schedules over questionnaires is that they allow for the incorporation of complex branching patterns (Saris 1991; Beebe et al. 1998; Wright, Aquilino, and Supple 1998; Tourangeau and Smith 1996). For example, many delinquency checklists include a series of followup questions if the respondent answers affirmatively to having committed a particular type of delinquent behavior within a specified period of time. The branching of these followup items can be quite complex; respondents who are asked to read and follow the skip patterns can easily miss important items. Computer software can program the skip patterns and increase the probability that the respondent will answer all appropriate questions. An added advantage of computer-assisted presentation is that the respondent does not see the implications of answering in the affirmative to questions with multiple followups. Respondents may be reluctant to indicate that they have committed a delinquent act if they realize that an affirmative answer will trigger a series of followup questions (Thornberry 1989).

Computer software can also identify inconsistent and incomplete responses. Thus, if a respondent indicates that he has never been arrested but also indicates that he has spent time in a juvenile correctional facility, the program can identify this inconsistency and either prompt the respondent to clarify the issue or prompt the interviewer to ask for clarification. Computer-assisted administration can also decrease incomplete responses and reduce the number of “out of range” responses (Wright, Aquilino, and Supple 1998).

An audiotape on which questions are read to the respondent (ACASI) has two additional advantages. First, it circumvents the potential problem of illiteracy; the respondent does not have to read the questions. Second, in situations where other people might be nearby, the questions and responses are not heard by anyone but the respondent. Hence, the respondent can be more assured that answers to sensitive questions will remain private.

Although computer-assisted administration of sensitive questions provides some obvious advantages in terms of efficiency of presentation and data collection, the key question concerns the difference in the responses elicited when such technology is used. Tourangeau and Smith (1996) reviewed 18 studies that have compared different modes of data collection. The types of behavior examined include health problems (e.g., gastrointestinal problems), sexual practices, abortion, and alcohol and drug use. Tourangeau and Smith indicate that self-administered techniques generally elicit higher rates of problematic behaviors than those administered by an interviewer. Moreover, computer-assisted self-administered interviews elicit higher rates than either self-administered questionnaires or paper-and-pencil interviews administered by an interviewer. Also, ACASI (audio computer-assisted self-administered interviews) elicit higher rates than CASI.

In their own research, Tourangeau and Smith (1996) compared different modes of administration with a sample of adults ages 18 to 45. Respondents were asked questions regarding their sexual behavior as well as their use of alcohol and drugs. Data were collected using CAPI, CASI, and ACASI. With CAPI, the questions appeared on the computer screen and were read by the interviewer, who then entered the responses. With CASI, the respondent entered the responses. With ACASI, the questions appeared on the screen while a digitized recording was provided to the respondent via earphones. They found that ACASI and CASI elicited higher rates of drug use and sexual behavior than CAPI. For example, respondents who were administered CAPI reported a lifetime prevalence rate for drug use of 44.8 percent, compared with 58 percent under CASI and 66.3 percent under ACASI. The same trend was evident for other measures of drug use and for sexual activity, although in some cases the differences were not statistically significant. Tourangeau and Smith conclude that by allowing

respondents to interact directly with the computer, respondents are convinced of the “legitimacy and scientific value of the study” (p. 301). Other studies comparing administration modes have found that the level of reporting may be contingent on characteristics of respondents and the setting. For example, Wright, Aquilino, and Supple (1998) found that adolescents reported higher levels of alcohol and drug use in the computer mode than in the paper-and-pencil mode. However, mode effects were not evident for young adult respondents.

Estimates of prevalence rates of illegal and embarrassing behavior appear to be higher when computer-assisted techniques, particularly those involving self-administration, are used. The higher prevalence rates need to be externally validated. The added benefits of providing for schedule complexity and consistency in responses make these techniques attractive, and it is clear that they will continue to be used with increasing frequency.

Criminological research has increasingly come to rely on longitudinal panel designs using self-report measures of antisocial behavior to understand the dynamics of offending careers.

Self-Report Measures Across the Lifecourse

One of the most significant developments in criminology over the past 15 years has been the emergence of a “lifecourse” or developmental focus (Farrington, Ohlin, and Wilson 1986; Thornberry and Krohn forthcoming; Thornberry 1997; Jessor 1998; Weitekamp 1989). Theoretical work has expanded from a narrow focus on the adolescent years to encompass the entire criminal career of individuals. This can extend from precursors of delinquency manifested in early childhood (Moffitt 1997; Tremblay et al. 1998), through the high-delinquency years of middle and late adolescence, on into adulthood when most, but not all, offenders decrease their participation in illegal behavior (Moffitt 1997; Thornberry and Krohn forthcoming; Sampson and Laub 1990; Loeber et al. 1998). Research on “criminal careers” (Blumstein et al. 1986) has documented the importance of examining such issues as the age of onset (Krohn, Thornberry, and Rivera forthcoming) and the duration of criminal activity (Wolfgang, Thornberry, and Figlio 1987). In addition, a growing body of research has demonstrated that antisocial behavior is fairly stable from childhood to adulthood (Farrington 1989a; Huesmann et al. 1984; Olweus 1979; Moffitt 1993). Much of this work has relied primarily on the use of official data. However, criminological research has increasingly come to rely on longitudinal panel designs using self-report measures of antisocial behavior to understand the dynamics of offending careers. The use of self-report

techniques in longitudinal studies over the lifecourse introduces a number of interesting measurement issues.

Construct continuity

Although many underlying theoretical constructs, such as involvement in illegal behaviors, remain constant over time, their behavioral manifestations can change as subjects age. Failure to adapt measures to account for these changes inevitably leads to age-inappropriate measures with reduced validity and reliability. To avoid this, measures need to adapt to the respondent's developmental stage to reflect accurately the theoretical constructs of interest (Campbell 1990; Patterson 1993; Le Blanc 1989; Weitekamp 1989). In some cases, this may mean defining the concept at a level to accommodate the changing contexts in which people act at different ages. In other cases, it may mean recognizing that different behaviors at different ages imply consistency in behavioral style (Campbell 1990, 7).

In previous sections, our discussion has focused on the problems with how self-reported delinquent behavior has been defined and measured when sampling adolescents. When applying the self-report technique to both younger children and adults, these definitional issues are magnified. We recognize that different items may be needed to measure the same underlying construct to maintain the age-appropriateness of the measure. Therefore, the construct continuity of the different measures of delinquency or antisocial behavior becomes of paramount importance.

Self-report measures for children

Although antisocial behavior is quite stable, it has been likened to a chimera (Patterson 1993), with manifestations that change and accumulate with age. At very young ages (2 to 5 years), behavioral characteristics such as impulsivity, noncompliance, disobedience, and aggression are seen as early analogs of delinquent behavior. At these young ages, self-report instruments are not practical because of the age of the respondents. Rather, researchers have measured these key indicators either through parental reports or through observational ratings. Many studies of youngsters at these ages have used Achenbach's (1992) Child Behavior Checklist (CBCL), a parent-completed inventory with versions for children as young as 2 to assess "externalizing" problem behaviors.² Studies using either CBCL, some other parental or teacher report of problem behaviors, or observational ratings have demonstrated that there is a relationship between these early manifestations of problem behavior and antisocial behavior in school-age children (Belsky,

Woodworth, and Crnic 1996; Campbell 1987; Richman, Stevenson, and Graham 1982; Shaw and Bell 1993).

Starting at school age, the range of antisocial behaviors expands to include stubbornness, lying, bullying, and other externalizing problems (Loeber et al. 1993). School-age children, even those as young as first grade, begin to exhibit delinquent behaviors. However, self-report instruments of delinquent behavior have rarely been administered to preteen-age children (Loeber et al. 1989). There are some studies that have administered self-report instruments to youngsters as young as 10 or 11 years of age, slightly modifying the standard delinquency items (Elliott, Huizinga, and Ageton 1985).

Loeber and colleagues (1989) provide one of the few attempts to not only gather self-report information from children younger than the age of 10, but also examine the reliability of those reports. They surveyed a sample of 849 first grade and 868 fourth grade boys using a 33-item self-reported antisocial behavior scale. This is a younger age version of the self-reported delinquency index used by the three projects of the Program of Research on the Causes and Correlates of Delinquency. Items that were age appropriate were selected, and some behaviors were placed in several different contexts to make the content less abstract for the younger children. A special effort was made to ensure that the child understood the question by preceding each behavior with a series of questions to ascertain whether the respondents knew the meaning of the behavior. If the child did not understand the question, the interviewer gave an example and then asked the child to do the same. If the child still did not understand the question, the item was skipped.

The parents and teachers of these children were also surveyed, using a combination of the appropriate CBCL and delinquency items. To examine the validity of the child self-reported antisocial behavior scale, comparable items contained in the parent and teacher CBCL were compared with the self-report items.

Loeber et al. (1989) report that the majority of boys understood most of the items. First grade boys did have problems understanding the items regarding marijuana use and sniffing glue, and fourth grade boys had difficulty understanding the question regarding sniffing glue.

A substantial minority of the first grade boys reported damaging property (26 percent) and stealing (26 percent), while over half of the fourth grade boys reported vandalizing (51.2 percent) and stealing (53.1 percent). An even higher percentage of both first and fourth graders reported a violent offense (66.3 percent and 91.2 percent, respectively), but these items included hitting siblings and other students.

Loeber and colleagues (1989) recognized the difficulty of assessing the accuracy of self-reported delinquent behavior among elementary school children, who are unlikely to have court or police records. As an initial step, Loeber and colleagues compared the children's self-reports with parental reports about similar behaviors. They found surprisingly high concordance between children's and parents' reports about the ever-prevalence of delinquent behavior. This is especially true for behaviors that are likely to come to the attention of parents, such as aggressive behaviors and school suspension. Concordance was higher for first graders than it was for fourth graders, which Loeber and colleagues suggest would be expected, since parents are more likely to know about misbehavior at younger ages.

These findings are encouraging and suggest that self-report instruments, if administered with concern for the respondents' age, can be used for very young children. Loeber and colleagues (1989) suggest that another measure of the utility of these measures will be their predictive validity. If self-reports of delinquent behavior in the first or fourth grades predict later delinquency, there is further reason to be confident in this methodology's applicability for elementary school samples.

Self-report measures for adults

The interest in assessing antisocial behavior across the lifespan has also led to an increasing number of longitudinal surveys that have followed respondents from their adolescent years into early adulthood (e.g., Elliott 1994; Huizinga et al. 1998; Loeber et al. 1998; Farrington 1989b; Le Blanc 1989; Hawkins, Catalano, and Miller 1992; Krohn, Lizotte, and Perez 1997). The concern in constructing self-report instruments for adults is to include items that take into account the different contexts in which crime occurs at these ages (e.g., work instead of school), the opportunities for different types of offenses (e.g., domestic violence, fraud), the inappropriateness or inapplicability of offenses that appear on adolescent self-report instruments (e.g., status offenses), and the potential for very serious criminal behaviors, at least among small subset of chronic violent offenders.

Weitekamp (1989) has criticized self-report studies not only for being predominantly concerned with the adolescent years but, when covering the adult years, for also using the same items as for juveniles. He argues that even such studies as NYS (Elliott 1994) do not include many items that are more serious, and therefore appropriate for adults, than the items included in the original Short and Nye study (1957). Weitekamp asserts that we need to use different instruments during different life stages. Doing so, however, raises questions about

construct continuity similar to those discussed in constructing self-report inventories with very young children. If the researcher wants to document the change in the propensity to engage in antisocial behavior throughout the lifecourse, he or she must assume that the different items used to measure antisocial behavior at different ages do indeed measure the same underlying construct. Le Blanc (1989) suggests that a strategy of including different but overlapping items on instruments covering different ages across the lifespan is the best compromise.

The use of self-report studies in longitudinal research has generated a number of issues regarding the definition and measurement of antisocial behavior. If the researcher wants to examine the development of antisocial behavior across the lifespan, a definition and measurement of delinquent behavior limited to the standard used in research on adolescents will not suffice. Expanding that definition to encompass behaviors that take into account antisocial acts by very young children and more serious offenses by adults that may take place in different social contexts requires a well-considered definition of the construct that these different behaviors represent. Indeed, ultimately the resolution of this issue relies on a strong theoretical foundation that provides a clear definition of antisocial behavior. The utility of such a definition and the measurements that derive from it will be assessed in examining the correlations across different stages in the lifespan.

Panel or testing effects

Developments in self-report methods have improved the quality of data collected and have expanded the data's applicability to the study of antisocial behavior throughout the lifecourse. Although these advances are significant, they have increased the potential for the data to be contaminated by testing or panel effects (Thornberry 1989).

Testing effects are any alterations of the respondent's response to an item or scale that are caused by the prior administration of the same item or scale (Thornberry 1989, 351). With the use of self-reports in longitudinal research, respondents are administered the same or similar items across waves of data collection. Improvements in self-report instruments have led to the inclusion of a longer list of items to tap more serious offenses, and often, a number of followup questions are asked. The more acts a respondent admits to, the longer the overall interview will take. The concern is that this approach will make respondents increasingly unwilling to admit to delinquent acts because those responses will lengthen the interview. This effect would probably be unequally distributed because respondents with the most extensive involvement in delinquency would lose the most time by answering affirmatively to the delinquency items. Over successive administrations of the self-report instrument, respondents

would learn that positive responses lengthen their interview time, and the amount of underestimation of delinquency rates would increase.

It is also possible that the simple fact that a respondent is reinterviewed may create a generalized fatigue, decreasing the respondent's willingness to respond to self-report items. Research using the National Crime Survey of victimization found that the reduction in reporting was due more to the number of prior interviews than to the number of victimizations reported in prior interviews (Lehnen and Reiss 1978).

Three studies have examined testing effects in the use of self-report studies; all are based on data from NYS (Elliott, Huizinga, and Ageton 1985). They were conducted by Thornberry (1989), Menard and Elliott (1993), and Lauritsen (1998). NYS surveyed a nationally representative sample of 1,725 11- to 17-year-old youths in 1976. They reinterviewed the same subjects annually through 1981. These data allow researchers to examine the age-specific prevalence rates by the number of times a respondent was interviewed. For example, some respondents were 14 at the time of their first interview, some were 14 at their second interview (the original 13-year-old cohort), and some were 14 at their third interview (the original 12-year-old cohort). Because of this, a 14-year-old prevalence rate can be calculated from data collected when respondents were interviewed for only the first time, from data collected when they were interviewed a second time, and so on. If a testing or panel effect plays a role in response rates, the more frequently respondents are interviewed, the lower the age-specific rates should be. Thus the 14-year-old rate from a second interview would be lower than the 14-year-old rate based on a first interview.

Thornberry analyzed these rates for 17 NYS self-report items representing the major domains of delinquency and the most frequently occurring items. The overall trend seemed to suggest that either a panel or testing effect was occurring. For all offenses except marijuana use, comparisons between adjacent waves indicated that the age-specific prevalence rates decreased more often than they increased. For example, comparing the rate of gang fights from wave to wave, Thornberry found that for 67 percent of the comparisons, there was a decrease in the age-specific prevalence rates, whereas there was an increase in only 20 percent of the comparisons, and there was no change in 13 percent. The magnitude of the changes were, in many cases, substantial. For example, for stealing something worth \$5 to \$50, the rate drops by 50 percent for 15-year-olds from wave 1 to wave 4 (Thornberry 1989, 361).

NYS did not introduce the detailed followup questions to the delinquency items until the fourth wave of data collection. The data analyzed by Thornberry (1989) show the decline in reporting occurred across all waves. Hence, it appears

that the panel design itself, rather than the design of the specific questions, had the effect of decreasing prevalence rates. Thornberry suggests that panel and testing effects could be a serious threat to the use of self-reports in longitudinal research and calls for a more thorough investigation of this issue. The observed decline in the age-specific rates could be due to an underlying secular drop in offending during these years. Cross-sectional trend data from the Monitoring the Future (MTF) study, which cannot be influenced by a testing effect, do not indicate any such secular decline (see Thornberry 1989).

Menard and Elliott (1993) reexamined this issue using both the NYS and MTF data. They rightfully point out that comparisons across these studies need to be undertaken cautiously because of differences in samples, design features, item wording, and similar concerns. Menard and Elliott's analysis also shows that at the item level, declining trends are more evident in the NYS than in the MTF data (1993, 439). They go on to show that most of these year-to-year changes are not statistically significant, however. They then use a modified Cox-Stuart trend test to examine short-term trends in delinquency and drug use. Overall, the trends for 81 percent of the NYS offenses are not statistically significant, and about half of the MTF trends are. But, an examination of the trends for the 16 items included in their table 2 indicates that there are more declining trends in the NYS data, 9 of 16 for the 1976–80 comparisons and 7 of 16 for the 1976–83 comparisons, than there are for the MTF data, 3 of 16 in both cases. Menard and Elliott focus on the statistically significant effects that indicate fewer declining trends in NYS than is evident when one focuses on all trends, regardless of the magnitude of the change.

More recently, Lauritsen (1998) examined this topic using data from the first five waves of NYS. Specifically, she used hierarchical linear models (HLM) to estimate growth curve models for general delinquency and for serious delinquency. HLM models make fuller use of the data and include tests for statistical significance. She limited her analysis to four of the seven cohorts in NYS, those who were ages 11, 13, 15, and 17 at wave 1.

For those who were age 13, 15, or 17 at the start of NYS, involvement in both general delinquency and serious delinquency decreased significantly over the next 4 years. For the 11-year-old cohort, the rate of change was also negative but not statistically significant. This downward trajectory in the rate of delinquent behavior for all age cohorts is not consistent with theoretical expectations or what is generally known about the age-crime curve. Also, as Lauritsen points out, it is not consistent with other data on secular trends for the same time period (see also Thornberry 1989; Osgood et al. 1989).

Finally, Lauritsen examined whether this testing effect is due to the introduction of detailed followup questions at wave 4 of the NYS or whether it appeared to be produced by general panel fatigue. Her analysis of individual growth trajectories indicates that the decline is observed across all waves. Thus, she concludes, as Thornberry did, that the reduced reporting is unlikely to have been produced by a change in survey administration, namely, by the addition of followup questions.

Overall, Lauritsen offers two explanations for the observed testing effects. One concerns generalized panel fatigue, suggesting that as respondents are asked the same inventory at repeated surveys, they become less willing to respond affirmatively to screening questions. The second explanation concerns a maturation effect in which the content validity of the self-report questions would vary with age. For example, how respondents interpret a question on simple assault, and the type of behavior they consider relevant for responding to the question, may be quite different for 11- and 17-year-olds. Of course, both of these processes may operate.

The studies by Thornberry and by Lauritsen suggest that there is some degree of panel bias in self-report data collected in longitudinal panel studies. The analysis by Menard and Elliott indicates that this is still just a suggestion, as the necessary comparisons between panel studies and cross-sectional trend studies are severely hampered by lack of comparability in item wording, administration, and other methodological differences. Also, if there are testing effects, neither Thornberry nor Lauritsen is arguing that they are unique to NYS. It just so happens that the sequential cohort design of NYS makes it a good vehicle for examining this issue. The presumption, unfortunately, is that if testing effects interfere with the validity of the NYS data, they also interfere with the validity of other longitudinal data containing self-report information. This is obviously a serious matter, as etiological research has focused almost exclusively on longitudinal designs in the past 20 years. Additional research to identify the extensiveness of testing effects, their sources, and way of remedying them are certainly a high priority.

Conclusions

The self-report method for measuring crime and delinquency has developed substantially since its introduction a half century ago. It is now a fundamental method of scientifically measuring criminality and forms the bedrock of etiological studies. The challenges confronting this approach to measurement are daunting; after all, we are asking individuals to tell us about their own, undetected criminality. Despite this fundamental challenge, the technique seems to be successful and capable of producing valid and reliable data.

Early self-report scales had substantial weaknesses, containing few items and producing an assessment of only minor forms of offending. Gradually, as the underlying validity of the approach became evident, the scales expanded in terms of breadth, seriousness, and comprehensiveness. Contemporary measures typically cover a wide portion of the behavioral domain included under the construct of crime and delinquency. These scales are able to measure serious as well as minor forms of crime, with such major subdomains as violence, property crimes, and drug use; to measure different parameters of criminal careers such as prevalence, frequency, and seriousness; and to identify high-rate as well as low-rate offenders. This is substantial progress for a measurement approach that began with a half dozen items and a four-category response set.

The self-report approach to measuring crime has acceptable, albeit far from perfect, reliability and validity. Of these two basic psychometric properties, the evidence for reliability seems stronger. There are no fundamental challenges to the reliability of these data. Test-retest measures (and internal consistency measures) indicate that self-reported measures of delinquency are as reliable as, if not more reliable than, most social science measures.

Validity is much harder to assess, as there is no “gold standard” against which to judge the self-reports. Nevertheless, current scales seem to have acceptable levels of content and construct validity. The evidence for criterion validity is less clear cut. At an overall level, criterion validity seems to be in the moderate to strong range. Although there is certainly room for improvement, the validity appears acceptable for most analytic tasks. At a more specific level, however, there is a potentially serious problem with differential validity. Two of the major assessments of criterion validity, by Hindelang, Hirschi, and Weis (1981) and by Huizinga and Elliott (1986), found lower validity for African-American males. The more recent assessment by Farrington and colleagues (1996) did not. Additional research on this topic is imperative.

Although basic self-report surveys appear to be reliable and valid, researchers have experimented with a variety of data collection methods to improve the quality of reporting. Several of these attempts have produced ambiguous results; for example, there is no clear-cut benefit to mode of administration (interview versus questionnaire) or to the use of randomized response techniques. There is one approach that appears to hold great promise, however. Audio-assisted computerized interviews produce increased reporting of many sensitive topics, including delinquency and drug use. Greater use of this approach is warranted.

In the end, the available data indicate that the self-report method is an important and useful way to collect information about criminal behavior. The

skepticism of early critics like Nettler (1978) and Gibbons (1979) has not been realized. Nevertheless, the self-report technique can clearly be improved. The final issue addressed in this chapter is suggestions for future research.

Future directions

Much of our research on reliability and validity simply assesses these characteristics; there is far less research on improving their levels. For example, it is likely that both validity and reliability would be improved if we experimented with alternate items for measuring the same behavior and identified the strongest ones. Similarly, reliability and validity vary across subscales (e.g., Huizinga and Elliott 1986); improving subscales will not only help them but also the overall scale as they are aggregated.

Throughout this chapter, we discussed the issue of differential validity for African-American males. It is crucial to learn more about the magnitude of this bias and its source, if it exists. Future research should address this issue directly and attempt to identify techniques for eliminating it. These research efforts should not lose sight of the fact that the problem may be with the criterion variable (official records) and not the self-reports.

The self-report method was developed in and for cross-sectional studies. Using it in longitudinal studies, especially ones that cover major portions of the life-course, creates a new set of challenges. Maintaining the age-appropriateness of the items, while at the same time ensuring content validity, is a knotty problem that we have just begun to address. There is some evidence that repeated measures may create testing effects. More research is needed to measure the size of this effect and to identify methods to reduce its threat to the validity of self-report data in the longitudinal studies that are so crucial to etiological investigation.

One of the most promising developments in the self-report method is the advent of audio-assisted computerized interviews. This technique offers increased confidentiality to the respondent in an interview setting. Although somewhat expensive and complicated to design, the early studies indicate that it may be worth the effort.

Finally, we recommend that methodological studies be done in a crosscutting fashion so that several of these issues—reliability and validity, improved item selection, assessing panel bias, etc.—can be addressed simultaneously. It is particularly important to examine all of these methodological issues when data are collected using audio-assisted computerized interviewing. For example, studies that have found differential validity or testing effects have all used paper-and-pencil interviews. Whether these same problems are evident under the enhanced confidentiality of audio interviews is an open question. It is clearly a high-priority one as well.

There is no dearth of work that can be done to assess and improve the self-report method. If the progress over the past half century is any guide, we are optimistic that the necessary studies will be conducted and that they will improve this basic means of collecting data on criminal behavior.

Notes

1. This is particularly the case given the level of reliability of self-reported data (see the section "Assessing reliability"). By adding random error to the picture, poor reliability attenuates or reduces the size of the observed correlation coefficients.
2. CBCL also assesses internalizing problem behavior.

References

- Achenbach, T.M. 1992. *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington: University of Vermont.
- Akers, Robert L. 1964. Socio-economic status and delinquent behavior: A retest. *Journal of Research in Crime and Delinquency* 1:38-46.
- Akers, Robert L., Marvin D. Krohn, L. Lanza-Kaduce, and M. Radosevich. 1979. Social learning and deviant behavior: A specific test of a general theory. *American Sociological Review* 44:636-655.
- Akers, Robert L., James Massey, William Clarke, and Ronald M. Lauer. 1983. Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Social Forces* 62 (September): 234-251.
- Anderson, Linda S., Theodore G. Chiricos, and Gordon P. Waldo. 1977. Formal and informal sanctions: A comparison of deterrent effects. *Social Problems* 25:103-112.
- Aquilino, William S. 1994. Interview mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly* 58:210-240.
- Aquilino, William S., and Leonard LoSciuto. 1990. Effects of interview mode on self-reported drug use. *Public Opinion Quarterly* 54:362-395.
- Beebe, Timothy J., Patricia A. Harrison, James A. McRae, Jr., Ronald E. Anderson, and Jayne A. Fulkerson. 1998. An evaluation of computer-assisted self-interviews in a school setting. *Public Opinion Quarterly* 62:623-632.
- Belsky, J., S. Woodworth, and K. Crnic. 1996. Troubled family interaction during toddlerhood. *Development and Psychopathology* 8:477-495.
- Belson, W.A. 1968. The extent of stealing by London boys and some of its origins. *Advancement of Science* 25:171-184.

Blackmore, J. 1974. The relationship between self-reported delinquency and official convictions amongst adolescent boys. *British Journal of Criminology* 14:172–176.

Blumstein, A., J. Cohen, J.A. Roth, and C.A. Visher, eds. 1986. *Criminal careers and "career criminals."* Washington, D.C.: National Academy Press.

Braukmann, C.J., K.A. Kirigin, and M.M. Wolf. 1979. Social learning and social control perspectives in group home delinquency treatment research. Paper presented at the 1979 Annual Meeting of the American Society of Criminology, November, Philadelphia.

Broder, P.K., and J. Zimmerman. 1978. *Establishing the reliability of self-reported delinquency data.* Williamsburg, Virginia: National Center for State Courts.

Browning, Katharine, David Huizinga, Rolf Loeber, and Terence P. Thornberry. 1999. *Causes and correlates of delinquency program.* Fact Sheet, FS 99100. Washington, D.C.: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Campbell, S.B. 1990. *Behavioral problems in preschool children: Clinical and developmental issues.* New York: Guilford Publications.

———. 1987. Parent-referred problem three-year-olds: Developmental changes in symptoms. *Journal of Child Psychology & Psychiatry* 28:835–845.

Clark, J.P., and L.L. Tiftt. 1966. Polygraph and interview validation of self-reported delinquent behavior. *American Sociological Review* 31:516–523.

Clark, J.P., and E.P. Wenninger. 1962. Socioeconomic class and area as correlates of illegal behavior among juveniles. *American Sociological Review* 28:826–834.

Conger, Rand. 1976. Social control and social learning models of delinquency: A synthesis. *Criminology* 14:17–40.

Dentler, R.A., and L.J. Monroe. 1961. Social correlates of early adolescent theft. *American Sociological Review* 26:733–743.

Elliott, Delbert S. 1994. Serious violent offenders: Onset, developmental course, and termination—The American Society of Criminology 1993 presidential address. *Criminology* 32 (1): 1–21.

———. 1966. Delinquency, school attendance, and dropout. *Social Problems* 13:306–318.

Elliott, Delbert S., and S.S. Ageton. 1980. Reconciling race and class differences in self-reported and official estimates of delinquency. *American Sociological Review* 45 (1): 95–110.

Elliott, Delbert S., D. Huizinga, and S.S. Ageton. 1985. *Explaining delinquency and drug use.* Beverly Hills: Sage Publications.

- Elliott, Delbert S., and H.L. Voss. 1974. *Delinquency and dropout*. Lexington, Massachusetts: D.C. Heath.
- Empey, LaMar T., and Maynard Erickson. 1966. Hidden delinquency and social status. *Social Forces* 44 (June): 546–554.
- Erickson, M., and L.T. Empey. 1963. Court records, undetected delinquency, and decision-making. *Journal of Criminal Law, Criminology, and Police Science* 54 (December): 456–469.
- Farrington, David P. 1989a. Early predictors of adolescent aggression and adult violence. *Violence and Victims* 4:79–100.
- . 1989b. Self-reported and official offending from adolescence to adulthood. In *Cross-national research in self-reported crime and delinquency*, edited by Malcolm W. Klein. Los Angeles: Kluwer Academic Publishers.
- . 1977. The effects of public labelling. *British Journal of Criminology* 17:112–125.
- . 1973. Self-reports of deviant behavior: Predictive and stable? *Journal of Criminal Law and Criminology* 64:99–110.
- Farrington, David P., Rolf Loeber, Magda Stouthamer-Loeber, Welmoet B. Van Kammen, and Laura Schmidt. 1996. Self-reported delinquency and a combined delinquency seriousness scale based on boys, mothers, and teachers: Concurrent and predictive validity for African-Americans and Caucasians. *Criminology* 34 (November): 493–517.
- Farrington, David P., Lloyd E. Ohlin, and James Q. Wilson. 1986. *Understanding and controlling crime: Toward a new research strategy*. New York: Springer-Verlag.
- Gibbons, Don C. 1979. *The criminological enterprise: Theories and perspectives*. Upper Saddle River, New Jersey: Prentice Hall.
- Gold, M. 1970. *Delinquent behavior in an American city*. Belmont, California: Brooks/Cole Publishing Company.
- . 1966. Undetected delinquent behavior. *Journal of Research in Crime and Delinquency* 3:27–46.
- Hardt, R.H., and S. Petersen-Hardt. 1977. On determining the quality of the delinquency self-report method. *Journal of Research in Crime and Delinquency* 14:247–261.
- Hathaway, R.S., E.D. Monachesi, and L.A. Young. 1960. Delinquency rates and personality. *Journal of Criminal Law, Criminology, and Police Science* 50:433–440.

Hawkins, J. David, Richard F. Catalano, and Janet Y. Miller. 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): 64–105.

Hepburn, J.R. 1976. Testing alternative models of delinquency causation. *Journal of Criminal Law and Criminology* 67:450–460.

Hindelang, M.J. 1973. Causes of delinquency: A partial replication and extension. *Social Problems* 20:471–487.

Hindelang, M.J., T. Hirschi, and J.G. Weis. 1981. *Measuring delinquency*. Beverly Hills: Sage Publications.

———. 1979. Correlates of delinquency: The illusion of discrepancy between self-report and official measures. *American Sociological Review* 44:995–1014.

Hirschi, T. 1969. *Causes of delinquency*. Berkeley: University of California Press.

Huesmann, L.R., L.D. Eron, M.M. Lefkowitz, and L.O. Walder. 1984. The stability of aggression over time and generations. *Developmental Psychology* 20:1120–1134.

Huizinga, D., and D.S. Elliott. 1986. Reassessing the reliability and validity of self-report delinquent measures. *Journal of Quantitative Criminology* 2 (4): 293–327.

———. 1983. *A preliminary examination of the reliability and validity of the National Youth Survey self-reported delinquency indices*. National Youth Survey Project Report 27. Boulder, Colorado: Behavioral Research Institute.

Huizinga, David, Anne W. Weiher, Scott Menard, Rachele Espiritu, and Finn Esbensen. 1998. Some not-so-boring findings from the Denver Youth Survey. Paper presented at the 1998 Annual Meeting of the American Society of Criminology, November, Washington, D.C.

Jensen, G.F. 1973. Inner containment and delinquency. *Journal of Criminal Law and Criminology* 64:464–470.

Jensen, Gary F., Maynard L. Erickson, and Jack P. Gibbs. 1978. Perceived risk of punishment and self-reported delinquency. *Social Forces* 57:57–78.

Jensen, Gary F., and Raymond Eve. 1976. Sex differences in delinquency. *Criminology* 13:427–448.

Jessor, Richard. 1998. *New perspectives on adolescent risk behavior*. New York: Cambridge University Press.

Johnson, R.E. 1979. *Juvenile delinquency and its origins*. Cambridge: Cambridge University Press.

- Johnston, Lloyd D., Patrick M. O'Malley, and Jerald G. Bachman. 1996. *National survey results on drug use from the Monitoring the Future study, 1975–1995*. Rockville, Maryland: U.S. Department of Health and Human Services.
- Kaplan, H.B. 1972. Toward a general theory of psychosocial deviance: The case of aggressive behavior. *Social Science and Medicine* 6:593–617.
- Kelly, D.H. 1974. Track position and delinquent involvement: A preliminary analysis. *Sociology and Social Research* 58:380–386.
- Klein, Malcolm W., ed. 1989. *Cross-national research in self-reported crime and delinquency*. Los Angeles: Kluwer Academic Publishers.
- Krohn, Marvin D., Alan J. Lizotte, and Cynthia M. Perez. 1997. The interrelationship between substance use and precocious transitions to adult statuses. *Journal of Health and Social Behavior* 38 (March): 87–103.
- Krohn, Marvin D., Terence P. Thornberry, and Craig Rivera. Forthcoming. Later careers of very young offenders. In *Child delinquents: Development, interventions, and service needs*, edited by Rolf Loeber and David P. Farrington. Thousand Oaks, California: Sage Publications.
- Krohn, Marvin D., Gordon P. Waldo, and Theodore G. Chiricos. 1974. Self-reported delinquency: A comparison of structured interviews and self-administered checklists. *Journal of Criminal Law and Criminology* 65 (4): 545–553.
- Kulik, J.A., K.B. Stein, and T.R. Sarbin. 1968. Disclosure of delinquent behavior under conditions of anonymity and non-anonymity. *Journal of Consulting and Clinical Psychology* 32:506–509.
- Lauritsen, Janet L. 1998. The age-crime debate: Assessing the limits of longitudinal self-report data. *Social Forces* 76 (4) (June): 1–29.
- Le Blanc, Marc. 1989. Designing a self-report instrument for the study of the development of offending from childhood to adulthood: Issues and problems. In *Cross-national research in self-reported crime and delinquency*, edited by Malcolm W. Klein. Los Angeles: Kluwer Academic Publishers.
- Lehnen, R.G., and A.J. Reiss. 1978. Response effects in the National Crime Survey. *Victimology: An International Journal* 3:110–124.
- Loeber, Rolf, David P. Farrington, Magda Stouthamer-Loeber, Terrie E. Moffitt, and Avshalom Caspi. 1998. The development of male offending: Key findings from the first decade of the Pittsburgh Youth Study. *Studies on Crime and Crime Prevention* 7 (2): 1–31.

- Loeber, Rolf, Magda Stouthamer-Loeber, Welmoet B. Van Kammen, and David P. Farrington. 1989. Development of a new measure of self-reported antisocial behavior for young children: Prevalence and reliability. In *Cross-national research in self-reported crime and delinquency*, edited by Malcolm W. Klein. Los Angeles: Kluwer Academic Publishers.
- Loeber, R., P. Wung, K. Keenan, B. Giroux, and M. Stouthamer-Loeber. 1993. Developmental pathways in disruptive child behavior. *Development and Psychopathology* 5:101–133.
- Matthews, Victor M. 1968. Differential identification: An empirical note. *Social Problems* 14:376–383.
- Menard, Scott, and Delbert S. Elliott. 1993. Data set comparability and short-term trends in crime and delinquency. *Journal of Criminal Justice* 21 (5): 433–445.
- Merton, R.K. 1938. Social structure and anomie. *American Sociological Review* 3:672–682.
- Moffitt, Terrie E. 1997. Adolescence-limited and life-course-persistent offending: A complementary pair of developmental theories. In *Developmental theories of crime and delinquency*, edited by Terence P. Thornberry. Vol. 7 of *Advances in criminological theory*. New Brunswick, New Jersey: Transaction Publishers.
- . 1993. Life-course-persistent and adolescence-limited antisocial behavior: A developmental taxonomy. *Psychological Review* 100:674–701.
- Murphy, F.J., M.M. Shirley, and H.L. Witmer. 1946. The incidence of hidden delinquency. *American Journal of Orthopsychiatry* 16 (October): 686–696.
- Nettler, G. 1978. *Explaining crime*. New York: McGraw-Hill.
- Nye, F.I., and James F. Short, Jr. 1957. Scaling delinquent behavior. *American Sociological Review* 22 (June): 326–331.
- Nye, F. Ivan, James F. Short, Jr., and V. Olson. 1958. Socioeconomic status and delinquent behavior. *American Journal of Sociology* 63:381–389.
- Olweus, D. 1979. Stability and aggressive reaction patterns in males: A review. *Psychological Bulletin* 86:852–875.
- Osgood, D. Wayne, Patrick O'Malley, Jerald Bachman, and Lloyd Johnston. 1989. Time trends and age trends in arrests and self-reported illegal behavior. *Criminology* 27:389–418.
- Park, Robert E.K., Ernest W. Burgess, and Roderick D. McKenzie. 1928. *The city*. Chicago: University of Chicago Press.

- Patterson, G.R. 1993. Orderly change in a stable world: The antisocial trait as a chimera. *Journal of Consulting and Clinical Psychology* 61:911–919.
- Patterson, G.R., and R. Loeber. 1982. The understanding and prediction of delinquent child behavior. Research proposal to the U.S. Department of Health and Human Services, National Institute of Mental Health. Eugene: Oregon Social Learning Center.
- Polk, K. 1969. Class strain and rebellion among adolescents. *Social Problems* 17:214–224.
- Porterfield, Austin L. 1946. *Youth in trouble*. Fort Worth: Leo Potishman Foundation.
- . 1943. Delinquency and outcome in court and college. *American Journal of Sociology* 49 (November): 199–208.
- Reiss, Albert J., Jr., and A.L. Rhodes. 1964. An empirical test of differential association theory. *Journal of Research in Crime and Delinquency* 1:5–18.
- . 1963. Status deprivation and delinquent behavior. *Sociological Quarterly* 4:135–149.
- Richman, N., J. Stevenson, and P.J. Graham. 1982. *Preschool to school: A behavioural study*. London: Academic Press.
- Robison, Sophia Moses. 1936. *Can delinquency be measured?* New York: Columbia University Press.
- Rojek, D.G. 1983. Social status and delinquency: Do self-reports and official reports match? In *Measurement issues in criminal justice*, edited by Gordon P. Waldo. Beverly Hills: Sage Publications.
- Sampson, R.J., and J.H. Laub. 1990. Crime and deviance over the life course: The salience of adult social bonds. *American Sociological Review* 55 (October): 609–627.
- Saris, Willem E. 1991. *Computer-assisted interviewing*. Newbury Park, California: Sage Publications.
- Sellin, Thorsten. 1931. The basis of a crime index. *Journal of Criminal Law and Criminology* 22:335–356.
- Shaw, Clifford, and Henry D. McKay. 1942. *Juvenile delinquency and urban areas*. Chicago: University of Chicago Press.
- Shaw, D.S., and R.Q. Bell. 1993. Developmental theories of parental contributors to antisocial behavior. *Journal of Abnormal Child Psychology* 21:35–49.
- Short, James F. 1957. Differential association and delinquency. *Social Problems* 4:233–239.

Short, J.F., Jr., and F.I. Nye. 1958. Extent of unrecorded juvenile delinquency: Tentative conclusions. *Journal of Criminal Law and Criminology* 49:296–302.

———. 1957. Reported behavior as a criterion of deviant behavior. *Social Problems* 5:207–213.

Silberman, M. 1976. Toward a theory of criminal deterrence. *American Sociological Review* 41:442–461.

Skolnick, J.V., C.J. Braukmann, M.M. Bedlington, K.A. Kirigin, and M.M. Wolf. 1981. Parent-youth interaction and delinquency in group homes. *Journal of Abnormal Child Psychology* 9:107–119.

Slocum, W.L., and C.L. Stone. 1963. Family culture patterns and delinquent-type behavior. *Marriage and Family Living* 25:202–208.

Smith, Carolyn, and Terence P. Thornberry. 1995. The relationship between childhood maltreatment and adolescent involvement in delinquency. *Criminology* 33:451–481.

Stanfield, R. 1966. The interaction of family variables and gang variables in the aetiology of delinquency. *Social Problems* 13:411–417.

Sutherland, Edwin H. 1949. *White-collar crime*. New York: Holt, Rinehart and Winston.

———. 1940. White collar criminality. *American Sociological Review* 5:1–12.

———. 1939. *Principles of criminology*. 3d ed. Philadelphia: J.B. Lippincott.

Thornberry, Terence P. 1997. *Developmental theories of crime and delinquency*. New Brunswick, New Jersey: Transaction Publishers.

———. 1989. Panel effects and the use of self-reported measures of delinquency in longitudinal studies. In *Cross-national research in self-reported crime and delinquency*, edited by Malcolm W. Klein. Los Angeles: Kluwer Academic Publishers.

———. 1987. Toward an interactional theory of delinquency. *Criminology* 25:863–891.

Thornberry, Terence P., and Marvin D. Krohn. Forthcoming. The development of delinquency: An interactional perspective. In *Handbook of law and social science: Youth and justice*, edited by Susan O. White. New York: Plenum Press.

Thrasher, F. 1927. *The gang: A study of 1,313 gangs in Chicago*. Chicago: University of Chicago Press.

Tourangeau, Roger, and Tom W. Smith. 1996. Asking sensitive questions: The impact of data collection, mode, question format, and question context. *Public Opinion Quarterly* 60:275–304.

- Tracy, Paul E., and James A. Fox. 1981. The validity of randomized response for sensitive measurements. *American Sociological Review* 46 (April): 187–200.
- Tremblay, Richard E., Christa Japel, Daniel Perusse, Pierre McDuff, Michel Boivin, Mark Zoccolillo, and Jacques Montplaisir. 1999. The search for the age of “onset” of physical aggression: Rousseau and Bandura revisited. *Criminal Behaviour and Mental Health* 9:8–23.
- Turner, Charles F., Judith T. Lessler, and James Devore. 1992. Effects of mode of administration and wording on reporting of drug use. In *Survey measurement of drug use: Methodological studies*, edited by Charles F. Turner, Judith T. Lessler, and Joseph C. Gfroerer. Washington, D.C.: U.S. Department of Health and Human Services.
- U.S. Department of Justice, National Institute of Justice. 1990. *1988 Drug Use Forecasting annual report: Drugs and crime in America*. Research in Action, NCJ 122225. Washington, D.C.
- Vaz, E.W. 1966. Self reported juvenile delinquency and social status. *Canadian Journal of Corrections* 8:20–27.
- Voss, H.L. 1966. Socio-economic status and reported delinquent behavior. *Social Problems* 13:314–324.
- . 1964. Differential association and reported delinquent behavior: A replication. *Social Problems* 12:78–85.
- . 1963. Ethnic differentials in delinquency in Honolulu. *Journal of Criminal Law and Criminology* 54:322–327.
- Waldo, Gordon P., and Theodore G. Chiricos. 1972. Perceived penal sanction and self-reported criminality: A neglected approach to deterrence research. *Social Problems* 19:522–540.
- Wallerstein, J.S., and C.J. Wylie. 1947. Our law-abiding law-breakers. *Probation* 25:107–112.
- Weis, J.G., and D.V. Van Alstyne. 1979. The measurement of delinquency by the randomized response method. Paper presented at the 1979 Annual Meeting of the American Society of Criminology, November, Philadelphia.
- Weitekamp, Elmar. 1989. Some problems with the use of self-reports in longitudinal research. In *Cross-national research in self-reported crime and delinquency*, edited by Malcolm W. Klein. Los Angeles: Kluwer Academic Publishers.
- Williams, Jay R., and Martin Gold. 1972. From delinquent behavior to official delinquency. *Social Problems* 20 (2): 209–229.

Wolfgang, Marvin E., Robert M. Figlio, and Thorsten Sellin. 1972. *Delinquency in a birth cohort*. Chicago: University of Chicago Press.

Wolfgang, Marvin E., Terence P. Thornberry, and Robert M. Figlio. 1987. *From boy to man, from delinquency to crime*. Chicago: University of Chicago Press.

Wright, Debra L., William S. Aquilino, and Andrew J. Supple. 1998. A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly* 62:331–353.

Self-Report Surveys as Measures of Crime and Criminal Victimization

by David Cantor and James P. Lynch

Self-report surveys of victimization have become commonplace in discussions of crime and criminal justice policy. Changes in the rates at which residents of the country are victimized by crime have taken a place alongside the Federal Bureau of Investigation index of crimes known to the police as widely used indicators of the state of society and the efficacy of its governance. While a great deal has been learned about this method for producing data on crime and victimization, a number of fundamental issues concerning the method remain only partially explored. This paper outlines what we have learned about victimization surveys over the past 30 years and how this source of information has been used as a social indicator and a means of building criminological theories. It also identifies major methodological issues that remain unresolved and suggests some approaches to exploring them. The evolution of the National Crime Victimization Survey is used as a vehicle for this discussion, because the survey has been conducted continuously for 25 years and has been the subject of extensive methodological study.

A
B
S
T
R
A
C
T

85

David Cantor is an Associate Study Director with Westat in Rockville, Maryland. James P. Lynch is a Professor in the School of Public Affairs with American University in Washington, D.C.

A Review of Self-Report Surveys

Self-report measures of criminal victimization have become widely used social indicators and research tools in criminology and criminal justice. A great deal has been learned about the strengths and weakness of this methodology. Substantial improvements have been made in this methodology since its inception in the late 1960s, yet problems and limitations persist. This essay examines the evolution of the victimization survey methodology. It identifies (1) the contribution that these surveys have made to our understanding of crime and victimization, (2) what we have learned about the methodology, (3) what more we need to know, and (4) what additional research could be done to help us know it.

Assessing the self-report methodology in all its guises would require many more pages than we have been allotted.¹ Consequently, we will focus on household surveys of the general population for the purpose of continuing statistical series on the incidence and characteristics of criminal victimization. The National Crime Survey (NCS), conducted for the U.S. Department of Justice's (DOJ's) Bureau of Justice Statistics (BJS), will be at the center of our attention, along with its immediate precursors and their genesis.²

NCS has many design features that are not employed in other large-scale household surveys of victimization. Settling on many of these features involved conscious tradeoffs between data quality and the costs of administering the survey within the environment of the U.S. Census Bureau. Some of these decisions were supported by extensive testing to determine the effects of varying design features on the reporting of crime and the feasibility of fielding a survey with given features. Knowledge of this methodological work is important for understanding the current state of self-report victimization surveys. The uniqueness of NCS will be described briefly. A more detailed description of the history and evolution of the survey will follow.

Contributions of the survey method to our understanding of crime

Victim surveys substantially changed the definition of crime and the nature of the information available on crime events. Prior to the availability of victim surveys in the United States, much of our information on the volume and nature of crime came from the police, specifically the Uniform Crime Reports (UCR). Since 1929, UCR compiled statistics submitted by participating local police departments on offenses known to the police, persons arrested, and officers killed or assaulted. This information was collected on a subset of crimes that the International Association of Chiefs of Police (IACP) at the time considered

prevalent, serious, and well reported to the police. Consistent with the technology available at the time, local police departments submitted aggregate counts of offenses known for seven categories of crime: homicide, rape, robbery, aggravated assault, burglary, larceny, and motor vehicle theft. These annual counts by jurisdiction of crimes known to the police were the principal product of UCR, and they were used by many to assess the level and change in level of crime in the United States.³

The events defined as crime and the information collected on these events were shaped by the needs of police organizations. The surveys had a different set of limitations related to the survey enterprise. Police record systems available at the time included only those events reported and recorded by the police, collected data on a selected subset of crime, and data presented as aggregate counts of crimes. The victim surveys included events that were reported to the police as well as those that were not. They included extensive information on victims and the social context of the crime and made those data available on an incident or victim basis. These surveys gathered data from victims and nonvictims. All of these differences increased the utility of available data on crime as a social indicator and for research.

In this section, the major contributions that victim surveys have made to criminological theory and policy are reviewed. Given space limitation, this review takes a broad-brush approach. For more detailed reviews, at least through the mid-1980s, the reader is referred to several other excellent reviews (Gottfredson 1986; Sparks 1981).

Implications for crime as a social indicator

One of the major functions of crime statistics is to provide a social indicator. Crime statistics serve this function by providing estimates of the level and change in one aspect of the well-being of a nation, state, or locality. Victim surveys substantially improved the information available on the volume of crime. The data from victim surveys included many crimes that were not reported to the police or other criminal justice agencies (Biderman and Reiss 1967). Victim surveys also provided more detailed information on crime events than did national data systems based on police records. These surveys would ask respondents to provide information on themselves, the offenders, the nature of the crime, and the context in which it occurred. While this type of information may have been available in local police files, it was not assembled nationally by agencies like the Federal Bureau of Investigation (FBI) in a form that allowed easy access. Moreover, the detail available in police files varied substantially, depending on the willingness of police officers to ask victims systematically for the specifics of crime events. Sample surveys take much greater pains to ensure

that all respondents are asked to provide the same information on every crime. Finally, victim surveys provided event-level data, whereas UCR offered aggregate counts of eligible events in a given jurisdiction. As a result, victim survey data could be reported in various ways, while police counts could not. For example, the survey could present change estimates for lower class, black males in central cities or for females over 50, while UCR could present only a count of crimes by type nationally and for a jurisdiction.

These differences substantially improved our ability to estimate the volume and change in the volume of crime. At last, an estimate that included unreported crime could be made, and estimates could now be made for subgroups as well as for the population as a whole. This went a long way toward improving on police data as a social indicator. The volume of crime could be estimated for young males or whites or American Indians, for example, so that one could assess whether the volume of crime and the change in the volume of crime was the same for everybody. It became clear with the release of the survey data that this was not the case (U.S. Department of Justice, National Criminal Justice Information and Statistics Service 1976). This was a tremendous step forward for the use of crime data as a social indicator.

Finally, the surveys allow for the creation of new ways of classifying crimes other than the ubiquitous index crime classification, which had come to dominate and limit our understanding of the crime problem. The survey could produce estimates of "stranger to stranger" crime, crime "among intimates," "crime at work" or "vehicle-related crime" rather than staying with rape, robbery, aggravated assault, etc. These alternative crime classifications shed a whole new light on crime. Just as population-specific crime rates demonstrated that different groups of people had different crime problems, these alternative classifications showed that there were different problems as defined by the social context of the act that were not visible when events were classified by the criminal act alone.

The availability of victim surveys in tandem with data from administrative police records has raised the level of sophistication among consumers of crime information.

The benefits of the victim surveys as social indicator arose as much from the organization of the survey enterprise as from the enhancement in the information provided. Prior to the institutionalization of victim surveys, crime information was entirely under the control of the criminal justice system. This raised questions about the accuracy and scientific impartiality of the resulting data. Because the police have an immediate and specific interest in the crime problem, there is always the suspicion that they are "cooking the books." Victim surveys brought the "patina of

science” to crime statistics. The Census Bureau and survey research agencies were not interested parties with respect to the crime problem, and there was greater trust that the resulting crime estimates were not purposely manipulated.

The availability of victim surveys in tandem with data from administrative police records has raised the level of sophistication among consumers of crime information. They use both of these indicators to try to understand the crime problem and how it might be changing. Initially, the two indicators were pitted against each other as the one true indicator of crime, but gradually this is giving way to the complementary use of the two series (McDowall and Loftin 1992). BJS, for example, is issuing reports that include data on relevant topics from both UCR and NCS (Zawitz 1988). Journalists make references to both indicators in their routine crime stories, and disparities between the police and survey data are taken as issues to be explained rather than used to impeach one or the other statistic (*New York Times* 1981; *Washington Post* 1981). These are all positive signs that the consumers of crime statistics are appreciating the complexity of describing the crime problem and are treating these social indicators with appropriate caution. The depth and breadth of this sophistication is difficult to assess in a period when the two series have tracked each other for a number of years. It will become more clear when the series diverge again. Nonetheless, this movement toward greater sophistication in the production and consumption of crime statistics would not have occurred without the routine availability of victim surveys.

Finally, victim surveys have played a role in the evolution of UCR. In 1984, the FBI undertook a study of UCR for the purpose of improving the system (Poggio et al. 1985). This redesign effort may have been prompted directly by the NCS redesign that had been under way for several years at that time. The NCS redesign was uncovering a great deal about the survey, and it gave BJS the ability to deflect any criticism of the survey by pointing to the redesign as evidence that something was being done about it. UCR had no such protection unless similar efforts to improve the series were undertaken. Moreover, the redesign recommended that UCR adopt a number of the features of NCS, specifically incident-based rather than aggregate reporting. Although this recommendation was affected by the increased use of computing and management information systems in local police departments, it was also an attempt by UCR to match the flexibility of NCS in reporting crime rates.

Implications for building theories of crime and its consequences

Victim surveys have also had a profound effect on theories of crime causation. The availability of highly disaggregated information on crime events, including

events not known to the police, facilitated the development of a whole new way of looking at crime. Routine activity, opportunity theory, and even rational choice theories of crime flourished in large part because of the availability of victim survey data (Hindelang, Gottfredson, and Garofalo 1978; Gottfredson 1984; Kennedy and Forde 1999; Hough 1987; Cohen and Cantor 1981; Miethe and Meier 1994; Maxfield 1987; Lynch and Cantor 1992). The surveys also provided an opportunity to identify and investigate the consequences of crime. By documenting the durable and psychic harm that resulted from victimization, the surveys prompted researchers to investigate why the degree of harm differed across crimes and victims (Resnick et al. 1993). Self-report surveys also allowed responses to crime events to become the object of study. Much of the attention was focused on why victims call or do not call the police, but the mobilization of resources other than the police has also been investigated with these data (Skogan 1984).

Theories of crime causation

Self-report victim surveys have contributed to the building of criminological theory. The availability of these data encouraged the development and testing of victim-centered theories of crime (Hindelang, Gottfredson, and Garofalo 1978; Cohen and Felson 1979). These theories focused on the occurrence of crime events rather than criminal motivation. They emphasized the routine activities of victims as sources of opportunity for the motivated offenders. The social, structural, and spatial location of victims influenced their routine activities, which in turn affected their risk of criminal victimization.

Because victim surveys provided a wealth of disaggregated and detailed data on victims and the social context of victimization, they were ideally suited to the testing of routine activity theory. Attributes of persons and social contexts could be used to measure concepts within opportunity or activity theory. This type of detailed information on victims and events was not easily or reliably available from police data. The testing of these theories was facilitated further by the fact that the surveys collected the same information from samples of victims and nonvictims. Using the data from NCS and other surveys, researchers confirmed that the basic tenets of opportunity theory were consistent with the data. The higher the exposure and the lower the guardianship, the greater the probability of victimization.

While opportunity and routine activity theories are a major contribution of victim surveys to criminological theory, the recently revived interest in repeat victimization warrants mention (Pease 1998). Early on in the development of crime surveys, scholars observed that a small number of victims accounted for a relatively large portion of victimization (Sparks 1981; Nelson 1980). A flurry of activity followed, wherein several articles were published demonstrating

that the distribution of repeat victimization was greater than would have been expected if the risk of victimization had been equal across persons. This led to speculation that repeat victimization was due to the fact that some people have different risks of victimization, so that persons with higher risk will become repeat victims in higher numbers than others with lower risk. This heterogeneity of risk would account for the distribution of repeat victimization observed throughout the historical development of victim surveys (Biderman et al. 1967; Sparks 1981; Nelson 1980). An alternative hypothesis was that the first victimization exposed the victim to subsequent victimizations, as in the case where a burglary makes the offender aware of other property, which motivates the offender to break in a second time. This was referred to as "state dependence."

Recent research using the British Crime Survey (BCS), as well as police records, has refocused attention on repeat victimization, reasoning that because repeat victims accounted for so much of the cross-sectional crime rate, it would be efficient to target resources to repeat victims and thereby lower the crime rate (Ellingsworth, Farrel, and Pease 1995). This research has found that prior victimization substantially increased the risk of subsequent victimization, and that the time interval between the victimizations was extremely short (Polvi et al. 1990). This work was used to develop police intervention programs that would take advantage of the newfound knowledge about repeat victimization (Forrester, Chatterton, and Pease 1988).

The fact that BCS is a cross-sectional survey limits the degree to which it can be used to investigate the sources of repeat victimization. The British work rekindled interest in repeat victimization in the United States, where the longitudinal data from NCS and other surveys may be more useful in disentangling the relative importance of heterogeneity in risk versus state dependence in explaining repeat victimization (Lauritsen and Davis Quinnet 1995). This research bears watching as an area where victim surveys can contribute to our understanding of why crime occurs (Pease 1998).

Victim surveys have also been useful for shedding light on the composition of the offender population. As part of many victim surveys, respondents are asked about characteristics of the offenders involved in "contact" crimes (i.e., those where the victim actually saw the offender). These data provide a profile of offenders that had not been caught by the police. Analysis has compared offender characteristics collected from victim reports with those provided in official records (Hindelang 1978, 1981). This research found considerable similarity in the characteristics of victims and offenders. That is, "people tend to victimize people like themselves" (Gottfredson 1986, 268).

This line of research has evolved by explicitly linking victimization to offending (Singer 1981, 1986; Lauritsen, Sampson, and Laub 1991; Gottfredson

1984). Analyses of surveys that contain reports of both offending and victimization have shown that reporting offending is linked to reporting victimization, especially for violent events. This relationship is an important theoretical jump that moves toward unifying the disparate discussions of offender motivation and victim risk into a general theory of criminal events.

Responses to victimization

Self-report surveys also offer a unique opportunity for understanding which resources victims mobilize in response to victimization. The principal focus here has been on why the police are called in response to criminal victimization. The most prevalent answer to the question of why people call the police is the seriousness of the event in terms of loss or injury (Skogan 1974). This has been the finding across many different types of surveys in many different countries (Mayhew 1993; Kury 1993; Skogan 1984). This tradition of research using crime survey data suggests that citizens respond to the nature of the crime only so that other factors, such as the perceived legitimacy of the police, are not as important (Garofalo 1977). One of the more interesting findings from a victim survey in this area has been that the nature of prior service by the police affects subsequent willingness to call the police in response to victimization (Conway and Lohr 1994). This analysis was done with longitudinal data from NCS, and it raises anew the question of whether factors outside the crime event can influence the decision to call the police. Perhaps additional analysis of longitudinal data will reveal nuances not visible in cross-sectional data.

Consequences of victimization

Finally, self-report surveys of victimization have been essential to identifying and explaining the consequences of victimization. Here again, the fact that the victim surveys include crimes both reported and not reported to the police provides a more complete picture. The surveys provide a reasonably good picture of the immediate durable harm (i.e., injury and loss) resulting from crime (Harlow 1989). The cost (both to insurance companies and out-of-pocket) of a recent burglary, for example, is captured reasonably well in victim surveys. A number of surveys have assessed various forms of psychic harm that can result from criminal victimization, specifically, sexual assault (Gidycz and Koss 1991; Resnick et al. 1993; Norris, Kaniasty, and Thompson 1997; Finkelhor 1997). These studies have found that depression and posttraumatic stress syndrome are more prevalent among victims of crime than among victims of other traumatic events. Moreover, they have found that some categories of victims (e.g., rape victims) experience more enduring psychic harms than others.

Summary

A large number of the analyses referenced in this section were conducted with large-scale, ongoing household surveys such as NCS or BCS. This was the case in part because these data were easily accessible over a long period of time. These surveys were also extremely large compared with special purpose surveys in the social sciences, and these large samples were required for the study of rare events such as violent crime. NCS had the additional advantage of being the subject of a great deal of methodological study to inform design decisions made both at the inception and over the life of the survey. This methodological work outlined the error structure of the survey so that users of the data could use them appropriately. The error profile of NCS contributed to our understanding of the victim survey method more generally. The next section describes the unique features of the NCS design and subsequent sections review the methodological work done to inform decisions about specific design features of the survey.

NCS had the additional advantage of being the subject of a great deal of methodological study to inform design decisions made both at the inception and over the life of the survey.

Unique Features of the NCS Design

NCS employs a rotating panel design of addresses in which persons in sample households are interviewed at 6-month intervals over 3½ years. All members of the household 12 years of age and older are asked about their victimization experience in the previous 6 months. In addition, one household member is asked to report on the theft of common property as well as on his or her own personal victimization. The survey includes a screening interview in which respondents are asked to recall and report potentially eligible crime events and to fill out an incident form that contains questions about the details of the event. This detailed information is used to determine if the events mentioned in response to the screening interview are within the scope of NCS and how the crime should be classified. The survey data are most commonly used to estimate both the level and change in the level of crime for the seven UCR Index crimes and simple assault.⁴

No other victim survey in the world has the same design as NCS. NCS employs a continuing rotating panel as opposed to a cross-sectional design. Occupants of each housing unit in the NCS sample who are 12 years of age or older at the date the household is contacted (household members, as they reach age 12, are added throughout the survey's 3½-year duration) are interviewed

seven times over 3½ years. Other victim surveys interview sample units only once and rarely include both children as young as 12 and adults as respondents in the same survey.

NCS employs an address sample taken from the list of addresses compiled in the decennial census and updated throughout the decade by the U.S. Census Bureau. Many other victimization surveys in the United States use lists of telephone numbers or random-digit dialing to obtain their samples. In many other countries, excellent universal national lists simplify sampling. NCS attempts to interview everyone in the housing unit who is 12 years of age or older; most other surveys interview only one (often randomly selected) person in the household, and generally that person is an adult (i.e., 16 or older). The Census Bureau works assiduously to keep noncompletions to a minimum, and its completion rates are rarely matched by any other general population survey.

Respondents in NCS are asked to report all victimizations that occurred during the past 6 months or since the previous interview. In theory, victimization data are to be used only from respondents who had been interviewed 6 months previously. The respondent's recall and a record possessed by the interviewer of incidents reported in the prior "bounding" interview serve to exclude events that occurred during the prior reference period from the current one. Respondent mobility and noncompletions are so common, however, that intolerable data losses would incur were data to be used only from individuals who had been successfully interviewed 6 months earlier. In practice, bounding means merely that the unit was in the sample at the time of the prior interviews. Most other victimization surveys ask the respondent to report on an entire year or longer and do not employ a prior interview for temporal bounding.

The instrumentation employed in NCS is divided into a screening interview and an incident form. The interview presents cues to the respondents that are designed to help them recall and report possible criminal victimizations. Once a candidate event is mentioned, the respondents are asked detailed questions about the event to determine if it is a crime of interest to the survey and, if so, to provide information that can be used to classify the crime. All of the screening questions are administered before the incident form is administered. Some victimization surveys employ a screener/incident form logic, but many others do not. In those that do not, any positive response to a screening question would be considered a crime event. The type of crime event would be determined by the screening question that elicited the positive response. Moreover, the gathering of information about the incident occurs immediately after the respondent answers a screening question positively and before other screening questions are asked. The implications of these different approaches to screening will be discussed in detail later.

NCS asks only one respondent per household about thefts of certain kinds of things that are considered the common property of the household. All respondents are asked about thefts of their personal property. Specifically, these household respondents are asked screening questions about burglary, motor vehicle theft, and the theft of specific household property such as plants or lawn furniture. Because most other surveys have only one respondent per household, that person is asked about the theft of his or her personal property as well as the theft of common property of the household.

NCS uses a “series incident” procedure to accommodate victims who report a large number of victimizations and cannot report the details for each incident. Currently, if a respondent reports six or more incidents that are similar in kind and cannot provide the date and other details for each of the six events, then all of the events are treated as a “series incident.” This means that the interviewer notes the number of events but completes an incident form only for the most recent one. Some victim surveys count all the events that a person mentions without concern for the ability of the respondent to recall the date of or other specifics about the event. Still other surveys record the number of events but administer the incident form only on a set number of events (e.g., a maximum of five). Limiting the number of incident forms is an attempt to reduce the burden on respondents and interviewers.

There are other ways in which NCS differs from other large-scale household surveys of victimization, but the features mentioned previously are among the most consequential from a cost and error standpoint. These particular design features were adopted because those implementing NCS believed that a survey designed in this manner would minimize the error in the estimates of the change in the level of crime. Some of the evidence relevant to these design decisions and the evolution of NCS to its current form are described in the next section.

The Evolution of NCS

NCS evolved into its current design in a series of stages. The first stage set the foundation for what followed.⁵ In the early 1930s, the Wickersham Commission proposed a comprehensive national criminal justice statistics program under an independent central statistical agency. Although this plan did not achieve fruition, it led to making a cooperative national system for statistical reporting a Federal function under the FBI (National Commission on Law Observance and Enforcement 1931). Annual reports of a crime Index in UCR that the FBI compiled from these data became the most influential indicator for defining the seriousness of the Nation’s crime problem.

Thirty years later, this achievement of criminology and statistics was increasingly being called into question by official and journalistic investigations of police offense statistics and by critical social science analyses. In the early 1960s, a few social scientists speculated about the possibility of adapting self-report national household surveys to produce an indicator of the nature of and changes in the crime problem that would be less vulnerable to the vagaries attributed to UCR.

In the second stage, these ideas received a receptive hearing from two new presidential commissions appointed in 1965 for the reform of law enforcement and the administration of justice (hereinafter referred to as the President's Commissions). During this period, the fundamental idea that citizen self-reports of crime could be used as the basis for crime statistics was formulated, proposed, and accepted by government officials and the public. In addition, many of the important methodological and logistical issues required to field a victim survey were addressed by several pilot studies.

In the third stage, the Census Bureau addressed many of these issues within the context of a large-scale household survey in preparation for fielding NCS on an ongoing basis. Some of the lessons from the earlier field tests were incorporated into the Census version, but others were not. A number of methodological tests were done during this period, shedding additional light on the effects of various design features on reporting in victim surveys.

The fourth stage began with the launching of the actual survey. After the National Criminal Justice Information and Statistics Service (the precursor to BJS) published the first of its annual reports, *Crime in the United States 1973* (U.S. Department of Justice, National Criminal Justice Information and Statistics Service 1976), the survey immediately achieved prominence as an indicator in the public media and in academic research and discussion. As a major social survey, it attracted the attention of leading experts and organizations in the social sciences and statistics as well as the U.S. Congress. Specific problems which the leading experts and the BJS census team identified led BJS to sponsor a reevaluation of the National Crime Survey by a committee of the National Academy of Sciences (NAS). Shortly after the publication of the results of the NAS study, congressional hearings were held on the possible suspension of NCS (U.S. House 1977).

The NAS evaluation resulted in a fifth stage, during which a 5-year program of research, instrument development, and redesign planning was conducted to deal with the issues raised by the report. During this period, a large number of methodological tests were conducted with particular emphasis on underreporting and screening issues.

The sixth stage of development began with the implementation of the changes in the design of NCS that were recommended as part of the redesign research. Again the Census Bureau engaged in extensive testing of various designs in preparation for implementing the new design.

In the remainder of this section, the last five of these six stages will be discussed in more detail by noting the advances made for designing and conducting victimization surveys.

Setting the stage in the 1960s

The confluence of several forces made the 1960s an auspicious time for the development of victim surveys. The brief period of *détente* in the Cold War moved defense-related issues off the front page. Demographic changes, both in terms of both the Baby Boom and the movement from rural to urban areas, moved crime to the forefront as a public issue. The waning of defense issues freed research and development professionals to seek other pursuits, and survey research enjoyed rapid growth. At the same time, UCR was coming under fire for not accurately reflecting how the crime problem was affecting society (Biderman 1966). These factors provided the skills, technology, and motivation for exploring the possibility of a victim survey.

The proposals for using interview surveys of samples of the general population, or polling methods, for measuring crime incidence rested on the belief that a vast reservoir of crime was not reflected in the statistics on offenses known to the police. It was recognized that many crimes were not reported to the police, and that officers at all levels of report processing could exercise great discretion in recording events. Reforms of several of the Nation's metropolitan police departments were accompanied by exposés of the previous practice of killing crime on the books. It was suspected that more reports would make their way through to published statistics when police departments believed crime was not being controlled properly because they were not allowed adequate resources or freedom of action.

For a sample survey to be practical and useful as a measure of levels and changes in rates of crime, two things had to be true (Biderman 1967):

1. The existing statistical indicators had to be found to be so inadequate and potentially misleading that it was worthwhile to develop and test an untried and expensive alternative.
2. The existing indicators had to be erring in the direction of massively understating crime rates. Were crime incidence not much higher than the official

statistics suggested, extravagantly large and expensive samples would be required to achieve sufficient numbers of incidents for statistically reliable results. In 1965, the total rate for all Index offenses combined for that year was a bit under 1,500 per 100,000 persons. If this was the true rate of crime, the expected number of robbery victims in a sample of 10,000 persons would be fewer than 10. The chances of encountering even one rape victim in such a sample would be quite remote. At that time, there was only one continuing national survey with a sample that large.

Research would be needed to demonstrate that both of these conditions were true. In the remainder of this section, we highlight the key milestones, both political and methodological, that led to the creation of NCS.

The President's Commissions' studies

The development of the crime victimization survey began in earnest with explorations for two commissions appointed in 1965 by President Lyndon B. Johnson: the President's Commission on Crime in the District of Columbia and the President's Commission on Law Enforcement and Administration of Justice. In cooperation with the President's Commissions, the first research grants by the Office of Law Enforcement Assistance, precursor of the National Institute of Justice, included the research and development of interview surveys to illuminate public experiences with crime and with justice agencies. A goal of both President's Commissions was to reduce the amount of crime that eluded the attention of the police. This was to be accomplished, in part, by increasing citizen cooperation with law enforcement (e.g., increase the amount of crime reported to police). The President's Commissions realized, however, that this goal had to be coupled with developing measures of the incidence and impact of crime that were independent of the efforts of the police. Otherwise, reliance on national or local statistics on offenses reported to the police or which otherwise became known to the police might be paradoxically affected by the President's Commissions' successes.

The victimization survey developed rapidly in its early stages. The idea of the survey was first broached in writing to the D.C. Crime Commission in September 1965 and to the President's Commission on Law Enforcement and Administration of Justice shortly thereafter. Independently of the President's Commissions, the National Opinion Research Center (NORC) incorporated victimization items in its ongoing omnibus amalgam survey in November 1965. The initial pilot survey, conducted by the Bureau of Social Science Research (BSSR) for the two President's Commissions in three Washington, D.C., precincts, began in January 1966. Field work for a supplementary BSSR precinct study that began in July 1966 was integrated with precinct studies by the University of

Michigan in Boston and Chicago using the same instrument and method. Interviewing for a national survey of 10,000 households by the NORC was conducted that same month. By the end of the year, all three organizations (NORC, BSSR, and the University of Michigan) had completed their exploratory studies. Their reports were published in three separate volumes by the President's Commission on Law Enforcement and Administration of Justice (Biderman et al. 1967; Ennis 1967; Reiss 1967b). The BSSR pilot survey is given prominence in the following discussion because it clearly shows a link between this early research and many of the methodological issues that have continued to emerge to the present day.

The BSSR pilot studies

The initial set of BSSR pilot studies identified many issues that persisted throughout the development of NCS. Many of these issues, especially with regard to screening and scope, remain controversial among researchers to this day. Much of this work was the result of stating the case for why and how such surveys might be done. Other valuable information came from collection and analysis of the data.

Lessons from conceptualization and planning. From the start of the planning process, two contrasting aims had to be reconciled. On one hand, there was the need to present results that could be compared directly with those from police statistics. This restricted much of the planning to the conceptual structure, definitions, and perspectives of police statistics, and to the UCR Index offense rates, in particular. On the other hand, it was important for the survey planners to incorporate in their instruments provisions for information on incidents and their victims that had not been collected before. In some cases, these two goals were in direct conflict.

The provisions of the surveys for comparing police and survey statistics had to permit adjusting survey *victimization* rates of individual persons (which UCR once claimed to be, but in many key respects was not) and infer from them *offense* rates for specific jurisdictions. This was no easy task because it required the designers of these surveys to stretch the methodology in a variety of ways, including:

- **Place.** Police statistics provide rates of occurrence within a jurisdiction, not for residents of that jurisdiction. What befell residents from the suburb while in the city had to be discounted for survey comparisons. Using a national survey for comparisons with national UCR rates is simpler, but not where comparisons are made for subnational places or types of places.

- **Residential mobility.** There was recognition that some means would be needed to deal with persons moving into and out of the areas under study. Although necessity might dictate the assumption that the premove victimizations of in-movers balanced out the postmove experience of out-movers, this assumption was tenuous.
- **Multiple and collective victims.** For offense classes where police statistics count only one incident even when there may be multiple victims, the survey needed provisions to identify events that someone else eligible for the survey might also give, if sampled.
- **Offenses against organizations.** As a sample of households and persons, the survey was an inappropriate vehicle for collecting information on crimes against businesses and other organizations. Separate exploratory surveys were undertaken of samples of businesses that are not under discussion here (Reiss 1967a, 1969; Aldrich and Reiss 1970). However, UCR does not consistently distinguish between residential and commercial crime. This made the comparison of UCR with the BSSR pilot data difficult. The decision was made to include robberies and other offenses against the person carried out against a respondent at a business or who is performing an organizational role, with the harm done to the individual distinguished from that done to the organization.

Although the requirement of making victimization estimates from the survey comparable to police offense rates constrained those designing the BSSR pilot survey, this survey was different from UCR in several important respects. First, the BSSR pilot survey would not attempt to validate crime reports in ways similar to police records. Theoretically, police reports are backed by the officer who fills out an incident report. The information from persons claiming to be witnesses or victims is subject to evaluation, and the report, in turn, is subject to evaluation at higher organizational levels and may be labeled unfounded on many grounds. The survey method, by contrast, places its ultimate trust in the unsupported testimony of the individual citizen respondent. It is assumed that the pledge of anonymity and the absence of material consequences, positive or negative, for the information given, should leave respondents with scant motive for deceit, invention, embroidery, reticence, or other departures from disinterested performance (Biderman and Reiss 1967). Rather, the survey exercises quality control by trying to identify miscomprehension or incorrect execution of the procedures.

A second difference lay in the scope of offenses covered. Provision was made for the interview to cover victimizations by a far more extensive range of offenses than the set making up the UCR Index. It included any acts of which

the respondent was a victim and that the respondent thought was a crime in that it could be punished by imprisonment or fine. It included cues to a variety of frauds, forgeries, swindles, extortions, defamations, false accusations, and official misconduct as well as arson and vandalism. Proponents of direct UCR–NCS comparisons viewed this expansion to be counter to the goals of the survey. As will be noted in a later section, this feature of the BSSR pilot survey was greatly curtailed in later implementation by the Census Bureau and DOJ.

A third area expanded by the BSSR pilot survey was measuring the impact of crimes. The offense classifications used by UCR were highly constrained by its need to provide the least able cooperating departments with a set of categories and instructions for sorting and hand-tallying offenses in each category. The survey was not restricted by these categories. Planning for the survey could envision more refined discriminations within the traditional common-law categories and categorizations along other dimensions, as well. The survey instruments could explore the significance of victimizations from the victim's perspective, and they could cover many variates of relevant social values or policy issues.

Lessons from fielding and analysis. Once the interviews for the BSSR pilot survey began, a number of fundamental conclusions, both substantive and methodological, emerged. The first related to the salience of victimization events. The BSSR pilot survey found that most victimizations were not readily recalled by respondents, including victimizations that are classifiable as Index offenses or have high scores on the Wolfgang-Sellin seriousness scale (Biderman et al. 1967). Increments in the specificity of questions, prompts, and pauses for reflection brought forth large increments in the number of victimizations recalled. The first BSSR pilot survey questionnaire employed 70 discrete probes for victimizing incidents. Although these facts figure in the literature primarily for their methodological significance, their substantive significance for criminology is also important. That crimes are not highly salient events in memory implies that they do not rank high relative to many other life events in their importance for individuals. If we reflect upon how crowded lives can be with trials and tribulations of everyday life, even the most serious crimes are, apparently, paltry. The earliest report gave other reasons that so many victimizing incidents were apparently forgotten:

Forgetting these events also stems from the unpleasant and embarrassing aspects of the experience. . . . Further, few of the incidents led to a path of action that might serve to reinforce the ability to recall the event. The large majority of the . . . [events] are happenings that would have been difficult to avoid—measures to prevent repetition . . . would usually involve greater cost than . . . the risk deserves. In very few of them is the victim known; hence there is no individual target on whom the victim can fix

whatever affect the event may arouse. In most instances, there is nothing to do to gain either material or emotional indemnification for the loss. (Biderman 1966, 12)

The final report of the BSSR pilot study went on to explain why the low recall salience of crime incidents does not mean that they were unimportant events for the victims. Their importance, it was argued, resides in their being indicative of the fragility of the social order, and these experiences are assimilated to and may be outweighed by other signs of disorder.

A second important finding was that the incidence of victimization was far more frequent than existing statistics suggested. The feasibility of a national survey was asserted in a progress report 3 months after the pilot project began (Biderman 1966; also reproduced as appendix G in Biderman et al. 1967). The report was based on the high percentage of respondents giving victimization reports in pretest interviews and in the earliest interviews of the survey proper (only 183 interviews in all).

A third important substantive finding related to the great excess of the survey rates over those reported to the police. To compare police and BSSR pilot survey data, a procedure was applied for reconciling survey offense rates with those of police reports for the same precincts. Even after eliminating from the calculations those incidents that respondents said had not been reported or were not otherwise known to the police, the survey rates were far higher. The conclusion was that nonreporting *by* the police may account for more of the dark figure than nonreporting *to* the police.

These initial trials identified problems of interviewing for victimization that have continued to receive methodological attention to the present day:

1. A recency bias in recall so pronounced that a reference of period of no more than 6 months was recommended for future surveys.
2. The need for singularly focused incident recall tasks.
3. Far greater victimization reporting by self-respondents than by household members acting as proxy respondents for other members.
4. The problem of crimes against the household and of multiple-victim incidents.

Integrating results across surveys

The foregoing results of the BSSR pilot studies were both consistent with and contrary to the NORC pilots. By comparing and contrasting across these surveys,

several key findings emerged. The first was the revelation that the incidence of crime generated by victim surveys was sensitive to survey procedures. The Washington, D.C., study (Biderman et al. 1967) and the NORC survey (Ennis 1967) used different methodologies. The Washington, D.C., field test was organized around principles that would facilitate recall and reporting of crime events. In practice, this meant minimizing cognitive burdens that occur when the interview imposes official rules, terms, and definitions that hamper straightforward internal and conversational discourse. The procedures avoided complicating the respondents' memory work with filtering, composition, and decomposition tasks to make their thoughts and answers fit official categorical molds.

This contrasted with the tack taken by the NORC questionnaires. Those questionnaires used a battery of screening questions, each one devoted to a specific Uniform Crime Reports crime class and containing all the elements needed to define a victimizing event as belonging to that class. Screening questions were worded to exclude experiences that did not fit the official definition of the crime class to which the item was devoted. They included wording that sought to ensure that the item encompassed all the experience fitting the criteria for the class. By having the respondent answer positively to only one screening question for any incident, analysts could make its preliminary victimization counts by crime class simply by tallying "yes" answers to screening questions. The NORC survey then followed the next step of police statistics: further incident interviewing to inform an "unfounding" procedure for eliminating questionable reports. This included providing interviewer ratings of the veracity of the respondent's testimony and then a review by experts of a subsample of incidents, including police and lawyer raters of the incident report for inconsistencies and appropriate classification.

The NORC approach made for long screening questions, as illustrated by this one for robbery:

Within the last twelve months, did anyone actually take or try to take by force or threat of force from you personally or anyone in the household any money or property? This would include bicycles forcibly taken away from children, or a violent purse snatching. (Ennis 1967, appendix A, 3)

As previously noted, the BSSR Washington, D.C., pilot survey, by contrast, proceeded by orienting the respondent to the crime victimization recall task and then presented the respondent with a long list of short cues, largely of between one and five words, giving the respondent time to think between each one. The screening questions were not to be used as data (other than for methodological analyses), but simply triggered the execution of an incident form. The detailed incident questioning had the burden of getting the information needed to

The pilot surveys answered the basic questions about the need for and the feasibility of a survey-based indicator of crime. Moreover, they identified (and informed) many of the basic design issues in creating such surveys.

determine what offense(s), if any, had occurred in the incident(s) the respondent recounted, who the victim(s) was (were), and additional information about the incident and its aftermath.

The BSSR pilot survey procedure yielded far higher annual victimization rates (0.80/respondent) than that yielded for central cities by NORC (0.08). After taking account of what was learned in the initial Washington, D.C., pilot work, the BSSR instrument was modified in collaboration with the University of Michigan's Metropolitan Areas surveys for the President's Commission on Law Enforcement and Administration of Justice. The revised BSSR interview procedure yielded approximately 2.0 incidents per respondent (Biderman et al. 1967, 50). The greater productivity of the BSSR/Michigan survey suggested that one's approach to screening will dramatically affect the resulting incidence estimates.

More specifically, it suggested that organizing surveys in a manner consistent with the principle of facilitating the recall and reporting tasks was preferable to emphasizing legal principles which complicated the respondent's task.

Another indication of the dependence of the rates yielded by the survey on method was the positive correlation between education level and victimization by the types of crimes where it might not be expected. This suggested better performance as an interview subject of the better educated. Biderman (1967) wrote more generally of "class-linked under-reporting" in the survey.

A second key finding found across all the pilots was a severe recency bias in the data. This was observed by increased reports of victimizations at the earliest and most recent ends of the reference periods. Increased reports at the beginning of the period were thought to reflect incidents occurring outside the period being brought forward in time into the period. The increase at the end of the period was seen as a mix of telescoping and greater recall of events that are closer to the interview. These phenomena, identified earlier in a Census Bureau experimental housing survey (1965), were regarded as applicable here and figured in much of the future design research on NCS.

Summary of the pilot studies

The pilot surveys answered the basic questions about the need for and the feasibility of a survey-based indicator of crime. Moreover, they identified (and

informed) many of the basic design issues in creating such surveys. These studies confirmed suspicions that police data substantially underestimated the level of crime because both citizens failed to report and the police did not record eligible events. They also showed that there was a vast reservoir of crime that could be estimated with a household survey. Grappling with actually fielding such a survey identified the design issues that needed to be addressed. Principal among the lessons learned here was the inherent tension between the logic of police record systems, particularly UCR, and the logic of surveys. The constraints of the former would prohibit realizing the full potential of the latter. Asking about crimes using the legalistic framework of police record systems would probably inhibit complete reporting of events in the survey. Moreover, constraining the scope of crimes and the information collected about crime events to that which is customarily included in police record systems would fail to exploit the potential of these surveys.

Other valuable lessons were learned. First, respondents had trouble recalling and reporting events, so steps should be taken to facilitate the task. The recall task should be simplified and many cues should be provided to jog memories. Second, temporal placement of events within and outside of the reference period by respondents was problematic, so some attention should be given to making this easier. Third, self-respondents were preferable to proxy respondents. Fourth, some attention needed to be given to the problem of reporting the theft of collective property. Asking everyone in the household about these items would result in some duplicate reporting, but asking less than everyone would result in underreporting.

Implementation of the survey by BJS and the Census Bureau

It was clear from the pilot studies that large samples would be required to obtain reliable estimates of victimization for crime classes of intense interest (e.g., rape).⁶ The Census Bureau was the only organization that could field such a large survey and was chosen to conduct the ongoing NCS. In preparation for implementing the survey within the Census environment, some of the lessons from the field surveys were taken into account although others were ignored. In addition, the Census underwent an extensive program of pretests, trial surveys, and record check experiments beginning in 1970 (Lehnen and Skogan 1981).

Several important features of the current NCS design resulted from this work. One experiment in Dayton and San Jose (Kalish 1974) assessed the effectiveness of proxy reporting for the survey. This continued the line of work reported by the President's Commissions that proxy respondents were far less productive than self-respondents. The conclusion from the pilot studies, not surprisingly,

was that when a single respondent reported for the entire household, far fewer crimes were reported when compared with interviewing all members of the household as self-respondents. As a result, the idea of interviewing all members of the household was eventually adopted for NCS.

A series of reverse record check studies was also conducted in different cities. These surveys drew a sample of known victims from police records and then interviewed these individuals to see if the incident was reported on the survey. Theoretically, this type of study provides an external criterion to judge the accuracy of reports on the survey. These studies were used primarily to determine the accuracy of the recall of incidents by respondents. The Census researchers drew two main conclusions about the optimal length of the recall period:

If the objective is to determine whether a crime occurred, as opposed to placing it in a more accurate timeframe, then a 12-month reference period is as good as one of 6 months. . . . To the extent that it is desirable to place an incident in a specific timeframe, greater accuracy is obtained from a shorter reference period. Thus, a 6-month reference period is better than 12, and a 3-month period is better than 6. (Dodge and Turner 1981, 3)

These conclusions were used as a basis for a 12-month reference period for surveys done across cities (Hindelang 1976). However, this basic result was not accepted by a number of researchers (e.g., Biderman 1981a; Biderman and Lynch 1981, 31), partly due to the problems associated with a reverse record check design (see following text for problems). A 6-month reference period was eventually adopted for NCS.⁷

A second basic result from the reverse record check studies served as a precursor of issues that still haunt victim surveys. This result was the conclusion that

[recall] was very high for crimes involving theft of property (80 to 85 percent). With respect to personal crimes, robbery was well reported (75 percent and above), but rape and assault were less so ($66\frac{2}{3}$ percent and 50 percent, respectively). An important factor in the recall rates for cases of personal victimization is the relationship of the offender and victim. Recall rates vary directly with the nature of that relationship; that is, when victim and offender are strangers, recall rates are high. . . . Acquaintance, and even more kinship, results in lower reporting rates. (Dodge and Turner 1981, 3)

As will be noted, one of the primary faults found with victim surveys has been their inability to illuminate violence among persons that know one another.

A third feature of NCS was also adopted from these experiments: the use of a household respondent to report about crimes against household property. This was based on the conclusion that a single household member could report on crimes such as burglary, auto theft, and larcenies against household property (e.g., lawn furniture, plants). This resulted in arranging NCS screening so that a single person (the household respondent) is administered a set of screen items that specifically ask about these types of crimes. Once this part of the screen is complete, all household members are administered a set of questions that are meant to apply to personal crimes.

It should be noted that the final design of NCS was, in several ways, contrary to the recommendations initially made by the President's Commission on Law Enforcement and Administration of Justice based on the results of the field surveys. Specifically, the principle of facilitating recall and reporting was compromised somewhat in favor of some of the legal principles and the desire to classify crimes neatly (Dodge and Turner 1981, 4). The major impetus behind this was the attempt to mimic UCR. The Census survey restricted its screening to Part I crimes in UCR, such that questions were asked with the intent of eliciting mentions of these crimes and only these crimes. Although the Census instrumentation separated the screening task from the provision of detailed information for classification, there was a one-to-one correspondence between the screen questions and the UCR crimes. Related to this was that the NCS questionnaire departed from the "short cue" approach adopted in the BSSR/Michigan pilot studies in favor of a more rigid approach that attempted to direct attention to legal categories. Evidence from the pilot studies, as well as evidence that has been cumulated since (see the following discussion of the NCS redesign), suggest that all of these departures reduced the rate of reporting in NCS.

Other design features of NCS were occasioned by the need to fit into the organization of the Census Bureau and the Current Population Survey (CPS). CPS is the largest intercensal survey conducted in the world and, at the time, NCS was to be the second largest of these surveys. Sharing interviewers between the two surveys would mean great efficiencies for the organization. CPS employed a rotating panel design. This was viewed as an advantage to NCS for a number of reasons. One was the ability to use prior interviews to "bound" subsequent interviews (Neter and Waksberg 1964). A second was that the rotating panel design substantially increased the precision of the year-to-year change estimates. The panel design feature produces a natural positive correlation across annual estimates. This, in turn, substantially reduces the standard error on change estimates.

In addition to these decisions regarding the design of NCS, the Census Bureau also instituted a survey of commercial establishments and a set of cross-sectional

surveys conducted in a number of the largest U.S. cities. The commercial surveys were developed because exploratory studies of small business showed that these establishments had victimization rates several times that of households (Reiss 1967a, 1969; Aldrich and Reiss 1970). Moreover, the household survey was not a good vehicle for measuring this component of the crime problem. The city surveys were fielded in an effort to evaluate the impact of crime prevention and crime reduction programs implemented with DOJ funding in the largest cities. These surveys were intended to assess the change in the level and distribution of crime in these cities as a result of the programs.

The National Academy of Sciences report

When NCS began to produce information on crime, various groups began to question the quality and usefulness of these data. Groups supportive of police-based crime statistics were already suspicious of this new data collection system. Academics began to raise questions about a multimillion-dollar data collection with few variables that could be used in testing theories of crime and that could not produce estimates for local jurisdictions. They also worried that this new data collection would take funds away from criminological research. The sponsoring agency also began to wonder about its new creation when the first years of data began to show the same large increases in crime as UCR (Parkinson, Paez, and Howard 1977). While much of the concern was focused on the commercial and city surveys, not NCS, all aspects of this new data collection came under scrutiny.

In response to these concerns, the Law Enforcement Assistance Administration asked the Committee on Social Statistics of the National Academy of Sciences–National Research Council (NRC) to evaluate the surveys. The committee selected a panel that represented a variety of disciplines and recruited staff to carry out the investigations necessary to perform the work. The study took place between January 1974 and June 1976. The panel examined every aspect of NCS, from the goals of the survey to its staffing and management to the publications produced with NCS data. The panel's recommendations and deliberations were published in *Surveying Crime* (Penick and Owens 1976).

Many of the panel's recommendations were pertinent to the management of the survey within the Census Bureau and DOJ, and others sought the elimination of the commercial and city surveys. Among the recommendations that addressed survey design, procedures, and instrumentation were some familiar calls for improvements in screening procedures. The panel suggested that:

[1] The function of screen questions should be to facilitate . . . recall and reporting of happenings that fall within the scope of the survey. The

usefulness as data of screening responses themselves is restricted to their use in the analysis of the effectiveness of screen questions. The operating rule is that data for final tabulations come from the detailed exploration of the pertinent events by the use of subsidiary incident forms. (Penick and Owens 1976, 82)

[2] A set of well defined screen questions must recognize that there are not just seven crimes in which one is interested. . . . The screener can be a cue to any. . . . The screen questions can also be a cue to some element of place or some other circumstance of the victimization that may help bring about mention of an incident. . . . (Penick and Owens 1976, 84)

[3] Screening procedures should reflect less worry about redundancy and about eliminating ineligible events than about unnecessarily cluttering up discrete questions and the respondent's thinking. (Penick and Owens 1976, 87)

[4] The household informant should be limited to questions on breaking and entering and to household property items. . . . Alternatively for research purposes, everyone who is interviewed within the household should be asked about household as well as personal, crimes. Interviewers . . . should assume the burden of eliminating separate mentions . . . of identical incidents, including the theft of jointly owned property such as automobiles. (Penick and Owens 1976, 87)

The panel also recommended that the "screen questions take account of the large volume of incidents now classed as series" (Penick and Owens 1976, 88). The urgency of this issue became apparent with the availability of the first years of data from the survey, in which "an estimated 20–30 percent of reported personal victimizations were treated as a series and excluded from the personal victimization count." The panel took exception to the fact that the series incident procedure (1) excluded a large number of relevant events simply because they did not conform easily to the incident logic of the survey, (2) required the respondent to make difficult judgments about combining a set of incidents into a series and to estimate for each series the number of events involved, and (3) allowed the determination of whether the series procedure should be invoked to be made by the interviewer in the field rather than by data analysts.

In addition to these recommendations pertaining to screening, the panel called for research and development work on the best combination of reference period, frequency of interview, retention time of an address in sample, and bounding rules that would address the following questions:

[1] Should the reference period be 3 months, 6 months, or 12 months?

[2] Should a household address be interviewed once, twice, three times, seven times, or some other number?

[3] Is the bounding interview worth its cost and does it introduce a new significant bias into the results?

[4] What are the shapes of the reporting decay interviewer overload and telescoping functions? (Penick and Owens 1976, 68)

Surveying Crime and the work of the NAS–NRC panel raised questions about some of the design decisions made during the implementation of NCS within the Census Bureau. These questions would soon become the agenda for a program of research and development that would shed further light on the relative desirability of different approaches to surveying victims of crime.

The NCS redesign and other improvements

Since the publication of the NAS report, a number of studies have been completed that have informed the design of NCS as well as the conduct of victimization surveys more generally. In response to the NAS recommendations, BJS sponsored a long-term redesign of NCS by convening a consortium of Government, private, and academic experts in various fields relevant to the design of NCS (e.g., statisticians, criminologists, survey researchers). The result was to implement significant changes to the survey, the most drastic coming in 1992.⁸ Research related to the design of victimization surveys more generally was undertaken by other researchers interested in improving the method. In this section, we briefly review the results of this research.

Reference period

As noted in the NAS report, further work was needed to assess the optimum reference period for NCS. The reverse record check studies completed in the early 1970s were a first step in this process, but were not viewed as definitive. A reference period experiment (RPE) was conducted by the U.S. Census Bureau and sponsored by BJS in the 1978–80 period. In the experiment, portions of the NCS sample were randomly assigned to interviews with 3-month, 6-month, or 12-month periods. Analysis of these data found that aggregate level estimates increased substantially as the reference period was shortened (Bushery 1981). The 3-month period produced significantly higher rates than the 6-month period, which produced significantly higher rates than the 12-month period. This finding runs counter to the conclusions of Dodge and Turner (1981) that the production of incident reports did not vary by length of the reference period.

RPE also indicated that the 3-month period displayed significantly different relationships of victimization with key sociodemographic variables when compared with the 6-month period. In particular, the effect of age was found to be stronger for all personal crimes in the shorter period, and the relationship to race was stronger for serious assaults and robbery (Kobilarcik et al. 1983; Cantor 1985).

Czaja and Blair (1990) conducted an experiment that compared reference periods of 6 and 9 months for three different types of crimes (burglary, robbery, and assault). They found significant underreporting of all three types of crimes, with burglary (16 percent) and robbery (28 percent) having significantly lower underreporting than assault (71 percent). The authors attribute the significantly higher rate of underreporting for assault to conceptual issues related to whether the victim defined the event as a crime. They also found that underreporting varied by race. Nonwhites were significantly more likely not to report the crime than whites.

Czaja and Blair (1994) did not find reference period length to have a significant effect. Respondents seemed about equally able to report crimes across the two different reference periods. This is consistent with initial analysis of the early reverse record check studies previously discussed (Dodge and Turner 1981), but it is inconsistent with RPE. It seems likely that the differences between RPE and the reverse record check studies is due, at least in part, to the types of crimes that were investigated across the two studies. RPE asked about all types of crimes, while the reverse record check studies examined events that were reported to the police. The latter are most likely to be remembered by respondents, so the difference of 6 to 9 or 12 months in the reference period may not be critical for reporting these types of crimes. It may be more important for crimes that are never reported to the police.

Using the results of RPE, several individuals developed formal statistical models that quantified the error properties of designs with different length reference periods (Lepkowski 1981; Bushery 1981). These analyses are based on the assumption that any increase in the rate of reporting victimization is better. Under this assumption, these analyses concluded that using a 3-month reference period was the best alternative among the three tested in the experiment. However, the assumption that more is better has been questioned in a number of contexts (e.g., Skogan 1981, 12). Increased reporting rates may occur, for example, if respondents telescope more crimes into the reference period.⁹ Ultimately, NCS did not shorten the reference period, in part because of a fear that such a change would have a serious impact on the statistical power of key comparisons.¹⁰ Nonetheless, this line of research has led designers of victimization surveys to be cautious when trying to extend reference periods beyond 12 months and to prefer shorter reference periods whenever possible.

A second change considered by the redesign was to simplify the reference period for NCS respondents. NCS imposed both an early and recent boundary on the period. Respondents were asked to report for the 6-month period ending at the end of the previous month and beginning 6 months prior. For example, if interviews were conducted in August, respondents were asked to report for the time period between the end of July and the beginning of January. As previously described, this ran counter to the initial pilot designs, which specifically deemphasized all of the reference period boundaries. The idea was to emphasize the recall of any eligible incidents without placing many filters on the cognitive task of the respondent. The use of any specific boundaries ran counter to this basic premise. The use of two boundaries, as in NCS, further complicated the recall process.

The redesign of NCS found that instituting the most recent bound resulted in substantial telescoping of crimes from the month of interview to the last month of the reference period (Biderman et al. 1986, 80). For this reason, starting in 1992, NCS asked respondents to report victimizations up to the day of the interview.

Improvements in screening

As previously noted, the early pilot studies indicated that respondents need cues and examples to help them recall incidents of victimization (Biderman et al. 1967). Specific screen cues and questions serve both to orient respondents to the types of events covered by the survey and to jog their memories for incidents that do not immediately come to mind as instances of crime (Biderman et al. 1986, 88–103).

In addition to jogging their memories, it is desirable to reduce any inhibition respondents might have to report crimes that might be sensitive. Victims may be reluctant to report incidents that are a source of pain, fear, shame, or embarrassment. One way of coping with a painful experience, in fact, is to try to forget it. Reporting the incident in a survey forces the victim to reexperience it and, perhaps, disclose information that could become known to other household members. Of particular concern is the gross underreporting of domestic violence on household victimization surveys.

Procedures for conducting NCS were not set up to promote disclosure of incidents among household members. For example, the Census Bureau does not treat the guarantee of confidentiality as applying to other people who are present during an interview; as a result, many NCS interviews are not conducted in private. There is evidence that this does, in fact, inhibit reporting of violence. In the case of domestic violence, the offender may actually be present (Coker and Stasny 1996).¹¹

Experimental tests conducted as part of the redesign tested screening strategies using an enhanced cuing approach. It was found that this approach significantly increased the number of reports of all types of crimes relative to the screener used on NCS (Biderman et al. 1986, 104–165). Increases were thought to be due to widening the concepts respondents have about eligible events as well as facilitating retrieval from memory. In the late 1980s, the Census Bureau went on to test a revised version of this screening strategy in several field experiments. The tests uniformly showed increased reporting of all types of crimes except robbery and motor vehicle theft (Hubble and Wilder 1988; Kindermann, Lynch, and Cantor 1997). The strategy was eventually adopted by NCS. Particularly large increases occurred for crimes that were thought to be the most underreported, such as sexual assault (especially among nonstrangers) and simple assaults (many of which are attempts without completion).

The introduction of a computer into the survey process changed not only the way interviews were administered but also the way survey organizations were managed.

Computer-assisted interviews

The introduction of a computer into the survey process changed not only the way interviews were administered but also the way survey organizations were managed. These changes have had a dramatic effect on the quality of the information that is collected on surveys in general (Couper et al. 1998) and NCS in particular.

The past 15 years have seen the universal adoption of computer-assisted telephone interviews (CATI). This has generally been seen to have had a positive effect on data quality for three reasons. First, it allows for programming more complex skip patterns. This takes the burden of navigation off of the interviewer and allows her to concentrate on the respondent (at least theoretically). Second, the computer forces the interviewer to at least see and review all questions for all respondents. When administered by paper and pencil, interviewers have more control over what questions they will and will not administer to the respondent. If they view certain questions as burdensome or feel a respondent may not react well to them, they can easily skip over them. When a computer is used, they are at least forced to view the screen before passing through. This could be especially important in conjunction with the detailed screening strategies described previously. With the increased cuing, the screener is longer and could be viewed as especially burdensome on the respondent. There might be more of a tendency to skip parts of the screener if the interviewer does not believe the questions are worth asking.

A third advantage to CATI is that administration is centralized. Rather than interviewers working out of their homes, they work in a central facility. This allows much tighter quality control over their work. Interviews are routinely monitored by supervisors. This makes it much more difficult for interviewers to deviate from accepted protocols, not to mention fabricating data.

As part of the redesign of NCS, the Census Bureau conducted a series of split-ballot experiments investigating the use of CATI to conduct interviews. Cases were randomly assigned to be interviewed either by CATI or by an interviewer out of his or her home by telephone (the traditional method). The results indicated a substantial increase in the reporting of all types of crimes in the CATI condition (Hubble and Wilder 1988). It is not clear whether the increase was due to computerization of the instrument, the centralized monitoring, or both. Nonetheless, as the new methods of NCS were implemented in 1992, a significant proportion of the jump in the reported victimization rate was attributed to this aspect of the redesign.

Within the past 3 to 5 years, survey researchers have developed methods for the respondent to complete a survey using a computerized self-administered procedure. The primary motivation behind this has been to reduce response inhibition and distortion. If respondents do not have to report sensitive information to an interviewer, they are more likely to report socially sensitive incidents. Computer-assisted self interviews (CASI) were first developed for use in self-report drug surveys and have been applied to a wide range of sensitive behaviors (e.g., same-sex sexual activity, abortion). An enhancement of this method has been to add an audio component, audio computer-assisted self interviews (ACASI), which reads the questions to the respondent. Respondents wear a set of headphones while following the questions on the computer screen. The audio component assists in overcoming possible literacy problems as well as enhances the privacy of the interview (e.g., respondents are free to blank out the screen and use only the audio).

Experimental research has found that ACASI leads to more reports of sensitive information when compared with an interviewer-administered instrument (Tourangeau and Smith 1996; Turner, Ku, and Sonenstein 1996). Respondents seem to feel more comfortable reporting sensitive behaviors when interviewers or other observers are not involved.¹²

Computerizing a self-administered instrument is particularly convenient for victimization surveys, given the relatively complex skip patterns and questionnaire structures. The need to first administer a screener and then follow up each incident mentioned with detailed questions (e.g., What happened? When did it happen? Who did it? Where did it happen?) makes it extremely difficult,

if not impossible, to use a paper and pencil self-administered form. The skip instructions are simply too difficult to communicate and implement. A computer takes care of this problem without complicating the respondent's task.

ACASI has not been widely implemented on victimization surveys. The exception is the British Crime Survey (Mayhew 1995; Percy and Mayhew 1997), which uses it to administer questions on sexual assault and domestic violence. Results, although not experimental, indicate a large increase in the reporting of these incidents. Application of this new methodology is likely to spread as it becomes more available.

Revising the series incident procedure

Dodge and Balog (1987) examined series incidents in NCS largely for the purpose of determining if classification of events as series incidents was due to interviewers' unwillingness to collect data on a large number of incidents from a given respondent. To do this, they first identified respondents who had initially reported series incidents and then reinterviewed these individuals using two separate surveys modified from NCS. Dodge and Balog found that in most series incidents with five or fewer events, the respondent could give details of the event if the interviewer asked for those details. This was not the case for the majority of the series incidents with six or more events. On the basis of this study, the Census Bureau changed the requirements for invoking the series incident procedure from situations in which three or more events were reported to those in which six or more events were reported. Moreover, the other information necessary for using the series procedure was explicitly built into the interview. Interviewers were required to ask or verify if (1) the events were similar and (2) the respondent could not report the details of each event. These changes followed closely the recommendations of the NAS panel and substantially reduced the number of events treated as a series incident in NCS. They also help reduce the effects of interviewer discretion on the identification of high-volume victimization. We can be more confident that events treated as series incidents are different from those events that are more discreet and distinguishable for the respondent. This moves us closer to being able to treat series incidents as a distinct form of victimization to be investigated rather than as measurement error.

The Census Bureau also introduced questions about the interrelationship of events in a series incident. They asked if the events involved the same offender or different offenders, if they occurred in the same place or in different places, and if the victimization had stopped or was continuing at the time of the interview. These few questions add a great deal of information about series incidents. It provides an idea of whether the victimization is a repeated

The magnitude of the effects of differing design features on victimization rates is so striking as to raise serious questions about the implications procedural variations may have for uses made of the survey data.

encounter with the same individuals or a much less particular event. This, in turn, provides valuable insight into the genesis of this continuous event (Lynch, Berbaum, and Planty 1998).

These changes in the series procedure are noteworthy because they mark a break with the point-in-time assumptions of NCS regarding crime events. In some cases, it may be more appropriate to consider crime as part of an ongoing event or condition such that one event can be a precipitant of another. This was not known until the survey collected information on the relationship between crime events. This is a small break with the emphasis on incident rate estimation and the assumption that crimes are best viewed as point-in-time events rather than events of continuous duration.

Limitations and Future Research

The institutionalization of victim surveys has encouraged their use in many debates of controversial policies. For example, victim survey data figured prominently in the debates about the Campus Crime Act, the Violence Against Women Act (Gilbert 1992; McPhail 1995; Murray, n.d.), and gun use (Kleck 1991, 1996; Cook 1985). This intense use of victim surveys has identified a number of longstanding methodological issues as well as raised new ones (see Fisher and Cullen in this volume). In addition, a number of longstanding issues were not resolved in the research and development work of the past two decades; they, too, should be addressed.

Controversies with the design and analysis of victimization surveys

Researchers' opinions differ on the importance survey design has for interpretation of victimization surveys. The problem of understanding the implications of survey methodology for analysis is not unique to victimization surveys. However, the magnitude of the effects of differing design features on victimization rates is so striking as to raise serious questions about the implications procedural variations may have for uses made of the survey data.

The differences between the BSSR and NORC surveys, the differences in NCS before and after 1992, and more recent surveys on violence show that self-reports

of victimization vary by a factor of 2 or greater, depending on the design features of the survey. Not only is this variation extremely large, but some of it is related to characteristics of the respondent or the event itself. An important controversy, and area of research, centers on the implications this variation has for conducting research with victimization surveys.

Several examples illustrate this point more clearly:

- NCS incurs great expense to use a bounded 6-month reference period because of extensive evidence of improvements in data quality that bounded, brief periods provide superior data. The original city surveys (Hindelang 1976; Hindelang, Gottfredson, and Garofalo 1978), the British Crime Survey, and most of the surveys on violence against women use an unbounded 12-month period. What are the implications of this for interpretation of analysis of these surveys? For example, many of the analyses of multiple victimization (see our earlier discussion) were based on unbounded reference periods. Does this increase estimates of multiple victimization artificially?
- As noted previously, when NCS changed methods in 1992, the level of crime jumped by 50 percent to 200 percent, depending on the type of crime. This is generally attributed to the change in screening methods and the use of CATI (Persley 1995). If one assumes the postredesign data are better, does this argue against ever using preresign data?
- Screening, the context of the questions, and automation (CATI and ACASI) have been shown to increase reports of sexual assault by factors of at least 2, depending on the domain of interest. Does this invalidate surveys that use methods where underreporting is the greatest?

The use of different methods both within and across surveys is unavoidable, given the costs associated with data collection. In many cases, the design feature that is considered better is also more expensive to implement than alternatives. This includes, for example, using shorter recall periods, using bounded recall periods and self-reports rather than proxy reports. Because of the expense of these “best” design features and the need to survey fairly large samples of people to yield statistically reliable analyses, only a survey of the magnitude of NCS can hope to institute many of these procedures. Even NCS, however, treats respondents nonuniformly (see the following discussion). Consequently, it is of both scientific and practical interest to understand what sacrifices survey planners and users may make in adopting particular designs or their products.

When judging alternative designs, it is important to keep in mind two basic analytic goals. One is to estimate the actual level of particular crimes. This, for example, has been the main controversy surrounding the surveys focusing on

violence against women. A second goal is to look at relationships among different variables to evaluate particular policies or examine year-to-year change.

When estimating the level of crime, it is readily evident that variations in methods provide vastly different answers to questions. Which method or sets of methods provides the best estimate of the particular concept of interest? We discuss in more detail in the following sections issues of data validity along with the research needed to clarify key issues related to understanding these wide variations for purposes of estimating the level of crime.

For more elaborate analyses (e.g., analysis of change over time, relationships among variables, and evaluation of policies), judging validity is more complicated. The critical question is not only which methods increase the validity of level estimates, but also whether different methods produce different substantive conclusions. A particular data collection method may be better at estimating the true level of crime (at least as defined by a particular study), but if measurement error is uncorrelated with the domain of interest, substantive conclusions may be unaffected. For example, one might argue that the use of longer recall periods is legitimate if measurement error is not correlated with critical relationships that may be of interest. If the primary relationship of interest, say, is between the victim's race and rate of victimization and the underreporting associated with longer reference periods does not vary by race, then analysis may be unaffected by the use of a longer period.¹³

We know that measurement error is correlated with a number of important characteristics related to victimization reporting. Studies have shown, for example, that race is related to underreporting (Czaja and Blair 1990), differential error by race is associated with the length of the recall period (Kobilarcik et al. 1983), and blacks underreport simple assaults (e.g., Skogan 1981, 30–31). Respondent event dating and definitional problems are correlated with the saliency of the event, at least as indicated by reverse record check studies.

Little research has been done on how these types of relationships vary by design feature. Are the underreporting patterns by race and/or education different when the screener is modified to encourage more complete reporting? When the reference period is shorter? When the instrument is self-administered? Answers to questions like these would enhance both the design and interpretation of results across surveys.

The use of differential methods also exist within particular surveys. It may be more convenient, for example, to conduct interviews by telephone and conduct in-person interviews for those persons who do not have telephones. Similarly, it may be more convenient to interview particular household members using a

proxy interview than to use self-reports. These treatment nonuniformities may be correlated with measurement error and the domain of interest. In this regard, NCS is, perhaps, the most egregious culprit. For example:

- The panel design of NCS does not allow for an initial bounding interview for persons who move into a sampled unit after its panel's first time in the sample (i.e., at waves 2–7). Thus, data from persons who move into a housing unit are combined with data from those who had been living in the house in the previous interview. As noted earlier, unbounded data produce more reports of victimization and those most likely to move have higher victimization rates (Biderman and Cantor 1984). It follows that NCS will overestimate the relationship between mobility (and its correlates) with victimization. In addition, as the percentage of unbounded households changes from year to year, there is potential that the yearly change estimates may also be affected (Biderman and Lynch 1991).
- A single member of the household is administered screening questions devoted to crimes against the household. This screener, however, reveals more crimes against individuals as well (Biderman, Cantor, and Reiss 1985). The selection of household informants is negatively correlated with victim risk (the household member who tends to stay home is most likely to be selected as the household informant). This depresses relationships associated with risk.
- The use of CATI on NCS is restricted to those who have telephones and who are willing to participate using this mode. Since CATI increases reporting of victimizations (Hubble and Wilder 1988) and its use is negatively correlated with risk, relationships examining risk factors are depressed.

Future research might further explore differences across methods as they impact substantive relationships. Research along these lines can be done in several ways. The most elaborate, and expensive, is through experiments, much like those described in the development of NCS (e.g., Kobilarcik et al. 1983) and by the work of Czaja and colleagues (1990, 1994). Treatments can be randomly assigned across respondents and results compared across treatments.

A second line of research would be to conduct identical analyses across data sets that vary systematically by design features. There has been little detailed comparison across datasets for key analyses (e.g., NCS versus BCS). Such research might illuminate how the different designs affect key relationships. Surveys, unfortunately, do not typically differ by only one feature. Consequently, these comparisons could not assign definitively the reason for any observed difference. They might, however, suggest the magnitude of the effects of particular combinations of design features.

The suggestion of more methodological research is complicated by the absence of good criteria for assessing the validity of the resulting data.

Validation

The suggestion of more methodological research is complicated by the absence of good criteria for assessing the validity of the resulting data. Reverse record check studies using police reports have been shown to be flawed conceptually (Biderman and Lynch 1981). One concern is that police records cover only events that are, by definition, not in the "gray area" the survey is meant to cover (i.e., crimes that do not come to the attention of the criminal justice system). This problem is illustrated when comparing the reverse record check studies (Kalish 1974; Czaja and Blair 1990) and the reference period experiment (Bushery 1981). They came to different conclusions partly because the former covers police events, whereas the latter does not. A second criticism of the reverse record check methodology is the difficulty of matching across the two mediums. Information in police records about the event may not be reported by the victim. Consequently, determining whether reports by victims match a police report are difficult to determine. In one study, for example, Miller and Groves (1985) demonstrated that the conclusion is influenced by what matching rules are applied.

Comparative studies of different survey procedures are useful alternatives to external data for validation purposes, but require assumptions in order to say something about validity. The most common assumption has been that there is more underreporting than overreporting and that, as a result, more is better. As retrospective surveys appropriately cast broader nets in search of eligible events and screening procedures become more sophisticated (as in the case of NCS), this assumption becomes less tenable.

The use of both reverse record check and comparative studies should still play some role in the development of new procedures. These provide external validation measures that normally cannot be obtained in survey research.¹⁴ However, as more aggressive and "broad net" screening techniques are employed in victim surveys, much more pressure must be put on incident forms to filter out ineligible events and to classify events deemed eligible for inclusion. These methods increase the number of events that fall into the gray area; that is, events for which questions arise about *content validity*. In the case of rape, for example, where the question of consent is extremely important, victims may indicate a lack of consent when the circumstances of the event indicate ambiguity in that area (e.g., prior intimacy, absence of force). As Biderman (1981c) noted, critical to understanding interpersonal harm is an accurate assessment of

what actually happened. This includes not only the circumstances of the event (who, what, when, where, how), but also something about the sequence of the events, and, possibly, the motivations of each of the actors.

The best way to address this issue is to collect the appropriate attributes related to the event and to use those attributes to construct a crime classification that reflects what happened. For example, if the purpose of the survey is to measure the number of crimes that occur in a population, an appropriate classification scheme can be developed similar to what is currently on NCS. Violent events, such as rape, can be classified by their typical components (e.g., forcible sexual penetration by a stranger). Those possessing all of these components can be put in one class, while events that have only two of the attributes would be put in another. They could all be classified as rape, or not, depending on the goal of the study. Furthermore, other attributes could be used to distinguish between degrees of certainty.

This is particularly important for events about which there is intense interest in the prevalence of the event but little consensus about definitions. The more complex classifications possible with attribute-based classification can prevent citing statistics for a large and heterogeneous class of events while claiming that all of the events in that class have the attributes of a much smaller and much more serious subset of these events. Loftin, Logan, and Addington (1999) are working on this type of classification scheme for hate crimes, and more work of this type should be done for other types of crime.

NCS currently has an extensive set of characteristics used to classify events. This is one of its strengths. However, these characteristics are geared primarily to the purpose of classifying events into official classes of crime. More research is needed in the development of incident forms to reflect both broader screening strategies and other uses of the data.

This can be done in several different ways. One way would be to conduct more qualitative analyses through collection of verbatim incident descriptions, focused respondent debriefings, or more intensive cognitive interviewing methods. More quantitative approaches would involve reinterview studies to examine test-retest reliability. The focus of these studies would be twofold. One would be to examine the attributes of events reported by different screening items. This would begin to provide evidence of content validity for reports using different screening strategies. A second focus would be to match the qualitative descriptions with the picture presented by the attribute-based classifications. The latter would provide some sense of the accuracy of the incident form in characterizing the event (e.g., motivations, interactional sequences, intent of victim). These would also provide the survey designer with a sense of the response processes that are used to formulate reports.

Continuations in the development of screening procedures

Parallel to the developments in screening for NCS, researchers interested in violence against women have developed screening methods based on many of the same principles. These methods have raised questions about the content of the survey instrument and methods that should be used to understand the types of events captured by a victimization survey. (For a more detailed description of these studies, especially as they compare with NCS, see Fisher and Cullen in this volume.)

The approach used by a number of violence researchers (e.g., the Conflict Tactics Scale) is to rely on extensive, and quite explicit, cues that narrowly focus on violent events, including sexual violence (Strauss et al. 1995). In addition to its narrow focus and explicit cues, this approach sets a different context than NCS. This is done by the use of a different type of introductory statement. Family conflict studies (Strauss 1998) set up the survey as one concerned with family or marital problems. As noted previously, this contrasts with the design of NCS, which frames the survey around concerns with crime. A second variant of this approach is surveys on personal safety (Tjaden and Thoennes 1998). This strategy directly addresses problems of failures of concept by not asking the respondent to make a value judgment about whether the incident is a criminal event.

Studies using this strategy have also adopted a quite different approach from NCS when sampling and interviewing household members to minimize problems with response inhibition and distortion. These studies (Tjaden and Thoennes 1998; Koss 1996; Kilpatrick et al. 1987) typically interview one person in a household. This is done, in part, to prevent others in the household from knowing what is actually on the questionnaire (e.g., if an abuser knows what is on the questionnaire, it may endanger the victim). Before administering the questions, the interviewer makes sure the respondent is in a private room where no one can overhear the conversation. It is made clear to the respondent that if someone walks in during the conversation, the interview will be continued at another time. This might include, for example, abruptly ending a telephone interview if the respondent feels it is necessary to do so.

The family conflict and personal safety surveys find extremely high rates of violence, especially rates of violence against women by nonstrangers. Strauss (1998), for example, presents a table that shows family conflict studies finding an average rate of violence among family members to be 16 percent for family conflict studies, 2 percent for a personal safety study (Tjaden and Thoennes 1998), and 0.9 percent for NCS.

The differences in the estimates appear to be functions of both the context of the items and the cueing used. As previously noted, NCS explicitly sets the context of the interview as one concerned with crime, whereas the conflict and safety studies explicitly avoid the use of any legal connotations (Strauss 1998, 3). Whether the event is considered a crime by the respondent is irrelevant. One example of this is a question used from the Sexual Experiences Survey (SES) (Koss 1993), which asks: "Have you had sexual intercourse when you didn't want to because a man gave you alcohol or drugs?"

This contrasts to the question on NCS that asks about sexual assault by a nonstranger:

People don't often think of incidents committed by someone they know. Did you have something stolen from you OR were you attacked or threatened by: (a) someone at work or school, (b) a neighbor or friend, (c) a relative or family member, (d) any other person you've met or know?

SES asks about a situation that ignores any criminal intent, it simply asks about unwanted sex that was preceded by using alcohol or drugs. NCS asks about "incidents committed" within the context of being a victim of a crime.

Perhaps just as important as the context of the questions, the conflict and safety studies focus cues exclusively on violence, especially among nonstrangers. NCS screens on all types of crimes and only has one or two questions (with multiple cues) that specifically target (domestic) violence. The higher density of more specific cues will lead to reports of more events, as shown by the studies referenced earlier.

The discrepancies between these two approaches pose both conceptual and methodological challenges. Biderman (1981b, 1981c) makes the distinction between an indicator of crime and that of interpersonal harm (also see the discussion by Skogan 1981, 9–10). The former implicitly relies on the judgments of the respondent to report details about the culpability of the offender in the event. The latter does not, at least when initially asking the respondent to report the event. As Biderman (1981c, 49) notes, by restricting attention to events that are crimes, the survey may be leaving out events that are critically important for understanding the causes and consequences of interpersonal harm:

Victims apply their own conceptions of whether the act indeed was "criminal," whether it should be made a matter for official attention, and whether the official system would be likely to act sufficiently in accordance with the victim's view and desires were a complaint made. . . . These grounds for excluding events from the criminal justice process include all of the

classes of judgement that are the central objectives of victimology. . . . Victimological research that is based exclusively on officially recorded offenses thereby may be excluding most of the social phenomena with which it is particularly concerned.

As a measure of crime, however, the use of an indicator of interpersonal harm leaves out the formal criteria that separate events as criminal from others, such as accidents, legitimate retaliation, or other explanations that are not criminal.

The approaches taken by studies on violence against women have elaborated on the harm approach illustrated in the early work described by Biderman (1981b). They ask about actions and consequences without any reference to criteria related to criminal events. Many of these studies have not, however, taken the additional step of then asking details about the event to “establish sound actuarial knowledge of the magnitude of hazards various types of social situations present. [To do this] the data employed should be phenomenologically comprehensive and phenomenologically analyzable” (Biderman 1981c, 51). Further research needs to begin to move in this direction in order to begin conceptually relating harm to what society (and victims of harm) conventionally view as crimes.

While the context of the survey is important in defining the scope of eligible events for the respondent, the cues presented also serve this function, as well as influence the process of locating specific events in memory. Intensive cuing of particular types of events should yield more reports of these events, as illustrated by the NCS experiments, as well as the violence studies. This implies that reporting events will reflect, in part, the distribution of cues in addition to the distribution of crime events. Further evidence of this can be seen in a comparison of NCS and the National Violence Against Women Study. The latter cued extensively for rape and for crimes among intimates. This resulted in rates of physical assault that were generally higher in NCS, with the exception of rape and assaults by intimates (Bachman 1998). We know little about how cues interact with the survey context (e.g., harm versus crime). This requires more research into the effects of cues and how they should be allocated given the purpose of a particular survey instrument.

Sample design, coverage, and nonresponse

A number of issues related to the sample design, coverage, and nonresponse continue to be problematic for victimization surveys. These include developing efficient sample designs, improving coverage, and nonresponse imputation for groups at high risk of victimization.

Developing efficient sample designs

Conducting a victimization survey is an expensive undertaking. Because a relatively small percentage of people will report an event for a fixed time period, large sample sizes are needed to generate reliable population estimates. Some reduction in reliability can be compensated for by lengthening the reference period and improving the screening methods. However, both of these have their limitations and costs (as noted earlier). The former increases memory error, and the latter complicates the design and detail required for the instrumentation (and time needed to design and administer the instrument).

One of the major innovations over the past 10 to 15 years has been to increase the number of surveys that are done by telephone using samples generated by random digit dialing methods. By dispensing with expensive area-based, in-person designs, the project can increase the number of interviews per dollar spent. Many of the surveys referenced that have examined violence and the enduring effects of victimization have been conducted using this method. The disadvantage of this method is that it typically yields relatively low response rates (in the 60- to 70-percent range) and misses the population that does not have a telephone.

Yet to be fully exploited are less traditional methods, especially those using networks of victims. Network designs are based on the idea of using respondents as informants on other persons in the network. The respondent is asked to provide information on whether other persons that the respondent knows have been victims of a crime. If the information is accurate, and one can precisely enumerate the counting rules involved, then it is possible to develop estimates of victimization. Czaja and Blair (1990) conducted an evaluation of this method and did not find the network methods they employed to be better than a traditional approach using a mean square error criterion. However, they noted a number of problems with their design and recommended further research into this type of sampling process.

Other methods for reducing the expense involved in these surveys have been suggested, including the use of other types of multiplicity estimators. One variant of this logic is to use telephone prefixes or area-locations to find victims. Oversampling in areas yielding a large number of reports of victimization, using this logic, may produce more efficient sample designs. Of course the overall efficiency and utility of any such strategies depend on the goals of the survey (e.g., estimating population rates versus comparison across different subpopulations). Nonetheless, using some type of stratification or double sampling process needs to be explored when trying to reduce the costs of victimization surveys. If this were done, the utility of this method of evaluation and analysis would be greatly increased.

Nonresponse and coverage

A persistent observation in victimization surveys is relatively equal simple assault rates between whites and other minority groups. One possible explanation for this is differential error related to response problems such as comprehension and recall (Skogan 1981). Another explanation is differential coverage and nonresponse. There is evidence, in fact, that both nonresponse and coverage must be taken into account for victimization surveys (Reiss 1977; Biderman and Cantor 1984; Griffin Saphire 1984; Stasny 1991). These issues may be particularly problematic for telephone surveys, which are increasingly being used because of the economies they offer. With respect to coverage, research with NCS has shown that persons who do not own telephones are the most likely to be victimized (Woltman, Turner, and Bushery 1980). With respect to nonresponse, telephone surveys generally achieve response rates that are 15 to 20 percent lower than in-person surveys and that may be particularly vulnerable to issues of bias, especially for certain subgroups.

Even for in-person surveys, however, there is evidence that coverage and nonresponse biases are problematic. Particular problems have been found for certain minority groups, especially Hispanics and young black males. The hypothesis is that the surveys simply miss those who are most likely to be subject to crime. Using longitudinal data from NCS, Reiss (1977) found that persons who have high residential mobility have much higher victimization rates than those who are not mobile (see also Biderman and Cantor 1984). Further elaboration taking advantage of this correlation has found mobility to be an important covariate when imputing data (Griffin Saphire 1984; Stasny 1991). For cross-sectional analysis, it may be possible to adapt information from reports of mobility for respondents (e.g., how many times a respondent moved in the last year). Future research should elaborate on this correlation, as well as developing other indicators for uses in imputation.

The correlation with mobility should also be viewed as a proxy for coverage problems. Persons who are most likely to be missed are, in part, those persons with unstable living situations who either may not have a residence at any particular point in time or may not be considered part of the residence when the interviewer conducts the initial household enumeration (Martin 1996). Indirect evidence of coverage problems on NCS was found by Cook (1985), who compared estimates of gunshot victims with external records available from hospital emergency rooms. These data seemed to indicate a gross undercount of such injuries on NCS. One leading explanation is that NCS misses those individuals who are most likely to be a victim of this type of crime.

Going beyond the assumption of crime as a point-in-time event

One of the weaknesses of the victim survey method is its emphasis on crime as incidents occurring at a point in time. This approach to crime stems from the carryover from the attempts to have the survey mimic UCR. Many of the types of crimes that have the greatest social (as opposed to individual) import are more readily approached as conditions that endure rather than incidents that begin and end at a given point in time (Biderman 1975). Among the kinds of victimizations that could be conceived and measured in prevalence rather than incident terms are various forms of continuing terrorization and extortion. This might include, for example, a spouse or sexual partner in continuous fear of violence or school children who must routinely give up their lunch money to gangs of fellow students. Here the victim is in virtually a continuous state of threat and victimization, but the survey requires that this condition be divided into its component parts, which minimizes the disruption of life and social relations.

To some degree, victimization surveys provide information about these kinds of situations through tabulations of series victimizations. Historically in NCS these were defined as three or more similar incidents of victimization mentioned by a respondent, but which, because of their frequency or similarity, the respondent cannot individually date or differentiate from one another. The terrorized spouse then could be identified in NCS through repeated incidence of spouse beating or the terrorized school child by repeated robberies. It is not necessary for the specific acts defining victimization to exist for there to be continuous victimization. To make a threat credible to a victim and to continue the state of terrorization, the offender need not continually repeat his threat or actually inflict violence.

Even the series victimization is captive to the point-in-time logic in that most of the questioning regarding series incidents is done for the purpose of counting incidents and not for the purpose of establishing duration or patterns of events. It may be more useful for understanding conditions of continuous victimization to have respondents explain the interrelationship of events in a series or to talk about the factors that are contributing to persistence. One logical way to do this is with a longitudinal design. Directly asking respondents to draw linkages between events, if there are any, would certainly be useful in identifying who is in the condition of continuous victimization.

Longitudinal surveys are expensive and time consuming to complete. Linking events within the same reference period would provide a significant advance and would not be as expensive to implement. Recent changes to the series

victimization procedures in NCS have moved in this direction by asking respondents if all of the events in the series involved the same offender, if they occurred in the same place, and if the victimization is continuing. Unfortunately, these questions are asked only of respondents who satisfy the conditions of the series victimization procedure. It would be better if some provision were made for asking about the interrelationship of events in all instances of repeated victimization within a given interview period.

Another facet of victimization that might be better suited to a prevalence rather than an incidence approach to measurement is the durable and psychic consequences of point-in-time crime events. Here questioning that elicited the initiation and termination of conditions resulting from a crime event would be useful. Also, in surveys like NCS that involve more than one interview, asking about the conditions across interviews would be helpful in establishing the persistence of these conditions.

Conclusions

Self-report surveys of victimization have become an established feature of crime statistics in the United States and throughout the world. They are used routinely as social indicators and as tools for building criminological theory. Over the past 25 years, we have learned a great deal about asking persons to recount their victimization experiences. Much of that knowledge has come from the National Crime Survey and its antecedents. The process of selling, planning, and fielding the first surveys framed and informed many of the issues that needed to be resolved to conduct a household survey of victims. During the initial development there was a tension between a legalistic emphasis and one oriented to more traditional survey design concerns. Survey methodologists found that it was better to organize the survey to facilitate the recall and reporting task. Those interested in comparisons with police data were concerned with developing a social indicator comparable to police-reported measures of crime. Elements of both approaches to designing victimization surveys were retained by the Census Bureau and DOJ when they began fielding the ongoing NCS. However, a heavy emphasis was put on comparisons to the police record systems.

The subsequent methodological work moved NCS closer to an approach based on facilitating recall. Moreover, changes in the NCS design and the appearance of other victimization surveys have reiterated the lesson of the pilot studies: Data from victim surveys are heavily influenced by their design. The appearance of these alternative designs and the very large differences in reporting that resulted give us both the motivation and the ability to learn much more about the method. In the future, it will be critical to compare variations in design with

differences in reporting to better understand the implications of the methodology (Lynch 1996). Unless this additional research and development work is done, the substantial effect of design on the resulting data will raise suspicions about whether results are less a reflection of the crime problem than they are of the design of the survey. If this occurs, then the widespread acceptance of this method may decline.

This paper and the authors have benefited immeasurably from discussions with and the writings of Albert D. Biderman.

Notes

1. There are a number of excellent reviews of various aspects of the design and contribution of victimization surveys. For more detailed discussion of specific topics, see publications by Gottfredson (1986), Sparks (1981), Hindelang (1976), and Skogan (1981).

2. From 1973 to 1991, the survey was called the National Crime Survey. Since 1991, it has been referred to as the National Crime Victimization Survey. Because we often refer to the survey throughout its existence, we use National Crime Survey (NCS) throughout the text.

3. For a more complete discussion of UCR's organization and procedures, see *The Uniform Crime Reporting Handbook* (U.S. Department of Justice, Federal Bureau of Investigation 1984). For a discussion of the implications of these aspects of UCR for the quality of the resulting data, see Biderman and Lynch (1991).

4. For a more complete description of the design, see Rand and Taylor (1995).

5. For a detailed review of the early development of NCS, see Hindelang (1976, 21–76).

6. The size of the sample required will be affected by the productivity of the screening interview. The less productive the screening interview, the greater the projected sample size for the same level of precision. Although the more evocative screening procedures used in the BSSR pilot study may have reduced the sample size required, even in this case, extremely large samples would be required to estimate rare crimes like robbery and rape with any precision.

7. Several additional factors led to adopting a 6-month period. One was related to the observation that shorter time periods led to more accurately dated events. Because NCS relies on this dating to determine which incidents are within a reference year, this type of error had to be minimized. The second related to the timing of data releases. With a 12-month period, an additional 6 months would have to elapse (compared with a 6-month period) because of the need to interview all persons who could possibly report a crime within the appropriate calendar year. For example, for a 6-month period, interviews needed to generate an estimate for year t would have to wait until interviews in June of year $t+1$ are finished. For a 12-month period, one would have to wait until November interviews in year $t+1$ are completed.

8. At the time these changes were made, the National Crime Survey was renamed the National Crime Victimization Survey.
9. RPE was conducted with bounded reference periods. This should minimize this type of telescoping. Nonetheless, the assumption that more is better as a measure of improved quality may overestimate the true gains in quality achieved by shortening the reference period.
10. A shorter reference period would result in covering less of the calendar period in each interview (3 months rather than 6 months). This cuts the sample size by a significant proportion.
11. It is less clear how other types of events may be affected. For example, the presence of other household members may actually encourage reporting if the other household members actually know about the event.
12. It should be noted that the evidence related to self-administration does not link improvements in reporting to external validation criteria (e.g., biological tests, arrest records).
13. This perspective is somewhat simplistic. Methods have effects on not only the direction of relationships (bias), but also reliability and statistical power. For example, making concepts clearer or using a self- rather than a proxy-respondent may reduce sampling error by eliminating variation due to misunderstandings or faulty knowledge.
14. Other external criteria that should be considered, especially when using more broad-net approaches, could be records that capture more general sets of injuries or incidents. This might include, for example, records from emergency rooms, hospital records, and insurance claims.

References

- Aldrich, Howard, and Albert J. Reiss. 1970. The effects of civil disorder on small business in the inner city. *Journal of Social Issues* 26:187–206.
- Bachman, Ronet. 1998. Comparing estimates of violence against women from the National Survey of Violence Against Women and the National Crime Victimization Survey. Paper presented at the National Conference on Violence Against Women, 18 October, Arlington, Virginia.
- Biderman, Albert D. 1981a. Notes on the methodological development of the National Crime Victimization Survey. In *The National Crime Survey: Working papers, volume 1: Current and historical perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. NCJ 75374. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- . 1981b. A social indicator of interpersonal harm. In *The National Crime Survey: Working papers, volume 1: Current and historical perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. NCJ 75374. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

- . 1981c. When does interpersonal violence become crime?—Theory and practice. In *The National Crime Survey: Working papers, volume 1: Current and historical perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. NCJ 75374. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- . 1975. *Notes on the significance of measurements of events and of conditions by criminal victimization surveys*. BSSR 0003–58. Washington, D.C.: Bureau of Social Science Research.
- . 1967. Surveys of population samples for estimating crime incidence. *Annals of the American Academy of Social and Political Science* 374 (November): 16–34.
- . 1966. *Report on design of a national study*. Washington, D.C.: Bureau of Social Science Research, Inc.
- Biderman, Albert D., and David Cantor. 1984. A longitudinal analysis of bounding respondent conditioning and mobility as sources of panel bias in the National Crime Survey. In *American Statistical Association 1984 proceedings of the Section on Survey Research Methods*. Washington, D.C.: American Statistical Association.
- Biderman, Albert D., David Cantor, James P. Lynch, and Elizabeth Martin. 1986. *Final report of the National Crime Survey redesign*. Washington, D.C.: Bureau of Social Science Research.
- Biderman, Albert D., David Cantor, and Albert J. Reiss, Jr. 1985. A quasi-experimental analysis of personal victimization by household respondents in the NCS. Paper presented at the Annual Meetings of the American Statistical Association, Philadelphia.
- Biderman, A.D., L.A. Johnson, J. McIntyre, and A.W. Weir. 1967. *Report on a pilot study in the District of Columbia on victimization and attitudes toward law enforcement*. President's Commission on Law Enforcement and Administration of Justice, Field Surveys no. 1. Washington, D.C.: U.S. Government Printing Office.
- Biderman, Albert D., and James P. Lynch. 1981. Recency bias in data on self-reported victimization. *American Statistical Association 1981 proceedings of the Social Statistics Section*. Washington, D.C.: American Statistical Association.
- Biderman, Albert D., and James P. Lynch, with James Peterson. 1991. *Understanding crime incidence statistics: Why the UCR diverges from the NCS*. New York: Springer-Verlag.
- Biderman, Albert D., and Albert J. Reiss, Jr. 1967. On exploring the “dark figure” of crime. *Annals of the American Academy of Political and Social Science* 374 (November): 1–15.
- Bushery, J. 1981. Recall bias for different reference periods in the National Crime Survey. Paper presented at the Annual Meetings of the American Statistical Association, Boston.

- Cantor, David. 1985. Operational and substantive differences in changing the NCS reference period. In *American Statistical Association 1985 proceedings of the Social Statistics Section*. Washington, D.C.: American Statistical Association.
- Cohen, Lawrence, and David Cantor. 1981. Residential burglary in the United States: Lifestyle and demographic factors associated with the probability of victimization. *Journal of Research in Crime and Delinquency* 18:113–27.
- Cohen, Lawrence, and Marcus Felson. 1979. Social change and crime rate trends: A routine activity approach. *American Sociological Review* 44:588–608.
- Coker, Ann, and Elizabeth Stasny. 1996. *Adjusting the National Crime Victimization Survey's estimates of rape and domestic violence for "gag" factors*. NCJ 173061. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.
- Conway, Mark, and Sharon Lohr. 1994. A longitudinal analysis of factors associated with the reporting of violent crime to the police. *Journal of Quantitative Criminology* 10:23–39.
- Cook, Philip. 1985. The case of the missing victim: Gunshot wounding in the National Crime Survey. *Journal of Quantitative Criminology* 1:91–102.
- Couper, M.P., R.P. Baker, J. Bethlehem, C.Z.F. Clark, W.L. Nicholls, and J. O'Reilly, eds. 1998. *Computer assisted survey information collection*. New York: John Wiley & Sons.
- Czaja, R., and J. Blair. 1990. Using network sampling in crime victimization surveys. *Journal of Quantitative Criminology* 6:185–206.
- Czaja, R., J. Blair, B. Bickart, and E. Eastman. 1994. Respondent strategies for recall of crime victimization incidents. *Journal of Official Statistics* 10:257–276.
- Dodge, R.W., and F.D. Balog. 1987. *Series crimes: Report of a field test*. Technical Report, NCJ 104615. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Dodge, R.W., and A.G. Turner. 1981. Methodological foundations for establishing a national survey of victimization. In *The National Crime Survey: Working papers, volume 1: Current and historical perspectives*, edited by Robert G. Lehnen and Wesley G. Skogan. NCJ 75374. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Ellingsworth, Daniel, Graham Farrel, and Kenneth Pease. 1995. A victim is a victim is a victim: Chronic victimization in four sweeps of the British Crime Survey. *British Journal of Criminology* 35:360–365.
- Ennis, Philip H. 1967. *Criminal victimization in the United States: A report of a national survey*. President's Commission on Law Enforcement and Administration of Justice, Field Surveys no. 2. Washington, D.C.: U.S. Government Printing Office.

Finkelhor, David. 1997. The victimization of children and youth: Developmental victimology. In *Victims of crime*, edited by Robert Davis, Arthur Lurigio, and Wesley Skogan. 2d ed. Thousand Oaks, California: Sage Publications.

Forrester, D., M. Chatterton, and K. Pease. 1988. *The Kirkholt burglary prevention project*. Rochdale Crime Prevention Unit Paper no. 13. London: Home Office.

Garofalo, James. 1977. *The police and public opinion: An analysis of victimization and attitude data in 13 American cities*. NCJ 42018. Washington, D.C.: U.S. Department of Justice, National Criminal Justice Information and Statistics Service.

Gidycz, Christine, and Mary P. Koss. 1991. Predictors of long-term sexual assault trauma among a national sample of victimized college women. *Violence and Victims* 6:175–90.

Gilbert, Neil. 1992. Realities and mythologies of rape. *Society* 29 (4): 4–10.

Gottfredson, Michael. 1986. Victimization surveys. In *Crime and justice: An annual review of research*, edited by Michael Tonry and Norval Morris. Vol. 7. Chicago: University of Chicago Press.

———. 1984. *Victims of crime: The dimensions of risk*. Home Office Research and Planning Unit Report no. 81. London: Her Majesty's Stationery Office.

Griffin Saphire, Diane. 1984. *Estimation of victimization prevalence using data from the National Crime Survey*. New York: Springer-Verlag.

Harlow, C.W. 1989. *Injuries from crime*. Special Report, NCJ 116811. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Hindelang, Michael. 1981. Variations in sex-race-age-specific incidence rates of offending. *American Sociological Review* 46:461–474.

———. 1978. Race and involvement in common law personal crimes. *American Sociological Review* 43:93–109.

———. 1976. *Criminal victimization in eight American cities: An analysis of common theft and assault*. Cambridge, Massachusetts: Ballinger.

Hindelang, Michael, Michael Gottfredson, and James Garofalo. 1978. *Victims of personal crime: An empirical foundation for a theory of personal victimization*. Cambridge, Massachusetts: Ballinger.

Hough, Michael. 1987. Offender's choice of target: Findings from victim surveys. *Journal of Quantitative Criminology* 3 (4): 275–281.

Hubble, David, and B.E. Wilder. 1988. Preliminary results from the National Crime Survey CATI experiment. In *American Statistical Association 1988 proceedings of the Section on Survey Research Methods*. Washington, D.C.: American Statistical Association.

- Kalish, Carol. 1974. *Crimes and victims: A report on the Dayton-San Jose pilot survey of victimization*. Research Report, NCJ 13314. Washington, D.C.: U.S. Department of Justice, National Criminal Justice Information and Statistics Service.
- Kennedy, Leslie, and David Forde. 1999. *When push comes to shove: A routine conflict approach to violence*. Albany: State University of New York Press.
- Kilpatrick, D.G., B.E. Saunders, L.J. Veronen, C.L. Best, and J.M. Von. 1987. Criminal victimization: Lifetime prevalence, reporting to police, and psychological impact. *Crime & Delinquency* 33:479–489.
- Kindermann, Charles, James P. Lynch, and David Cantor. 1997. *The effects of the redesign on victimization estimates*. NCJ 164381. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Kleck, Gary. 1996. Crime, culture conflict, and the sources of support for gun control: A multilevel application of the general social surveys. *American Behavioral Scientist* 39 (4): 387–404.
- . 1991. *Point blank: Guns and violence in America*. New York: Aldine de Gruyter.
- Kobilarcik, E.L., C.A. Alexander, R.P. Singh, and G.M. Shapiro. 1983. Alternative reference periods for the National Crime Survey. In *American Statistical Association 1983 proceedings of the Section on Survey Research Methods*. Washington, D.C.: American Statistical Association.
- Koss, Mary P. 1996. The measurement of rape victimization in crime surveys. *Criminal Justice and Behavior* 23:55–69.
- . 1993. Detecting the scope of rape: A review of prevalence research methods. *Journal of Interpersonal Violence* 8 (2): 198–222.
- Kury, Helmut. 1993. Crime in East and West Germany: Results from the First Intra-German Victims' Survey. In *Fear of crime and criminal victimization*, edited by Wolfgang Bilsky, Christian Pfeiffer, and Peter Wetzels. Stuttgart: Enke Verlag.
- Lauritsen, Janet L., and K.F. Davis Quinnet. 1995. Repeat victimization among adolescents and young adults. *Journal of Quantitative Criminology* 11:143–166.
- Lauritsen, J.L., R.J. Sampson, and J. Laub. 1991. The link between offending and victimization among adolescents. *Criminology* 29:265–292.
- Lehnen, Robert G., and Wesley G. Skogan. 1981. *The National Crime Survey: Working papers, volume 1: Current and historical perspectives*. NCJ 75374. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Lepkowski, James. 1981. *Sample design issues from the National Crime Survey*. Ann Arbor: University of Michigan, Survey Research Center.

Loftin, Colin, Wayne A. Logan, and Lynn Addington. 1999. A proposed binary-based evaluative framework for the categorization of hate crimes. Paper presented at the Conference on Hate Crime Measurement, 23 April, School of Criminal Justice, University at Albany, State University of New York.

Lynch, James P. 1996. Clarifying divergent estimates of rape from two national surveys. *Public Opinion Quarterly* 60 (3): 410–29.

Lynch, James P., Michael Berbaum, and Mike Planty. 1998. *Investigating repeated victimization with the NCVS*. Final report, NIJ grant no. 97–IJ–CX–0027. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Lynch, James P., and David Cantor. 1992. Ecological and behavioral influences on property victimization at home: Implications for opportunity theory. *Journal of Research in Crime and Delinquency* 29 (August): 335–362.

Martin, Elizabeth. 1996. Household attachment and survey coverage. Paper presented at the Joint Statistical Meetings of the American Statistical Association, 4–8 August, Chicago.

Maxfield, Michael. 1987. Household composition, routine activity, and victimization: A comparative analysis. *Journal of Quantitative Criminology* 3:301–320.

Mayhew, Pat. 1995. Some methodological issues in crime victimization surveys. In *Crime victimization surveys in Australia—Conference proceedings*. Brisbane: Criminal Justice Commission.

———. 1993. Reporting crimes to the police: The contributions of victim surveys. In *Fear of crime and criminal victimization*, edited by Wolfgang Bilsky, Christian Pfeiffer, and Peter Wetzels. Stuttgart: Enke Verlag.

McDowall, D., and C. Loftin. 1992. Comparing the UCR and NCS over time. *Criminology* 30:125–132.

McPhail, Beverly. 1995. The term is “femicide” and society must do something. *Houston Chronicle*, 1 October.

Miethe, T.D., and R.F. Meier. 1994. *Crime and its social context: Toward an integrated theory of offenders, victims, and situations*. Albany: State University of New York Press.

Miller, Peter V., and Robert M. Groves. 1985. Matching survey responses to official records: Validity in victimization reporting. *Public Opinion Quarterly* 49:366–380.

Murray, David. N.d. *America’s phony rape epidemic*. Washington, D.C.: Statistical Assessment Service.

National Commission on Law Observance and Enforcement. 1931. *Police*. Washington, D.C.: U.S. Government Printing Office.

- Nelson, James F. 1980. Multiple victimization in American cities: A statistical analysis of rare events. *American Journal of Sociology* 85:871–891.
- Neter, John, and Joseph Waksberg. 1964. Conditioning effects from repeated household interviews. *Journal of Marketing* 29:51–56.
- New York Times*. 1981. Editorial, The conflict of UCR–NCS statistics, 19 April.
- Norris, Fran H., Krzysztof Kaniasty, and Martie P. Thompson. 1997. The psychological consequences of crime: Findings from a longitudinal population-based study. In *Victims of crime*. 2d ed. Edited by Robert Davis, Arthur Lurigio, and Wesley Skogan. Thousand Oaks, California: Sage Publications.
- Parkinson, R.P., A.L. Paez, and N.F. Howard. 1977. *Criminal victimization in the United States, 1974: A National Crime Survey report*. NCJ 39467. Washington, D.C.: U.S. Department of Justice, National Criminal Justice Information and Statistics Service.
- Pease, Ken. 1998. *Repeat victimization: Taking stock*. Crime Detection and Prevention Series. Paper 90. London: Home Office, Police Research Group.
- Penick, B.K.E., and M. Owens. 1976. *Surveying crime*. Washington, D.C.: National Academy Press.
- Percy, A., and Pat Mayhew. 1997. Estimating sexual victimization in a national crime survey: A new approach. *Studies on Crime & Crime Prevention* 6 (2): 125–50.
- Persley, Carolyn. 1995. The National Crime Victimization Survey redesign: Measuring the impact of new methods. In *American Statistical Association 1995 proceedings of the Section on Survey Research Methods*. Washington, D.C.: American Statistical Association.
- Poggio, E.C., S.D. Kennedy, J.M. Chaiken, and K.E. Carlson. 1985. *Blueprint for the future of the Uniform Crime Reporting program: Final report of the UCR study*. NCJ 98348. Washington, D.C.: U.S. Department of Justice, Federal Bureau of Investigation and Bureau of Justice Statistics.
- Polvi, N., Terah Looman, Charles Humphries, and Ken Pease. 1990. Repeat break-and-enter victimization: Time course and crime prevention opportunity. *Journal of Police Science and Administration* 17:8–11.
- Rand, Michael, and Bruce Taylor. 1995. The National Crime Victimization Survey redesign: New understandings of victimization dynamics and measurement. Paper presented at the Joint Statistical Meetings of the American Statistical Association, 13–17 August, Orlando, Florida.
- Reiss, Albert J. 1977. *Victim proneness by type of crime in repeat victimization*. New Haven, Connecticut: Institute for Social Policy Studies.

———. 1969. Impact of crime on small business. U.S. Senate. Select Committee on Small Business. *Hearings before the Select Committee on Small Business*. 91st Cong., 1st sess. 21–23 May and 22 July.

———. 1967a. Impact of crime on small business. U.S. Senate. Select Committee on Small Business. *Hearings before the Select Committee on Small Business*. 90th Cong., 1st sess. 24–26 April.

———. 1967b. *Measurement of the nature and the amount of crime: Studies in crime and law enforcement in major metropolitan areas*. President's Commission on Law Enforcement and Administration of Justice, vol. 1 of Field Surveys no. 3. Washington, D.C.: U.S. Government Printing Office.

Resnick, Heidi S., D. Kilpatrick, B. Dansky, B.E. Saunders, and Connie Best. 1993. Prevalence of civilian trauma and post traumatic stress disorder in a representative sample of women. *Journal of Consulting and Clinical Psychology* 61 (6): 984–999.

Singer, S.I. 1986. Victims of serious violence and their criminal behavior: Subcultural theory and beyond. *Violence and Victims* 1:61–70.

———. 1981. Homogeneous victim-offender populations: A review and some research implications. *Journal of Criminal Law and Criminology* 72 (2): 779–788.

Skogan, Wesley G. 1984. Reporting crimes to the police: The status of world research. *Journal of Research in Crime and Delinquency* 21:113–137.

———. 1981. *Issues in the measurement of victimization*. NCJ 74682. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

———. 1974. The validity of official crime statistics. *Social Science Quarterly* 55:35–48.

Sparks, Richard F. 1981. Multiple victimization: Evidence theory and future research. *Journal of Criminal Law and Criminology* 72:762–778.

Stasny, Elizabeth A. 1991. Hierarchical models for survey and nonresponse probabilities. *Journal of the American Statistical Association* 86:296–303.

Strauss, M.A. 1998. The controversy over domestic violence by women: A methodological, theoretical, and sociology of science analysis. In *Violence in intimate relationships*, edited by X.B. Arriaga and S. Oskamp. Thousand Oaks, California: Sage Publications.

Strauss, M.A., S.L. Hamby, S. Boney-McCoy, and D.B. Sugarman. 1995. *The Revised Conflict Tactics Scales (CTS2-form A)*. Durham, New Hampshire: University of New Hampshire, Family Research Laboratory.

Tjaden, P., and N. Thoennes. 1998. *Stalking in America: Findings from the National Violence Against Women Survey*. Research in Brief, NCJ 169592. Washington, D.C.: U.S. Department of Justice, National Institute of Justice, and U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Tourangeau, R., and T. Smith. 1996. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60:275–304.

Turner, C.F., L. Ku, and F.L. Sonenstein. 1996. Impact of ACASI on reporting male-male sexual contacts: Preliminary results from the 1995 Survey of Adolescent Males. In *Health survey research methods*, edited by R. Warnecke. Hyattsville, Maryland: U.S. Department of Health and Human Services, National Center for Health Statistics.

U.S. Department of Justice, Federal Bureau of Investigation. 1984. *Uniform Crime Reporting handbook*. Washington, D.C.

U.S. Department of Justice, National Criminal Justice Information and Statistics Service. 1976. *Criminal victimization in the United States 1973*. NCJ 34732. Washington, D.C.

U.S. House. 1977. Committee on the Judiciary. *Suspension of the National Crime Survey: Hearing before the Subcommittee on Crime*. 95th Cong., 1st sess., 13 October.

Washington Post. 1981. Editorial, Two reports on violent crime in the U.S., 5 May.

Woltman, Henry F., Anthony Turner, and John M. Bushery. 1980. A comparison of three mixed mode interviewing procedures in the National Crime Survey. *Journal of the American Statistical Association* 75 (371): 534–543.

Zawitz, Marianne W., ed. 1988. *Report to the Nation on crime and justice*. 2d ed. NCJ 111096. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Theory, Method, and Data in Comparative Criminology

by Gregory J. Howard, Graeme Newman, and William Alex Pridemore

The goal of this chapter is to provide an overview of the recently rejuvenated field of comparative criminology. It begins by considering the context and history of comparative criminology and continues by outlining the contemporary comparative perspective. After identifying several goals for comparative criminology that are often advanced, including theory elaboration and testing as well as policy evaluation and critique, the chapter describes the common approaches to comparative criminological research. The main theoretical traditions of comparative criminology are examined first, with particular attention directed to metanarratives such as modernization, civilization, opportunity, and world system theories and to structural theories based on culture, social bonds, and the distribution of economic resources. Taking up methodological concerns next, the chapter summarizes some of the more common dependent variables studied by comparative criminologists, noting how these variables have been operationalized in the literature, then explores the three methodological approaches most typically deployed in the field, specifically meta-level, parallel, and case studies.

Gregory J. Howard is Assistant Professor of Sociology at Western Michigan University; Graeme Newman is Distinguished Teaching Professor of Criminal Justice at the University of Albany, State University of New York; and William Alex Pridemore is Assistant Professor of Sociology at the University of Oklahoma in Norman.

A
B
S
T
R
A
C
T

139

**A
B
S
T
R
A
C
T**

With the growth of international “transparency” and the capacity of the World Wide Web to disseminate information, data about crime and justice around the world are more accessible than ever. The chapter discusses the three most common types of data on international crime and justice (i.e., official, victimization, and self-report data), describing the threats to the reliability and validity of each type and directing interested readers to existing sources of data relevant to frequently employed explanatory concepts.

In conclusion, the chapter observes that, while comparative criminology is a growing area of study owing to the influence of globalization and concerns about transnational crime, the relative neglect of systematic comparative work in criminology throughout the 20th century means that the field is still in its infancy. Growth in this promising area of inquiry should be nurtured with a renaissance in theory so that research is driven by theory and not by the mere existence of more data.

The History and Goals of Comparative Criminology

The context and history of comparative criminology

Comparative criminology is as old as criminology itself. Beccaria, Bentham, Voltaire, Helvetius, Quetelet, and many others of the 18th-century Enlightenment compared and contrasted their own systems of justice with those of other nations. Their recommendations and findings were often influential in bringing about change in countries other than their own. Indeed, the U.S. Constitution owes some of its language and ideas to the writings of these thinkers (see Granucci 1969; Schwartz 1971). Yet, for most of the 19th century and much of the 20th century, comparative criminology was neglected as nations looked inward for solutions to their specific crime problems. It was not until the middle and late decades of the 20th century that interest again emerged in comparing and contrasting the problems of crime across nations. There are many reasons for this renewed interest. The most obvious is that the latter half of the 20th century saw the world become a smaller place, a transformation initiated by revolutions in communication, transportation, and information technology. At the close of the 20th century, nations are increasingly pressured to account for their actions, and the activities of nations are transparent as never before.

One can reasonably argue that transparency began in economic institutions, where trade and commerce demanded it. But the availability of information about various facets of national social life has flourished as well, some have argued, because of an abiding concern with the health of democracy. Kenneth Prewitt, current Director of the U.S. Census Bureau, has suggested, "A healthy democracy needs a healthy number system, and anything that erodes that number system undermines democracy" (American Sociological Association 1999, 3). Gradually, countries have collected and made available to the international community statistics on a wide range of subjects relevant to the interests of comparative criminologists (see, for example, United Nations Development Programme 1998). Among these data are statistics on crime and criminal justice, which have only recently become widely accessible at the international level (Newman 1999). Although nations formerly guarded information on crime and criminal justice zealously, many nations now provide these data on the Internet, where they are available to anyone with adequate technological resources. The transparency and availability of such information have created a climate in which the promises of comparative criminological research may be realistically pursued.

Although many theoretical, methodological, and philosophical problems certainly have dogged comparative criminology since its inception, there is little doubt that this field of investigation is currently in a state of rapid expansion. While this chapter outlines some of the main problems that confront comparative criminology, the discussion also focuses on what cross-national research has accomplished and what it can do for the field of criminology in the future. We begin with two questions often asked of comparative criminology: What is the comparative perspective, and why employ it? Following this discussion, we move to a consideration of the substantive and theoretical issues that lie at the root of comparative criminological inquiry. We must begin with theory, because the plethora of databases and other information now available from many countries provides an environment that tempts rash comparisons and sometimes unsubstantiated conclusions based on what may be incomparable data. Faced with such a challenge, theoretically informed research supported by sound methodology is the wisest defense. Consequently, we look at the theoretical perspectives that have been brought to bear in understanding crime from a comparative perspective. Following this, we consider crime as a dependent variable in comparative work, then stake out the methodological approaches that are often used in this type of investigation. We then consider the data available to researchers interested in pursuing comparative studies and conclude with some observations about the future of comparative research in criminology.

The comparative perspective

Globalization

In its broadest sense, all social science research is comparative. As Durkheim noted: “Comparative sociology is not a special branch of sociology; it is sociology itself” ([1895] 1938, 157). To the extent that the scientific method depends on comparison, Durkheim is no doubt correct. But comparative criminology demands more than comparison. Comparative investigations do not involve so much a method as a perspective, one that demands a gestalt that in today’s terminology might be called a global view. It may be argued that applying today’s global perspective to comparative criminology is misleading because comparative studies of crime have been in existence for two centuries. Yet, comparative research has received relatively little attention from scholars in our field. Marsh noted in 1967 that a tiny number of articles in social science journals at that time dealt with crime in two or more cultures, and Beirne (1997, xiii) pointed out that this lack of interest persisted into the 1980s. It is clear, however, that engagement in comparative criminology has increased significantly in the last decade (see Adler, Mueller, and Laufer 1994; Ebbe 1996;

Moore and Fields 1996; Reichel 1994). One major reason is simply that a global perspective on all popular subjects now dominates the world.

The nation-state and the universality of the criminal justice system

It has become increasingly clear that much of what we call criminal justice depends to a large extent on the operations and structure of modern nation-states. It is the modern nation-state that has given us the tripartite criminal justice system (i.e., police, courts, and corrections) as it is commonly studied today (Newman 1999). Though it may rest within diverse legal traditions and cultures, this basic tripartite structure is similar in all modern nation-states (for authoritative discussions on legal traditions, see David and Brierley 1985 and Wigmore 1936), and investigations that compare nations do so on the assumption that these nations have similar structures of criminal justice. Thus, more attention has been given recently to comparing whole nations in addition to comparing cultures. This is partly the result of globalization, but much is due to the more easily definable boundaries and structures of nation-states when compared with cultures. (The latter, of course, often transcend national boundaries.)

The single most important difference between cultures and nation-states is that nation-states are political entities while cultures are ways of life. Nation-states may be composed of many cultures, as the recent war in the former Yugoslavia attests, and a single culture (depending on how broadly the term is defined) may span several countries (e.g., Roman Catholicism, Islam, Judaism, American popular culture). Because the nation-state is the operational basis for the crime control activities of criminal justice agencies, and because it is more easily defined than the elusive concept of culture, most recent comparative work in criminology has examined similarities and differences with respect to the nation-state (or relevant political subdivisions such as the state or province).

Definitional diversity of crime

One endemic problem confronting comparative criminology is the enormous diversity in the way different cultures and nation-states define crime, justice, and other relevant concepts. As we shall see, this has not stopped many researchers from conducting a wide variety of studies comparing crime and delinquency

The greatest advantage—although often considered an impediment—for research in comparative criminology is the great diversity that exists cross-nationally with regard to social, economic, and political indicators.

around the world. Many of these researchers have attempted to resolve this problem by adjusting legally defined crime categories into socially defined categories, many of which conform to various common theories about the functions and patterns of cultures and social structures in society (see the discussion of victimization surveys by van Dijk 1999). Others have managed to reclassify the official legal definitions of national crime categories into general crime categories to which nations can match their crime definitions (e.g., the United Nations Surveys of Crime and Criminal Justice Systems). But these solutions are only partial and many problems remain.

Advantage of diversity for comparative criminology

The greatest advantage—although often considered an impediment—for research in comparative criminology is the great diversity that exists cross-nationally with regard to social, economic, and political indicators. Though their structure and organization may vary, basic social and cultural categories such as family, urban and rural life, and community are universals of human existence, so they may be used as fundamental classifications when comparing one cultural group with another. Similarly, as previously noted, the criminal justice system deploys basic categories that may be used to guide comparisons of one nation to another. All nations have police, all have courts, and all have a prison system. However, elements of these parts of the criminal justice system differ widely, as do their relationships with each other, and recent comparative studies of criminal justice focus especially on these differences to derive policy and management implications. Of course, the scientific significance of the findings of cross-cultural and cross-national research depends on the deeper questions that one wishes to answer, and these questions are related to the aims and goals of comparative (or any kind of social science) research.

In sum, the comparative perspective is an approach that employs basic unifying concepts of human groups and seeks to compare cultures and nation-states to highlight the similarities and differences between each class with respect to these universal concepts. These comparisons are achieved in many ingenious ways, often depending on the home discipline of the researcher, and have produced rich information. Some of this information is explanatory in a traditional causal sense (i.e., the ordering in space and time of variables and events) and other information provides descriptive evidence of diversity, which also serves as the source of many questions for future research. The scientific merit of this information may well be questioned, however, depending on the methodology and data sources employed. It is a great problem of comparative research that the data sources are, virtually by definition, influenced by the cultures and nation-states from which the information is extracted and by the cultural commitments of investigators themselves. These are standard problems of scientific

research, however, and attention will be paid to these issues in cross-national terms in the pages to come.

The goals of comparative research

There are several goals of comparative research in criminology. Some are obvious applications of the traditional canons of the scientific method, and some are unique to the study of crime in an international setting. Although comparative criminology attends mainly to understanding criminal and deviant behavior as it is manifested globally, these studies will inevitably yield useful insights about the control of antisocial activity. Thus, the study of criminology will naturally intersect with the field of criminal justice if criminological observations are taken to their logical policy conclusions. With respect to the scientific import of comparative work in criminology, a few important goals are noted here.

Extending theories beyond cultural and national boundaries

Comparative research provides an opportunity for criminological theories, which are typically generated within the context of particular nation-states, to be given a wider hearing (Mueller and Adler 1996). Do the theories developed to explain crime rates in the United States, Finland, Japan, or South Africa serve with equal force to account for criminal violations in other nations around the world? Do theories that try to account for the police use of force in the United States, Russia, Australia, or Brazil help to understand police behavior in other countries? These are questions of replication, and they stand at the heart of the scientific enterprise. Beside permitting criminologists to assess the generalizability of important theoretical propositions, comparative research also assists in the elaboration and specification of theory. If a specific theoretical model does not account for variations in crime in other nation-states, perhaps some refinement can be identified on the basis of this investigation, thereby improving the explanatory power of the theory. Finally, as the world becomes a smaller place through the expansion of globalization and crime and criminal justice become increasingly transnational as a result, comparative investigations ensure that theories of crime and criminal justice will remain relevant to the exigencies of history.

Assessing the performance of national criminal justice systems

Another important goal for comparative work in criminology is the assessment of national criminal justice systems. For example, an article by Maguire, Howard, and Newman (1998) developed an index by which the performance

If the various institutions of criminal justice (i.e., police, courts, and corrections) are to work as a system charged with the control of criminal behavior, there must be some way to assess their performance as an operational unit.

of national criminal justice systems can be validly compared across nations. The idea of a criminal justice *system* is of relatively recent vintage, most forcefully espoused in the 1967 report of the President's Commission on Law Enforcement and Administration of Justice, *The Challenge of Crime in a Free Society*. Despite this conceptualization of criminal justice, which aims to make the practices of crime control more efficient and less fragmented, the promise of the system idea has hardly been realized more than 30 years later, as Inciardi's (1984) characterization of the U.S. criminal justice system as a *nonsystem* suggests.

An index of national criminal justice systems' performances, however, can serve to overcome the fragmented operations of the administration of justice that have been so widely attacked in the academic literature and popular media alike. In short, if the various institutions of criminal justice (i.e., police, courts, and corrections) are to work as a system charged with the control of criminal behavior, there must be some way to assess their performance as an operational unit. The performance measures of national criminal justice systems establish this type of benchmark. Moreover, comparisons of the performance of national criminal justice systems inevitably raise questions of policy and, in the long run, of justice. These are questions for which comparative criminologists should be able to field reasonable answers.

Evaluating national criminal justice policy

Comparative criminology and criminal justice also promise to yield insights into the efficacy of various policy initiatives. For instance, are high levels of gun violence inevitable in the United States because it harbors a gun culture? Perhaps there are other countries that have a high level of gun ownership but a low rate of gun crime (see Killias 1993). Would the legalization of drugs lead to an epidemic of drug use, as is often argued? Perhaps other countries have had a different experience. This is not to say that experiences with crime and its control in one nation should be copied wholesale to another. But when we see different and successful ways of dealing with crime in other countries, we at least know that it is possible and that the current state of affairs in a particular country is not preordained or inevitable. The work of the International Center for Crime Prevention has done much to highlight success stories in crime prevention throughout the world (Waller and Welsh 1999).

Coordinating the fight against transnational crime

Another response often provided to the question of “Why do comparative criminology?” maintains that the globalization of crime, as expressed in the increasingly popular notion of transnational crime, points to the need for a coordinated or transnational criminal justice response. Here the benefits of comparative criminology extend beyond the merely provincial and become more fully universal (Reichel 1994). Central to the prosecution of coordinated efforts, Moore and Fields (1996, 6) contend, is “greater international understanding” because “the more one knows about another people, society, or culture, the greater the potential for understanding their actions and responses to problems and situations.” Put more pragmatically, a coordinated law enforcement response to transnational crime such as money laundering or drug trafficking requires that the interested parties understand something of the characteristics of transnational criminals and recognize the operational strengths and weaknesses of one another’s crime control systems. Providing this type of information is one important goal of comparative criminology.

Critique

In contrast to the previous point, a final reason for pursuing cross-national studies of criminology is to provide critical scrutiny and a reasoned voice to counter what may often seem to be a knee-jerk embrace of all things global. In the rush to counter what is rather vaguely referred to as the threat of transnational crime, significant freedoms and human rights may be sacrificed in the name of the common, universal good. As those who have studied the historical developments of crime and justice in the United States must be painfully aware, the road to hell is paved with good intentions. Therefore, comparative work must serve as the critical conscience of the public to ensure that the widely promoted threat of transnational crime does not lead us down a road to a dystopia that would shock even the likes of George Orwell. Drawing on the words of Ralf Dahrendorf (1970, 55), comparative scholars of criminology can usefully serve the role of “intellectual court jester” or the “fool,” questioning that which is taken for granted and doubting “everything that is obvious, [making] relative all authority, [asking] all those questions that no one else dares to ask.” The goal of the comparative researcher in this capacity is not to be a thoroughgoing naysayer but to strengthen policy agendas and ensure they are defensible in terms of principle and fact. Without such a critical conscience, we are in danger of becoming, like the denizens of Samuel Butler’s ([1872] 1985, 227) *Erewhon*, “a meek and long-suffering people, easily led by the nose.”

Having specified some parameters of the comparative perspective in criminology and identified some of the main goals of this area, we presently move to a discussion of theory in this line of investigation.

Theoretical Explanations for Criminal Behavior in a Cross-National Context¹

Three general theoretical frameworks are commonly employed in explaining the variation of crime rates among nations. The first—grand theories—are metanarratives that entail a high level of abstraction and usually assume that one major theoretical construct, such as a nation's level of modernization or its placement in the world's political economy, has the greatest impact on its level of crime.² Structural theories, on the other hand, traditionally have been employed with smaller units of analysis, such as cities or states within a particular nation, and attempt to explain the spatial variation in rates of offending via subcultures (i.e., social learning), status-induced strain, or social control. These theoretical explanations are increasingly being tested at the cross-national level. Finally, a nation's demographic characteristics, such as its age and sex structure, may also be employed in an attempt to explain levels of violence and property crime. Each section following contains a brief description of theoretical concepts, common measurements for each of these concepts, and a selected list of studies that analyze crime from the particular approach. In all cases, these studies represent theories that were originally developed within one cultural and/or national tradition but that have been recast and applied to cross-national research designs. Thus, many of the theoretical approaches will initially appear familiar to criminologists. However, the employment of these theories in cross-national research promises either to extend the explanatory power of these theories or to demonstrate their limitations.

Grand theories

Modernization theory

Durkheim's notion of anomie is the basis for contemporary theories that pinpoint the effects of modernization as the main causes of crime. Durkheim ([1893] 1964) argued that, as nations develop, they are characterized by an increasingly intricate web of social and economic relations. These complex divisions are suspected of undermining mechanical solidarity and its control over the collective conscience. Thus, rapid social change engenders the breakdown of traditional values, resulting in, among other things, a higher crime rate. Eventually, however, organic solidarity and more formal mechanisms of social control should halt rising crime rates, although they are expected to remain at higher levels than before development.

Within cross-national criminology, Clinard and Abbott (1973) and Shelley (1981) have provided notable contributions to Durkheim's idea of modernization and

crime. In short, the contention is that each nation experiences similar phases of development. The stimulus for modernization is technological advancement, and this catalyst leads to political, economic, and demographic changes within a society (Strasser and Randall 1981). Industrialization and urbanization are key elements of this social transformation. Their effects—which include the tension between groups that accompanies increased social differentiation and the socioeconomic inequality many presume to follow modernization—are viewed as the main contributors to rising crime rates (Heiland and Shelley 1992). The outcomes of this process are expected to provide stronger explanations of crime than any distinct national or cultural characteristics have; thus, all nations are expected to experience similar trends in crime rates as they develop.³ Finally, changes in both types and rates of crime are expected as a result of development. In general, the overall rate of crime is expected to increase but eventually level off as modernization progresses, while another expected result is the predominance of property and economic crimes over crimes against people.

Recent examples of international research on modernization and crime include:

- Bouley and Vaughn's (1995) study of violent crime in Colombia, in which regression analysis revealed support for grand theories with respect to the crimes of theft and robbery but not for the more violent crimes of assault and homicide.
- Mahabir's (1988) work on urban gangs in the Caribbean.
- Industrialization and crime in the Russian region of Tuva (Balakina 1994), which official data show has the highest homicide rate in the country (Pridemore 1999).
- Huang's (1995) multivariate analysis of 29 countries employing United Nations data, which showed support for the modernization hypothesis.
- Ortega and colleagues' (1992) analysis of 51 nations using International Criminal Police Organization (Interpol) crime data, which revealed support for the Durkheimian-modernization thesis.
- Johnson's (1990) work, which examined crime and industrial development in Germany and concluded that modernization theory did not hold in this case.
- Matsuoka and Kelly's (1988) study of the negative impact of resort development and tourism on Native Hawaiians.
- Neuman and Berger's (1988) evaluation of modernization, Marxian world system, and ecological opportunity theories, which lent only weak support to the modernization theory.

- An application of Blau's (1977) macrostructural theory by Messner (1986), who claimed that the inequality that accompanies development is a major reason for heightened crime rates. (Inequality and heterogeneity have been popular topics with cross-national researchers; see Avison and Loring 1986; Hansmann and Quigley 1982; Krahn, Hartnagel, and Gartrell 1986).
- Schichor's (1985) examination of homicide and larceny rates, which indicated that development is likely to be accompanied by increased property and decreased violent crime.

Where modernization assumes that crime rates will increase and then level off over time, civilization theory expects decreasing crime rates as governments and their citizens become more humane and civilized.

Civilization theory

Where modernization assumes that crime rates will increase and then level off over time, civilization theory expects decreasing crime rates as governments and their citizens become more humane and civilized. Civilization theory is based largely on the work of Elias (1982), who argued that social norms and manners have become increasingly refined over the centuries. As this civilizing process occurs, individuals learn to inhibit their urges and societies become less violent as a result. Self-restraint, therefore, has become the hallmark of control, not external (i.e., state) threats or punishments.

As it relates to crime and control, proponents of this theory usually focus on how the forms of official social control and punishment have changed, but individual behavior obviously plays an important role.

Similar to the Durkheimian ([1893] 1964) notion of modernization, these theorists contend that as industrialization expands, it creates a complex division of labor that demands a high degree of interdependency. As this organic solidarity grows stronger, people exercise a higher degree of internal control over their behavior because others increasingly depend on them (Heiland and Shelley 1992). This internalization of control is expected to lead to a decrease in crime rates, especially in violence. As individuals increasingly repress their urges, however, they are likely to experience an increase in psychological pathologies (Freud 1962) and self-inflicted victimization (e.g., suicide and drug abuse).

Empirical studies by criminologists with direct references to the civilization hypothesis are rare. However, one study that does claim support for the theory is Gillis' (1994) work on literacy and violence in 19th-century France. The author employs data on violent crimes and suicide in France from 1852 to

1914, revealing that the climbing literacy rate was associated with a decrease in the rate of crimes of passion (e.g., homicide and certain other types of violence) and an increase in the suicide rate. Gillis argued that this information represents strong support for the hypothesis, which predicted that, as the civilization process occurs, interpersonal violence will decrease and violence directed against the self will increase.

Opportunity theories

In recognizing the complex social-structural changes that accompany societal evolution, opportunity theories are similar to both modernization and civilization theories. However, while the modernization hypothesis focuses on how changes compromise traditional values and the civilization thesis expects the development of internal self-restraint (as opposed to the external social control of the state) to inhibit harmful behavior, opportunity theories suggest that modern economies and social organization provide increased opportunities to engage in criminal behavior (Cohen and Felson 1979). For example, expanding economies create an increase in expendable income in the average household, which people can then spend on a growing variety of consumer goods, which in turn are increasingly available for theft. At the same time, technological gains produce smaller and more portable electronic devices that are easily stolen. Similarly, work (e.g., both spouses working instead of only one) and leisure activities (e.g., a larger amount of expendable income to spend on entertainment) may mean less time spent at home for many, which results in less guardianship over household items that may be stolen. At the same time, increased residential mobility and cultural heterogeneity may lead to weaker community ties than in the past, resulting in communities composed of people who are less willing to guard the personal safety and private property of neighbors they barely know.

It should be made clear that opportunity theorists do not expect the conditions of modern society to create forces that cause criminal behavior; rather, they assume that we are all motivated offenders who will act criminally in a situation given the presence of a suitable victim and the absence of a capable guardian (Cohen and Felson 1979; Felson and Cohen 1980). Thus, opportunity theory posits an increase in property crimes and a decrease in violent crimes over

Opportunity theorists do not expect the conditions of modern society to create forces that cause criminal behavior; rather, they assume that we are all motivated offenders who will act criminally in a situation given the presence of a suitable victim and the absence of a capable guardian.

time. Studies that directly test opportunity theories are not widespread at the cross-national level (for example, see Bennett 1991a, 1991b; Gartner 1990; Kick and LaFree 1985; Ortega et al. 1992). However, though it may be difficult to measure the amount of time citizens of different countries spend away from home or the average weight of their televisions, studies that examine the effect of modernization and crime may also be interpreted in terms of this thesis. For example, while Bennett (1991b); Groves, McCleary, and Newman (1985); Neapolitan (1994, 1996); and several others show no significant increase in the homicide rate with economic development, others, such as Stack (1984) and Hartnagel (1982), have shown that increases in property crime accompany economic development and/or urbanism, which can result in an increase in the number of victims and, due to mobility and heterogeneity, a decrease in guardianship.

World system theory

World system theory borrows from the Marxist perspective to explain the impact of an ever-expanding capitalism on nations that vary in their level of development (see Chirot and Hall 1982). This theory recognizes that the legitimacy of a market economy is spreading around the globe and that its expansion is uneven, meaning that (1) nations are no longer autonomous political and economic entities but are instead actors in an international political-economic system, and (2) weaker countries are politically and economically exploited by stronger ones (Smith 1984). This uneven expansion results in nations that are placed into one of three categories:

- *Core* nations are industrialized, and market relations in them are highly advanced.
- *Periphery* nations are characterized by a history of colonialism and the possession of natural and human resources that are underdeveloped but available for exploitation by the industrialized sectors of core nations. These nations are viewed as economically dependent on the core nations.
- *Semi-periphery* nations are underdeveloped, perhaps only partially industrialized, and at the mercy of both the core and the periphery (Evans and Timberlake 1980; Walton 1982).

As capitalism expands, world system theory maintains, it disrupts indigenous cultures and traditional means of subsistence, producing exploitation from the outside and new inequalities within. Political and legal formations are disrupted, and social dislocations become widespread. The rural population begins to migrate to cities in search of employment, creating class conflict and

competition for scarce resources (Castells 1977; Gilbert and Gugler 1982). Social relationships are replaced by market relations, and consumerism replaces traditional use patterns (see Fromm 1976).

World system theorists argue that shifts in political, economic, and social organization, together with the poverty, inequality, and poor living conditions resulting from this process, produce criminogenic conditions of all sorts. First, as a result of the human toll, the process itself is viewed as illegal by some, and multinational corporations and governments are labeled criminal (see Reiman 1998). Second, informal economies that deal in illegal goods develop as a way to produce income for the poor, but they may also help create destructive behavior and higher victimization rates among the lower classes. Third, the social and psychological strain that results from cultural shock and poor living conditions, together with culture-specific inequities that distinguish “acceptable” scapegoats for victimization, result in increased rates of violence (Messerschmidt 1986). Finally, collective responses such as protests, riots, and even some types of theft and violence might be viewed as an attempt to (1) foment political unrest, (2) create class consciousness out of the power and economic inequities, and (3) spark revolution—or at least forms of primitive rebellion (Hobsbawm 1959, 1969; O’Malley 1980).

As with most conflict-based theories, the tenets of world system theory lend themselves more to descriptive studies and are difficult to test empirically, especially at the cross-national level. Of course, the general findings of work testing other metanarratives, such as modernization, may be interpreted from the world system theory approach. Commonly, world system theory is advanced through case studies of one or a few nations in which the processes mentioned previously are described and general economic and political indicators are used to classify nations as core, periphery, and semi-periphery so that the relations between the indicators (and the resulting social problems they may create) can be discussed. This literature, however, is usually found in the disciplines of political science, anthropology, and sociology, and criminological issues often are considered only in tangential fashion.

Although there is no space here for a complete critique of each of these theories, there are two important points, one theoretical and one empirical, that should at least be introduced. Theoretically, these are metanarratives that attempt to explain crime causation largely in terms of a single, albeit broad, issue. As a result, an appreciation for a multicausal understanding of crime may be lacking. Empirically, these metanarratives are, in fact, theories of social change and of the effects of societal evolution on the nature and quantity of crime in a society. However, nearly all the studies mentioned previously employ cross-sectional designs to evaluate theories of temporal variation. For example,

it may be argued that a relationship between economic distress and crime within a country reveals support for a Marxist approach, but this does not necessarily speak to a larger world system and whether this system in fact has created the poor economic conditions within a nation. A cross-sectional design is not an entirely inappropriate methodology, and its predominance has perhaps been necessary due to the limited availability of data on nations over time. However, given the increasing availability, reliability, and validity of cross-national data and the growing sophistication of scientific techniques employed by criminologists, we must take care to design studies that best answer the proposed theoretical hypotheses and recognize design limitations when drawing our conclusions.

Structural theories

There are three general approaches to the study of the etiology of crime: social learning, strain, and social control. The structural analogs of these theoretical strategies are culture, strain (usually represented by absolute or relative economic deprivation), and social disorganization. These three theories, together with attempts to measure them and identify their impact on crime at the cross-national level, are outlined briefly.

Culture

The attempt to explain crime via cultural variation is a longstanding enterprise. In the first half of the 19th century, Guerry blamed the high rates of violence in the south of France on regional differences in culture that resulted from migration and settlement patterns (Corzine, Huff-Corzine, and Whitt 1998). In Italy, many argued that southern Italians, such as the Neapolitans and Sicilians, possessed cultural traits responsible for the high levels of crime in the regions they inhabited. In the United States, Redfield (1880) began a research tradition that continues today, when his systematic research revealed heightened rates of homicide in the American South.

Despite the fact that southerners everywhere seem to be viewed as a rather impetuous lot, cultural theories of crime are actually grounded in social learning processes and cultural norms. Researchers do not believe, for example, that southerners are born violent; instead, they submit that residents of the South learn violent traits from those in close proximity to them.⁴

Modern cultural theories, then, build on Sutherland's (1947) notion of differential association, in which norms conducive to violence are transmitted to individuals and across generations via processes of social learning. Having acquired these cultural values, the individual is provided with a "tool kit" for living (Swidler 1986), as culture provides one with the means to interpret interpersonal

interactions while at the same time providing an accepted repertoire of responses appropriate for each situation (Wolfgang and Ferracuti 1967). Individuals in one culture, for example, may be offended by an action that someone in another culture dismisses as unimportant. Similarly, Luckenbill and Doyle (1989) argue that individuals in some cultures are more likely to view a negative interaction, no matter how slight, as an injustice that demands revenge. Thus, cultural norms may promote or condone violence in certain situations, meaning that the attitudes and values of individuals within cultures that have higher rates of crime or violence should be distinguishable from those with lower rates.

Measuring and analyzing the effects of culture are extremely difficult tasks for the researcher. Some suggest that culture dictates the form of social institutions within a society (Lynch 1995a), while others argue that institutional/structural conditions have a hand in creating (sub)cultural values and that these values mediate the effects of social structure on behavior (Curtis 1975). These two views are not mutually exclusive and both are probably correct. However, attempts to measure culture at this level of analysis have been weak at best—usually taking national, regional, or ethnic group membership as a measure of culture—thereby making efforts to separate the discrete effects of culture and structure nearly impossible.

Research in the United States has focused on *subculturally* violent groups (see Wolfgang and Ferracuti 1967; Messner 1983a, 1983b; Nisbett and Cohen 1996), but those interested in cross-national comparisons must focus on the nation as the unit of analysis. This makes it difficult to examine cultural influences on crime because a cross-national sample requires an assumption that countries are culturally homogeneous, which is rarely the case. Therefore, although the literature (especially anthropological) abounds with case studies that examine the relationship between sociocultural attributes and levels and types of violence, there have been few empirical attempts to measure the effects of culture on crime with a cross-national sample.⁵ This is not a futile task, however. It is possible to devise a standard survey instrument that might measure beliefs concerning situational acceptance of violent behavior and other similar attitudes among members of different nations.⁶ To say that cultural variation does not affect levels and types of violence among countries makes little sense; most agree that it does. Cross-national researchers, however, have yet to

Thus, cultural norms may promote or condone violence in certain situations, meaning that the attitudes and values of individuals within cultures that have higher rates of crime or violence should be distinguishable from those with lower rates.

include an adequate measure of culture in structure-level models. This seems like a fruitful avenue of research, but we must be careful in how we analyze the issue. Just because one country or culture consistently exhibits higher crime rates than others, for example, does not make it culturally criminal (as has been suggested with the subculture of violence thesis in the United States) because, as noted, culture is integrally linked to history and to political and economic structure.

Strain

Some researchers argue that, regardless of cultural attributes, crime rates will vary spatially based on local structural composition. Strain, usually represented as absolute or relative deprivation in structure-level models, is one of the most widely tested elements in cross-national research on crime. For example, many suggest that the social and psychological strains generated by poverty lead to higher rates of crime in areas possessing a higher proportion of people facing these conditions (see Williams and Flewelling 1988). Some contend, however, that individuals' perceptions that others are somehow better off creates frustration over this inequitable distribution of resources, which is eventually expressed through aggression and violence (Blau and Blau 1982; Fowles and Merva 1996; Merton 1938; Messner 1982). This is contingent on the assumption that community members recognize these inequities, regard them as unfair, and respond violently to them.

Oddly enough, given the consistent findings in U.S. studies relating poverty to violence, this relationship is rarely tested at the cross-national level. Instead of using a measure of the extent of poverty within a nation, most researchers choose to employ a gross domestic product (GDP) or gross national product (GNP) per capita measure as an indicator of modernization (see the previous section on modernization; other examples include Bennett 1991a; Fiala and LaFree 1988; Groves, McCleary, and Newman 1985; Ortega et al. 1992). Because a measure of central tendency—such as average income—is only a rough measure of the magnitude and depth of poverty in a population (see LaFree, Drass, and O'Day 1992; McDowall 1986) we must be careful in interpreting this in terms of a poverty-violence relationship. However, several cross-national studies have shown that as per capita GNP or GDP increases, homicide rates tend to decrease (Krohn and Wellford 1977; McDonald 1976; Neapolitan 1994, 1996). This is the opposite of what modernization theorists expect, but it may lend tentative support to the poverty-violence thesis.

Although theoretically defined in terms of anger and frustration resulting from the inequitable distribution of resources, inequality is most often measured in terms of the Gini coefficient, which measures income distribution but not anger

or frustration. The Gini coefficient (often referred to as the index of income concentration), ranges from 0 to 1, with 0 indicating perfect equality and 1 indicating perfect inequality within a population.⁷ A few cross-national studies that employed the Gini coefficient as a measure of inequality include Avison and Loring's (1986) study of population diversity and homicide, Gartner's (1990) work on homicide victims, Messner's (1989) analysis of economic discrimination and homicide rates, and Messner and Rosenfeld's (1997) exploration of the relationship between institutional anomie and homicide rates. Alternative measures of inequality are defined in terms of the percentage of overall income received by a specific proportion of the population or as a ratio of the earnings of one segment of the population to that of another. For example, Fiala and LaFree (1988) used both the percentage of total income received by the bottom 20 percent of the population and the ratio of the percentage of income received by the top 20 percent to that received by the bottom 40 percent of each nation.

Social disorganization

While cultural models are founded in social learning principles and strain theories argue that economic and other forms of distress may propel people or groups toward criminal behavior, the theory of social disorganization posits that crime occurs as a result of a breakdown in social bonds. Structural forces act to disrupt social ties and group solidarity, thus interfering with community mechanisms (both formal and informal) of control. Detached from their social bonds and in the absence of the community's ability to control the behavior of its members, people are free to become involved in criminal behavior (Bursik 1988; Sampson, Raudenbush, and Earls 1997). Thus, social disorganization is an extension of the modernization thesis outlined previously (e.g., one situation in which normative controls may be broken is when shifts in urbanization and the political economy occur during development). This view of community disorganization is similar to the notion of social control discussed by Park and Burgess (1924) in their work on human ecology.

The community or neighborhood is usually considered the most appropriate level of analysis for testing social disorganization models. However, structural forces outside the community, such as political-economic shifts that redirect the distribution of jobs and services (Bursik 1988), obviously have an impact on social cohesion within a neighborhood. Thus, this theory is often tested at higher levels of analysis, including cross-nationally. The commonly accepted elements of these macrolevel models of social disorganization are poverty, population density, ethnic heterogeneity, residential mobility, and family disruption. Each is suspected of disrupting social integration and cohesion within communities, thereby weakening controls and allowing increases in crime rates (Sampson, Raudenbush, and Earls 1997).

The structural elements of social disorganization are described theoretically in a manner that makes them relatively easy to operationalize in measurement models for cross-national analysis. For example, poverty can be measured as the proportion of the population living below the poverty line; population density can be captured as the percentage of the population that resides in urban areas (see Krahn, Hartnagel, and Gartrell 1986; Messner 1989; Ortega et al. 1992) or even the density of the whole country (see Avison and Loring 1986; Neapolitan 1994). Several definitions of heterogeneity are available, including ethnic and linguistic differentiation (commonly consisting of the percentage of the population that is not of the same ethnic background—or that speaks an alternative first language—as the majority; see Gartner 1990; Hansmann and Quigley 1982; Messner 1989). Residential mobility can be operationalized in terms of either urban (Braithwaite and Braithwaite 1980; Fiala and LaFree 1988) or total population (see LaFree and Kick 1986; Schichor 1990) growth or decline. Family disruption is most often measured as a nation's divorce rate (see Gartner, Baker, and Pampel 1990; Rosenfeld and Messner 1991).

Within the past 15 years, elements of social disorganization have become widely tested in cross-national empirical studies. Even if not directly discussed in terms of this theory, the concepts involved are commonly employed as control variables. No doubt there are problems—in terms of theory, measurement, and the congruence of the two—with each of the measures discussed in the previous paragraphs, which authors of these studies sometimes address. However, the theoretical model for social disorganization is constructed in such a way as to make the production of a measurement model relatively simple, especially because data on these concepts are often readily available cross-nationally. Finally, tests of this model result in similar findings at several levels of aggregation (e.g., community, city, nation), suggesting that it might be a viable model that deserves further scientific attention at the cross-national level.

The demographic correlates of crime

Although they do not provide causal mechanisms, the demographic attributes of a nation's population are often used as control variables and, more recently, as elements of opportunity theories. In cross-national research, these demographic correlates are commonly accepted to be sex and age.⁸ In most studies, the sex and age categories are combined, the suggestion being that young males have the highest offending and victimization rates, and thus, as their proportion of the population increases, so will crime rates. Empirically, the cross-national findings examining this relationship have been inconsistent at best. Using various measures (age categories include younger than 15 years, 15–19, 15–24,

and 15–29, with some employing only males of these ages and others including both males and females), findings ranged from positive (see Hansmann and Quigley 1982; Ortega et al. 1992), to nonsignificant (see Gartner 1990; Messner 1989; Neapolitan 1994), to negative (see Bennett 1991a). At this stage, these inconclusive findings are of relatively little concern because they may be explained by any or all of the theoretical processes outlined previously or by the cross-sectional methodology usually employed. Given the commonly accepted association between these correlates and crime, however, any empirical test should continue to employ them as controls.

Methodology

This section of the chapter contains two main sections. The first is a survey of comparative research that examines specific substantive issues of crime (e.g., violence, property crime, genocide, transnational crime). The second section examines the general types of studies (i.e., metalevel, parallel, and case) normally undertaken by comparative criminologists.

Crime as a dependent variable

When comparative researchers undertake their studies of crime in a cross-national context, they often consider only one category of crime. In the next few pages, we first address some of the most common types of crime considered in comparative inquiries, such as violent and property crimes, then move to a discussion of a few types of crime that are only now beginning to receive serious scrutiny. Obviously, this brief overview of crime as a dependent variable is far from complete, as the proper subject of criminological investigation cannot be so summarily dictated given its constant state of development. For instance, we do not directly consider the import of Beirne's (1999) recent call for animal abuse as an object of criminological study, although provocative comparative work certainly could be pursued on this topic. Still, the dependent variables that we do identify are those most commonly deployed in comparative criminology, and the domains that we signal as areas of criminological inquiry in the future should alert scholars to the significant promise of these arenas of study.

Comparative studies of violent crime

Comparative studies of violence appear more often than studies about any other category of crime. Although violent behaviors such as assault, rape, and robbery have been the subject of much comparative research, studies of homicide are probably the most popular because of the mortal nature of the offense and

the higher availability, reliability, and validity of homicide measures for a large number of nations (see LaFree 1998 for a review of cross-national studies of homicide). A nation's homicide rate is also considered by many to be a fairly accurate indicator of its overall level of criminal violence (Fox and Zawitz 1998).

Comparative studies of violence may be able to answer several questions for criminologists. For example, do a nation's levels of modernization and industrialization increase its level of violence? Do different political-economic structures exhibit varying rates of violence (e.g., are homicide rates under communism and a state-run economy significantly higher or lower than nations with rule-of-law and free markets)? Do rates of violence vary with the levels of ethnic, linguistic, and religious heterogeneity within a nation? The answers to these and other questions posed by comparative criminologists can provide insight into the fundamental effects of cultural and structural organization on a nation's amount and forms of violence. Comparative studies of violence may also aid us in understanding whether varying manifestations of violence (e.g., rape, assault, homicide) are discrete forms of behavior that require separate causal models, or if they should all be contained under the general category of violence.

Comparative studies of property crime

Empirical work on property crime at the cross-national level has faced significant difficulty in the past due to definitional problems. The legal protection of private property, even the demarcation between private and public property, varies across nations. In a communist country, for example, all property in theory belongs to the state; thus, legal protection of private property is limited. Further, the likelihood of the development of an alternative economy for goods not provided by the centralized economy is high in such nations, and state ideology might label such transactions as speculation (for which there is no room in a strictly communist political economy) and police them accordingly. Both this ideology and the parallel economy confuse what Westerners normally call property crime. Legal definitions of property crime also vary among countries. Robbery in one country might be coded as a simple or aggravated theft in another. Bicycles may be the major mode of transportation in a nation and expensive for citizens to replace: Should theft of bicycles in this case be considered equivalent to motor vehicle theft elsewhere? Reporting practices also vary across nations, depending on several factors discussed elsewhere in this chapter. This presents further difficulty to researchers interested in property crime. Finally, the availability, reliability, and validity of data from a variety of different types of countries (e.g., developed and developing, socialist and free market) are far from ideal.

Cross-national study of property crime presents several intriguing questions, however. How much money do citizens of the world lose every year due to property crime? Does development really increase theft, as opportunity theorists suggest, simply because there are more attractive and more portable goods to steal? Do poverty and inequality have different levels and types of effects on the rates of property crime? Does property crime take different forms in different nations, especially those with disparate political-economic structures or levels of development? Are the antecedents of property crime around the world the same as those of violent crime? In what way might a nation-state's culture mediate the effects of its structure on property crime? Recent advances in methodology that rely on behavioral rather than legal definitions of crime (as described in the data sources section) make undertaking research on these questions less difficult. The answers to these questions, however, must be based on careful consideration of many fundamental issues sometimes not considered by criminologists (e.g., social-structural and political-economic conditions, the presence or absence of insurance companies). Keeping in mind these caveats, the comparative literature on property crime looks to expand in the coming years. The increasing availability and validity of data, coupled with increasingly sophisticated theoretical models, should allow criminologists to make significant gains in the study of property crime cross-nationally.

Genocide: National crimes with international implications

Although certain wartime behaviors and genocide have been recognized as international crimes since the end of World War II, these two subjects have received little systematic criminological attention (Adler, Mueller, and Laufer 1994). Questions about war crimes and genocide fall roughly under the rubric of macrocriminology (Shoham 1995). This term refers to crimes committed by whole governments—indeed, whole nations—against either their own citizens or those of other nations. In recent years, there have been attempts to establish an international legal machinery that defines the actions of individuals who perform genocide in the name of governments as crimes against humanity. It is rare, however, for a whole nation to be held accountable for crimes. Perhaps Iraq is an exception to this general rule, given the harsh sanctions imposed on it after it lost the Gulf War. However, the question of the criminality of that nation remains in dispute, especially as the tendency to hold individuals accountable for the acts of nations persists. Thus, Saddam Hussein, the “Butcher of Baghdad,” is vilified as the criminal leader of Iraq. In a similar vein, Slobodan Milosevic is denounced as the “Tyrant of Serbia” and Augusto Pinochet, the criminal “Dictator of Chile.” Attempts to apply criminal law in the international court are therefore limited, probably by the model of criminal law itself. Criminal law is essentially constructed to apply blame and attribute

responsibility to individuals or small groups of individuals (e.g., participants in conspiracies, corporations) but not to whole nations (see Barak 1991; Chambliss 1993; and Ermann and Lundman 1992 for intriguing discussions on the subject of state and governmental crime).

However, the inclusion of these topics in the field of cross-national criminology is crucial for an important philosophical reason. It has been argued in some quarters that comparative criminology is essentially impossible because the definition of crime is relative to particular cultures in particular times and places. This argument is most often put forward by the relativists of social science and criminology (Beirne 1997). Briefly stated, the claim is that by definition crime is a social phenomenon, defined by the culture and history unique to each nation. Thus, they claim that what is criminal in one country may not be defined as criminal in another. For evidence, they point to the wide variety of legal definitions and legal systems existing across the world.

While we need not go into the philosophical debates concerning relativism and social science (for discussions on this topic relevant to criminology, see Beirne 1983; DiCristina 1995; Leavitt 1990; Newman 1976), we simply point out the consequences of pursuing such a position. Put simply, the relativist position maintains that it is impossible to make judgments about human rights, the tyranny of dictators, or genocide committed against innocent citizens anywhere in the world. The relativist view is that, because these acts are not defined as crimes within a particular country (instead, they are lauded as ethnic cleansing or favorable to national security), they cannot be judged as more or less criminal than any other act. Nonetheless, the atrocities of this century have given rise to attempts to hold individual tyrants from particular countries responsible for their actions (the best examples are the Nuremberg trials after World War II and the establishment of the International Court of Justice in The Hague). These efforts suggest a recognition that some crimes are universally abhorrent (in the sense that the whole world suffers when they are committed). We merely ask the relativists whether they would prefer to not make such judgments. Thus, we conclude that the study of genocide, war crimes, and human rights is nascent but involves extremely challenging and legitimate research for criminologists, and we urge the development of a new field that can be termed macrocriminology. We are aware of some works emerging in this field, such as those concerned with comparative human rights (Bouloukos 1999) and genocide (Shoham 1995), and we encourage further efforts in this line of inquiry.

Domestic violence

It is often the case, especially in cross-national work, that the topics that most interest researchers are the most difficult to study. This is especially true for

domestic violence,⁹ a general term we use here to encompass spouse and child abuse, incest, infanticide, and similar family-related violence. An understanding of the variation in levels of domestic violence cross-nationally could yield valuable insight into the status of women and children around the world as well as help unearth the etiological factors responsible for this type of violence. Researchers might also be able to determine the association between domestic violence and overall violence, perhaps revealing whether the former is a special case of the latter or if different models are needed to explain each. Researchers can test any of the theories discussed previously (e.g., opportunity or strain theories) to determine their explanatory power in terms of spouse and child abuse. In short, an understanding of the cross-national variation of domestic violence could yield important revelations about both culture and social structure.

Unfortunately, empirical research of domestic violence at the cross-national level is extremely difficult for several reasons. First, disparate historical experiences, economic structures, religious beliefs, and other cultural factors have resulted in widely varying definitions of spouse and child abuse. Further, even given similar legal definitions, cultural norms may preempt legal dictates in regard to tolerance for actual behavior and the reporting and recording of incidents. Also, the availability and accessibility of institutions created to assist victims, punish and/or counsel offenders, and generally respond to these crimes vary widely throughout the world and are themselves a viable area of study. The availability of these institutions is likely related to some of the factors that produce varying levels of domestic violence, and their presence or absence also likely affects the true rate (and the reported and recorded levels) of victimization within a nation, possibly creating spurious associations between independent and dependent variables. Given cultural and structural differences, the form that spouse and child abuse takes likely varies from country to country, making measurement even more difficult. Finally, the structural, cultural, and religious factors that might mediate or exacerbate domestic violence are themselves difficult to capture in quantitative measures. These challenges have not stopped researchers from undertaking several qualitative case studies of different countries and from offering general comparisons of rates among nations. Also, as data increasingly become available, the number of comparative studies and scientific examinations of the etiology of domestic violence is rising, which should provide valuable insight into the various aspects of family-related violence throughout the world.¹⁰

Transnational crime

Perhaps the most recent and popular addition to the domain of comparative criminologists is transnational crime. Tremendous increases in trade and commerce

have produced considerable movement of people, goods, and ideas across borders at lightning speed. The enhanced interdependence among nation-states demanded by globalization has created a world in which transnational crime is not only possible, but perhaps inevitable. Of all the categories of crime described in this section, this type is the most fluid. Adler, Mueller, and Laufer (1994, 533) have defined transnational crime as “criminal activities extending into, and violating the laws of, several countries.” For their part, Martin and Romano (1992, 1, 4–5) specify transnational crime, or what they prefer to call “multinational systemic crime,” as “crimes by various kinds of organizations that operate across national boundaries and in two or more countries simultaneously. . . . It is crime by networks operating within a multinational arena, often with state support.”

Given its very nature, transnational crime would seem to require some organizational sophistication; for this reason, it is often linked to the idea of criminal organizations such as the Russian Mafia (although nation-states themselves cannot be excluded from the realm of transnational criminality). The types of behaviors that can be grouped into this category of behaviors are vast, and the diversity of activity considered to be transnational crime has probably hampered systematic efforts to study it. Nevertheless, a number of types of transnational crime have garnered recognition, including terrorism, espionage, drug trafficking, arms trafficking, environmental crimes by multinational corporations, motor vehicle theft, trafficking in humans and organs for transplants, fraud, money laundering, and art theft. Because these criminal activities take advantage of the interstices between nation-states, little official attention has been directed toward them, and consequently, there is a lack of data with which to assess the problem. Moreover, efforts at stemming this type of crime are further hampered by the fact that criminal justice systems are developed with the aim of policing within national borders.

Nevertheless, the increased attention presently devoted to this variety of crime by the United Nations, individual nations, and scholars is increasing the information available to criminologists who wish to undertake investigations of transnational crime. Surely, this will be an area of criminological investigation that will begin to bear fruit in short order. For the moment, however, much of the information is anecdotal and depends to a large extent on media accounts of transnational crime, and these data sources bring with them considerable difficulties in terms of validity, as noted by a number of researchers (Passos 1995; Williams 1999).

Methodological approaches to comparative studies of crime

We divide our discussion of methodological approaches to comparative studies in criminology into three general varieties. We begin with what we call metalevel studies, move to a consideration of parallel studies, and conclude with a consideration of case studies.

Metalevel studies

We use the generic term metalevel studies to describe research that employs the nation as the unit of analysis to quantitatively compare criminological issues in several countries. This work usually takes the form of multivariate regression analysis undertaken to test one or more of the theories discussed earlier. For example, a researcher may wish to examine the effects of poverty on homicide rates around the world or discover if a nation's level of development increases its rate of property crime. This is the most common category of empirical research done cross-nationally and, with the increasing availability, reliability, and validity of data, is becoming a more sophisticated and popular approach.

Examples of this type of comparative analysis are usually quantitative in nature, either simply describing trends, patterns, similarities, and differences between nations or employing statistical techniques to test criminological theories and search for correlates of crime at the cross-national level. The examples included here have been randomly chosen. Almost all have been published within the past 15 years, most during the 1990s, and works cited elsewhere in this article are not repeated here. The examples have been selected to present the variety of work undertaken by comparative criminologists, and their inclusion in no way suggests support for the validity of the theory, methodology, measurement, or findings involved.

Because homicide is commonly accepted as the most reliably measured crime, work on this topic has dominated cross-national research on violence. Examples include LaFree's (1998) summary of cross-national studies of homicide, Gartner's (1990) examination of the victims of homicide, Lester's (1991) test of the opportunity thesis as an explanation of European homicide rates, and Neapolitan's (1994) study of homicide in Latin American countries. Work by Krug, Powell, and Dahlberg (1988) and Killias (1992, 1993) addresses firearm-related deaths, gun ownership, and violence throughout the world. Recent research on child homicide includes studies by Fiala and LaFree (1988) and Briggs and Cutright (1994) and a comparison of the levels of child homicide in developed countries by Unnithan (1997). Junger-Tas (1996) and Pfeiffer (1998) have both published studies of juvenile violence in Europe. Examples of work on violence other

than homicide include research on family violence by Bowker (1985), who examined the effects of modernization on spouse abuse in developing countries; an early review of the literature on spouse abuse in several different nations by Cornell and Gelles (1982); a more recent and extensive cross-national bibliography on family violence by Patrignani and Ville (1995); and an edited volume on child abuse and neglect by Gelles and Lancaster (1987), which includes essays on several aspects of child abuse in different countries.

Cross-national variation in definitions and differences in the reporting and recording of property crime makes it difficult to examine this topic empirically, but the recent addition of victimization surveys and self-reports is a boon to this area of study. Cross-national research has examined the impact of development on property-related crime in Africa (Arthur 1991), on property crime patterns in general (Schichor 1990), on gender and property crime (Anderson and Bennett 1996; Widom and Stewart 1986), and on cross-national differences in theft in less developed nations (Neapolitan 1995). Zvekic and Alvazzi del Frate (1995) presented a volume that provides discussions on criminal victimization in developing countries based on the International Crime Victimization Survey (ICVS) (see also Alvazzi del Frate 1998). Kick and LaFree (1985) examined the social determinants of theft in 40 nations, and Stack (1984) provided cross-national evidence for a relationship between income inequality and property crime. In another volume, Kangaspunta, Joutsen, and Ollus (1998) employed data from the Fifth United Nations Survey to examine levels of property crime offenses (including burglary, motor vehicle theft, and car vandalism) in European and North American countries. Luikkonen (1997) provided a more indepth look at motor vehicle theft in Europe, and van Dijk and van Kesteren (1996) used ICVS to study criminal victimization in European cities. Property crime committed by juveniles was the topic of empirical research completed by Bennett and Basiotis (1991) and Bennett and Lynch (1990), with both examining the structural correlates of juvenile crime cross-nationally, and by Junger-Tas, Terlouw, and Klein (1994), who provided a volume that includes research based on the International Self-Report Delinquency study.

Finally, a few studies that do not fit exactly into these categories but that may be of interest to some scholars of comparative criminology include a review by Rummel (1994) that addresses several issues of genocide in the 20th century, including which types of regimes might be more prone to this behavior; Lester's (1994) study of interpersonal violence in bellicose nations, which examines the relationship between nations' levels of participation in wars and their levels of interpersonal violence; data released by the United Nations International Drug Control Programme (UNDCP) (1997) on the supply of and trafficking in narcotics around the world; and Farrell, Mansur, and Tullis' (1996) use of UNDCP's

data in a cross-national comparison of cocaine and heroine prices and trafficking in Europe.

The advantages and disadvantages of this methodological approach are widely discussed in the literature on comparative criminology, so they are summarized only briefly here. As for advantages, first, both explanatory and crime data for nations are becoming increasingly available to researchers. Second, descriptive studies employing nations as units of analysis can illuminate patterns and trends in violent, property, and other types of crime throughout the world. For example, do certain areas or countries of the world exhibit significantly higher or lower crime rates than elsewhere? Are these rates undergoing significant increases or decreases over time? Finally, by testing the rigor of the theoretical paradigms developed to explain these differences and/or trends, we are able to draw conclusions about national-level correlates of crime that add to the criminological literature that previously was based on work completed in a single nation or culture.

The main disadvantages of this type of work revolve around aggregation and the validity of data. First, aggregating to such a high level of abstraction presents many problems for researchers. Most importantly, it masks what is likely to be significant spatial variation in both crime and explanatory factors throughout a nation. Similarly, with available data only about the attributes of the nation as a whole, researchers are unable to recognize the more proximate causes of crime within the country. Second, measurement issues can present serious threats to validity. It is often difficult to construct precise operational measures of theoretical constructs at this level, especially because cross-national researchers usually depend on secondary data collected by governments for administrative, not scientific, purposes. Even if we are able to find a measure that closely corresponds to the theoretical elements, the likelihood of obtaining valid measures for a broad range of nations throughout the world (based on factors such as geographic location, level of development, type of government, or predominant religion) are slim. One way to overcome these disadvantages is to look more closely at the internal workings of criminal justice systems and structural and cultural contexts of crime within individual countries. Parallel studies have succeeded to some degree in achieving this.

Parallel studies

Parallel studies generally focus on a close analysis of the criminal justice systems or the nature of crime within two nations. They may be divided into three general subtypes.

Crime rate/criminal justice system analysis. Many studies compare in detail crime rates or other types of generated official statistics about crime or criminal

justice systems in two nations. A good example of this type of study is Downes' (1988) examination of incarceration rates as a function of the penal policies of The Netherlands, England, and Wales. This study required attention to the fine detail of the complex legal and bureaucratic processes that produce incarceration rates. A more recent example is that of McClintock and Wikstrom (1992), who initially compared crime rates between Scotland and Sweden, followed by a study comparing violent crime between Stockholm and Edinburgh.

In most cases, these types of studies are able to overcome the many difficulties of comparing official statistics across countries in the larger cross-national studies reported elsewhere in this chapter. Differences in police recording of offenses can be noted and accounted for and differences in legal procedures and definitions of crimes examined. When only two countries are examined, more meaningful comparisons can be drawn and explanations for similarities and differences in crime rates convincingly made. Another example of this approach is a study by Langan and Farrington (1998) that employed official data to compare crime rates and other criminal justice statistics between the United States and England and Wales. Through the creative use of official statistics as well as victimization surveys in the two countries, these researchers were able to make persuasive conclusions concerning the comparative levels of particular types of crime and other aspects of the criminal justice system—such as incarceration rates—that are notoriously difficult to compare cross-nationally.

Topical comparison. Here, researchers generally follow a particular perspective or approach concerning the same topic or social problem in two countries. Studies of this kind are often anthropological and/or historical. An excellent example of this approach is a study by George DeVos (1980), in which he studied the minority status of delinquents in Japan and compared it with the minority status of delinquents in the United States. The questions driving the study were why Japanese-Americans had a very low rate of delinquency and how this may be related to delinquency rates in Japan. DeVos used a variety of anthropological observations in each country and combined them with various psychological assessment instruments administered in both nations as a parallel methodology. Other studies include Zehr's (1976) study of crime and development in 19th-century Germany and France; Gurr, Grabosky, and Hula's (1977) work on violence in four cities in a historical perspective; Bayley's (1976) comparison of policing in Japan and the United States; Kaiser's (1984) study of prison systems and correctional law in the United States and Europe; and Bouloukos' (1999) comparison of human rights and the law in the field of incarceration.

Replication of an experimental design. Studies that replicate an experimental design in two countries are rare, no doubt because they require considerable

coordination. An example of this type of study is Friday, Yamagami, and Dussich's (1999) construction of a questionnaire to measure the threshold in perceptions of violence among respondents in Japan and the United States. The samples were drawn according to the same research design, and the questionnaire was constructed with reference to the differing cultural requirements of each nation. Researchers attempted to carry out the studies during the same time period to the greatest extent possible. Similarly designed studies have been conducted studying policing behavior (Ivkovic and Klockars 1996).

In sum, the central advantage of parallel studies is that meaningful comparisons between individual nations can be made while controlling for the many known factors that may lead to spurious conclusions in comparative criminal justice. Each of the three types of parallel studies described here attempts to achieve this in a different way. Although ideally the third type, the experimental design, promises the most scientific control, it requires a narrow definition of the issue addressed, thus limiting its generalizability. Of course, this is a standard problem of any experimental design employed in social science. Conversely, topical comparisons and crime rate/criminal justice analyses offer excellent ways to develop comparisons, but they may do so at the risk of somewhat divorcing the subject matter from the overall context of the nation and culture in which the problems occur. Case studies seek to overcome this disadvantage.

Case studies

By the term case study, we mean research undertaken in a single nation. This type of work is usually, though not necessarily, qualitative and descriptive in nature and often includes a historical element. Others have suggested that work done in a single country has "no obvious comparative intent" (Beirne and Hill 1991, viii) and thus do not include this methodological approach in their definition of comparative criminology. Even when no direct comparisons are made with other nations, however, descriptions of crime or criminal justice within a single country obviously increase our knowledge about these subjects throughout the world. Thus, we include a brief discussion here.

As in the previous sections, these examples are chosen to represent a wide array of comparative research by criminologists and others. First, there are a few books that make a good starting point for single-country studies of crime and justice. Heiland, Shelley, and Katoh's (1992) volume presented essays on crime in several countries around the world, including socialist, developing, and developed nations. More recently, Barak's (1999) volume provided chapters on crime and its control in an extended list of nations that are geographically, economically, and politically diverse. Finally, the United Nations *Global*

Report on Crime and Justice (Newman 1999) included both crime data and topical essays on a broad range of crime and justice issues throughout the world.

The impact of development on crime is a popular topic for researchers performing case studies. One example is Skinner's (1986) study of development and crime in Iceland, which revealed a lower crime rate in that country than in other modernized nations; the author attributes this to Iceland's culture of egalitarianism and its low unemployment rate. Other examples include Hatalak, Alvazzi del Frate, and Zvekic's (1998) volume that reported on ICVS findings from nations in transition; Zvekic's (1990) volume on crime and development; and Arthur and Marenin's (1995) essay on crime in developing nations, which recognized the difficulties involved in this type of research and called for a case study approach in order to understand country-specific experiences with crime and development. The series of volumes produced by the European Institute for Crime Prevention and Control has also, for a number of years, supplied case study material as a background for interpreting larger scale quantitative analysis.

Abel (1987) provided an annotated bibliography with many references to single-country studies of homicide. Other examples of case studies on violence include examinations of alcohol and homicide in Copenhagen (Gottlieb and Gabrielsen 1992); patterns of homicide in Greece (Chimbos 1993); an indepth examination of several violence-related issues among American Indian populations (Bachman 1992); a study of the effects of the drug trade on violence in Brazil (Zaluar and Ribeiro 1995); and an examination of the spatial, temporal, and demographic variation of homicide rates throughout Russia (Pridemore 1999). Examples of research on sexual assault and rape include studies of wartime rape in Yugoslavia (Stojsavljevic 1995) and a survey of survivors of sexual assault in Australia (Eastal 1994). Studies of spouse and child abuse in different nations and cultures can provide insight into the variation of cultural norms concerning the status of women and children within society. Examples include research on family violence in Canada (DeKeseredy and Hinch 1991), South Africa (Adams and Hickson 1993), India (Natarajan 1995), Zimbabwe (Khan 1995), and Russia (Gondolf and Shestakov 1997) and of Chinese immigrants in the United States (Chin 1994).

Case studies that include research on property crimes include Wu's (1995) look at declining gender differences in crime in Taiwan, Arthur's (1992) study of social change and crime rates in Puerto Rico, Helal and Coston's (1991) examination of Islamic social control and low crime rates in Bahrain, the results of victimization surveys in Estonia and Finland (Aromaa and Ahven 1993), research on school crime in Sweden (Lindstrom 1997), and rising crime rates in Switzerland (Niggli and Pfister 1997).

Juvenile delinquency is another popular topic among comparativists, and case studies of delinquency throughout the world include Mutzell's (1995) examination of troubled youths, drug abuse, and violent crime in Stockholm, as well as Hartjen and Priyadarsini's (1984) indepth study of delinquency in India. Finckenauer (1995) has done extensive research on Russian youths, and Pridemore and Kvashis (1999) provided a discussion of social problems and juvenile delinquency in post-Soviet Russia. Rechea, Barberet, and Montanes (1995) presented findings about Spain from the International Self-Report Delinquency survey; and Murty, Al-Lanqawi, and Roebuck (1990) explored self-reported delinquency of Kuwaiti males. Other examples include case studies of juvenile delinquency in Korea (Lee 1990), Cameroon (Ade 1995), Japan (Miyazawa and Cook 1990), and Nigeria (Ogunlesi 1990) and the impact of antiamphetamine laws on juvenile offending in Taiwan (Wu 1996).

Finally, some case studies do not fit exactly into any one of the previous categories but are likely to be of interest to comparative criminologists, including Morales' (1986) examination of how the drug trade is affecting the social organization and culture of peasants in the Peruvian Andes; the effects of warfare on interpersonal violence in Israel (Landau and Pfeffermann 1988) and Japan (Lunden 1976); the sex trade directed at international tourists in Southeast Asia (Fish 1984); and Maria's (1990) volume on parallel economies (i.e., the black market) in Marxist states.

The obvious advantage to this methodological approach is its contextual analysis of one nation or culture. This type of work is able to incorporate a deeper understanding of subnational processes, as well as historical- and cultural-specific information on the country or culture under study. Thus, case studies are best situated to avoid the cultural imperialism tag sometimes aimed at comparative criminology. Similarly, researchers are able to examine the effects of a significant event on a nation—such as the transition toward a free-market economy or a change in political regimes—on crime rates or types of crime being committed. This approach is invaluable as both an exploratory tool that can lay the foundation for more statistically sophisticated work and a mechanism for providing contextual information that quantitative analysis simply is unable to supply. Further, indepth study of one country can result in data disaggregated to a level lower than that of the nation, presenting researchers with the ability to examine the reliability of theories developed to explain crime in the United States or other Western nations.

The main disadvantage of this methodological approach is the inability to generalize findings to a broader population. Disaggregated empirical studies of a nation can test the reliability of theories in disparate settings, but most case studies are qualitative in nature and employ unique definitions and measures of

deviance, crime, and justice. Thus, although such work is contextually rich, it is difficult to replicate elsewhere or compare it directly with studies undertaken in other nations. A final disadvantage is the fault of comparative criminologists ourselves, not the method. Other disciplines, most notably political science and anthropology, have created wide-ranging literatures that directly or indirectly take on subjects such as deviance, justice, culture, and social organization in scores of different nations and cultures that are relevant to our field. Unfortunately, however, disciplinary boundaries often keep us from incorporating the results of this work into our own.

So far, we have discussed the major theoretical approaches employed in comparative criminology as well as the main methodological strategies that have been used to investigate crime in the cross-national setting. Our aim in the following section is to identify some of the primary sources of data that are likely to be of greatest use to comparative criminologists.

Sources of Cross-National Data for Research on Crime and Criminal Justice

At the present time, comparative criminologists can draw on a wide variety of data sources to inform their investigations. Official crime and criminal justice data are collated and disseminated by a number of international organizations, and researchers can also retrieve official data directly from national statistical agencies. Moreover, recent years have seen the development of victimization and self-report data collection efforts at the cross-national level. We will discuss these sources of information on crime and criminal justice in the coming pages. Of course, most comparative criminologists are interested in testing theories about the nature of crime and the social response directed at it by nations around the globe. In other words, these investigators also desire an assembly of explanatory variables at the nation-state level. In the last pages of this chapter, we will indicate some of the best sources for political, social, and demographic indicators about nations around the world.

Data on crime and criminal justice

Official data

There are three main sources of official crime data at the cross-national level. First, for those researchers interested in a specific country or for comparative analyses of a small set of countries, the best approach is to gather information

directly from the nations themselves. Although the problems of using official data remain (on the difficulties associated with official data, see Newman 1999), this strategy allows the researcher to become more familiar with the definitions of specific crimes and the idiosyncrasies associated with data collection in each nation. This clearly presents the investigator with a more thorough understanding of measures and promises stronger, more informed research. Familiarity with these agencies and their data also allows the researcher access to a broader range of information, including data aggregated to subnational levels.

These agencies are not always accessible to social scientists, however, especially if researchers wish to include a large number of nations in their study or if they are not proficient in the languages of the target countries. In these cases, investigators can take advantage of two other sources of cross-national databases on crime: Interpol and the United Nations crime surveys. They are outlined briefly here (for a thorough review of cross-national crime and explanatory data as well as information about gaining access to these data, see Neapolitan 1998).

Interpol has collected crime statistics from its member countries since 1950 and now publishes them annually in its publication *International Crime Statistics*, which contains information from approximately 100 countries each year (Interpol 1995). A standard form is sent to each country, with instructions provided in French, English, Spanish, and Arabic. Data provided to Interpol represent police and judicial statistics and are limited in scope. No attempts are made to evaluate the validity of the data. For this reason, Interpol is clear in publications that its data should not be used as a basis for making comparisons among nations. The Interpol database contains information on the volume of crime and the persons responsible for these offenses in general categories of crime, including murder, assault, robbery, burglary, fraud, and drug and sex offenses. Volume of crime figures include the total number of cases known to the police, the percentage of these cases that are attempts, the percentage of the total number of cases solved, and the rate of each offense per 100,000 population. Interpol also requests information on the total number of offenders and the percentage of known offenders who are females and/or juveniles¹¹ and who are not citizens of a particular country.

A second, more extensive source of official crime data is provided by the United Nations Survey of Crime Trends and Operations of Criminal Justice Systems (UNS). Five surveys have been completed thus far, covering the periods 1970–75, 1975–80, 1980–86, 1986–90, and 1990–94. These surveys gather qualitative and quantitative information on crime and criminal justice systems in the member countries of the United Nations, with the goal of improving the dissemination of this information to a global audience of researchers and administrators (Joutsen 1998; United Nations Criminal Justice Information Network 1999).

Although not all countries reply to each survey and no country answers all the questions on the form, the number of countries responding to the questionnaire rose in the fourth and fifth surveys, and the United Nations urges each nation to return the survey even if it is able to provide only limited information.

Given the nature of the data collection process, the material provided by UNS comes from the official criminal statistics of each nation. Thus, the origin of this information is similar to that from the Interpol database. However, because nations report these data to the United Nations (of which each is a member and by which each has been urged to respond to the questionnaire) whereas Interpol collects its information from police chiefs, the UNS data might be considered a more official statement by each nation about crime and its criminal justice system (Neapolitan 1997; Newman 1999). The survey's section on crime includes items related to intentional and nonintentional homicide, assault, rape, theft, robbery, burglary, fraud, embezzlement, drug-related crimes, bribery, and corruption (United Nations Criminal Justice Information Network 1999). Among other things, the survey asks for the number of crimes recorded by police and the age and gender of arrestees.

Unlike Interpol, the collectors of UNS data employ several methods of validating the information they receive as well as making the survey as user friendly as possible. First, in an attempt to minimize errors due to cross-national differences in categories of crime, the United Nations provides a standard definition for each crime and, although few countries respond to the query, they are asked to note any discrepancies between their definition and that of the United Nations. Second, if there is a 30-percent change in any reported number (e.g., the number of rapes or burglaries) from year to year or if the numbers at one stage of the criminal justice system do not match those at another (e.g., considerably more people were admitted into prison than were arrested), the United Nations contacts the reporting agency to account for these potential inaccuracies. The countries are also encouraged to report any situations, such as wars, political turmoil, or accounting practices, that might be responsible for significant changes. Finally, the United Nations reviews its experience with the process after each survey, as well as comments from the reporting agencies in each country, to improve on future questionnaires.

Researchers wishing to employ these two sources of information on criminal behavior across nations face not only difficulties common to all official crime data but also unique problems associated with the cross-national nature of the sample. First, no matter which database is chosen, researchers are in effect working with little more than a convenience sample (i.e., we can only use data for those nations that respond to Interpol and UNS queries), and it is likely that the sample of nations used will affect the results of cross-national comparisons

(Kohn 1989; Neuman and Berger 1988). Second, only national-level trends and relationships can be detected with this type of data. We can assume that aggregation to such a high level will mask the tremendous variation in crime rates and their social-structural correlates likely to be present throughout a country (Lynch 1995a). This has led many researchers to call for country-specific research that is capable of analyzing these issues at subnational levels (Archer and Gartner 1984; Arthur and Marenin 1995; LaFree and Kick 1986; Neapolitan 1997). Third, police data result from legal standards and administrative needs and practices, which can diverge extensively among different cultural, political, and economic systems. Policing agencies also vary in their level of professionalism, efficiency, recording procedures, and ability to collect data from all jurisdictions within their respective countries, making it difficult to ensure comparability in these data across nations.

For both administrative and cultural reasons, the number of crimes that come to the attention of the police and the category in which they are coded are also likely to vary. For example, although some crimes such as homicide and theft are consistently defined and perceived as serious across cultures (Kick and LaFree 1985; Scott and Al-Thokeb 1977), not all categories of crimes (or even subcategories of homicide and theft) are comparable cross-nationally. What is an assault? Are attempts included in the reported frequencies? At what age does a juvenile become an adult? Are data presented for reported crimes, recorded crimes, or arrests? These are fundamental questions that are not easily answered, and research has found that tests for relationships may yield divergent results depending on the database (e.g., Interpol or UNS), the crime category, and the level of analysis employed (Bennett and Lynch 1990; Huang and Wellford 1989).

Also, similar issues are likely to affect the rate at which victims report crime to the police. Given cultural variation, citizens' reporting of less serious victimizations, as well as the police response to these victimizations, probably varies considerably among nations (Lynch 1995a; Vigderhous 1978). Research in the United States has shown that reporting rates differ depending on the community's trust in the police (Biderman and Lynch 1991), and people's trust in the police likely varies throughout the world as well as within individual nations (Block 1984). Similarly, the status of women and the cultural response to sexual victimization is also likely to create differential reporting rates of rapes and

Policing agencies also vary in their level of professionalism, efficiency, recording procedures, and ability to collect data from all jurisdictions within their respective countries, making it difficult to ensure comparability in these data across nations.

sexual assaults (Ali Badr 1986; United Nations 1995). In fact, victimization data reveal differential reporting among nations in several crime categories, including assault, theft, and sexual offenses (van Dijk and Mayhew 1993).

Victimization surveys can also provide more contextual information concerning the nature of the criminal event, as well as victim attributes, their fear of crime, and their experiences with and view of the criminal justice system.

Victimization data

As just discussed, researchers experience several difficulties with official data when attempting to determine the etiology of crime. Not only are these data predicated on legal, rather than behavioral, definitions of harmful actions that can vary from nation to nation, but police data are constructed and maintained for administrative, not scientific, purposes and they do not include the large number of crimes that do not come to the attention of police. In response to these difficulties, victimization surveys have been undertaken in several countries in an attempt to gain a more accurate picture of the extent of criminal behavior. These victimization surveys can also provide more contextual information concerning the nature of the criminal event, as well as victim attributes, their fear of crime, and their experiences with and view of the criminal justice system.

As with official data, researchers interested in one or a few countries should collect information from country-specific victimization surveys. Again, this ensures that the instrument is more culturally relevant and provides information to the researcher that may not be available elsewhere. Victimization surveys are a relatively new and evolving tool, however, so the availability, reliability, and validity of country-specific surveys are limited. Likewise, if the researcher wishes to compare a large set of countries, then the variation in the samples drawn, the questions asked, and the methodologies employed make country-specific surveys unwieldy. In this case, the best available instrument is ICVS, which employs a standard survey instrument with all respondents in each of the participating nations. Beyond the goals it shares with other victimization surveys, ICVS aims to (1) be sensitive to each nation's unique experience with crime while at the same time providing an appreciation for shared patterns of and problems with criminal behavior; (2) provide the administrators and policy-makers of participating countries with valuable information so they can make informed decisions; and (3) provide social scientists with an alternative source of crime data with which they can track trends and test theories of crime causation (Zvekic and Alvazzi del Frate 1995).

The first ICVS was completed in 1989 and involved 17 nations, most of them developed (van Dijk, Mayhew, and Killias 1990). Subsequent surveys in 1992 and 1996 increased the number and type of nations included. In 1996, for example, several countries in transition and developing countries participated in the survey. In most developed nations, the samples are drawn from the whole country and computer-assisted telephone interviews are completed (Kangaspunta, Joutsen, and Ollus 1998). Results from these nations are weighted to make them as representative as possible. In developing countries and countries in transition, the samples are normally drawn from the largest (most often the capital) city, the households in the sample are chosen through random walk techniques, and interviews are conducted face to face. Finally, the sample size varies from country to country but with few exceptions, usually consists of at least 1,000 people. Exhibit 1 lists the countries that have participated at least once, the years they participated, the population from which the sample was drawn, and the sample size for each.

ICVS queries one respondent from each household in the sample about crimes affecting the household in general and about victimizations that he or she has personally experienced (see exhibit 2 for a list of the crimes about which each respondent is asked). The timeframe involves the 5 years leading up to the time of the interview; those respondents who report victimizations are asked to provide further details about the event. Demographic data are obtained from the respondents, and they are also asked a series of questions concerning their attitudes toward police and actions taken to protect against victimization.

Although ICVS provides a much-needed alternative to official data and is able to uncover important information about victims and criminal events that official crime data cannot, it still faces several challenges. First, as with any victimization survey, it is likely that issues such as respondent memory decay and telescoping will present difficulties (Block 1993; Lynch 1993; Skogan 1986), especially since the interviews are not bounded and respondents are asked to recall victimizations for a timeframe of 5 years. Respondents' willingness to reveal sexual victimizations is also likely to be a delicate issue (Skogan 1981), and this willingness is likely to vary across the sample of nations based on cultural norms. In fact, Zvekic and Alvazzi del Frate (1995) revealed complications with validity in measuring these types of victimization with ICVS.

Sampling issues and interviewing techniques present concerns as well. First, ICVS covers a limited, although growing, number of nations. Second, Killias (1990) argued that sample sizes of at least 5,000 are needed in most European nations given the relatively low victimization rates in these countries. The ICVS sample sizes are only a fraction of this number (see exhibit 1). Third, in most developed countries, nationwide samples are drawn, whereas in developing

Exhibit 1. Nations participating in the International Crime Victimization Surveys

Nation	Years of participation	Survey type	Sample size
Albania	1996	City	1,200
Argentina	1992	City	1,000
Austria	1996	National	1,507
Belarus	1997	City	999
Belgium	1989, 1992	National	2,060; 1,485
Brazil	1992	City	1,000
Bulgaria	1997	City	1,076
Canada	1989, 1992, 1996	National	2,074; 2,152; 2,134
China	1992	City	2,000
Costa Rica	1992	City	983
Croatia	1997	City	994
Czech Republic	1992, 1996	National	1,262; 1,801
Egypt	1992	City	1,000
England and Wales	1989, 1992, 1996	National	2,006; 2,001; 2,171
Estonia	1992, 1995	National	1,000; 1,173
Finland	1989, 1992, 1996	National	1,025; 1,655; 3,830
France	1989, 1996	National	1,502; 1,003
Georgia	1992, 1996	City	1,395; 1,137
Germany (West)	1989	National	5,274
Hungary	1996	City	756
India	1992	City	1,000
Indonesia	1992	Several cities	4,550
Italy	1992	National	2,024
Kyrgyzstan	1996	City	1,750
Latvia	1996	City	1,411
Lithuania	1997	National	1,176
Macedonia	1996	City	700
Malta	1997	National	1,000
Netherlands	1989, 1992, 1996	National	2,000; 2,000; 2,008
New Guinea	1992	Three cities	1,583
Northern Ireland	1989, 1996	National	2,000; 1,042
Norway	1989	National	1,009
Philippines	1992	City	1,503
Poland	1992, 1996	National	2,033; 3,483
Romania	1996	City	1,091
Russia	1992, 1996	City	1,002; 1,018

Exhibit 1 (continued)

Nation	Years of participation	Survey type	Sample size
Scotland	1989, 1996	National	2,007; 2,194
Slovakia	1992, 1997	National, city	508; 1,105
Slovenia	1992, 1997	City, national	1,000; 2,053
South Africa	1992	City	1,000
Spain	1989, 1993, 1994	National, city, city	2,041; 1,634; 1,505
Sweden	1992, 1996	National	1,707; 1,000
Switzerland	1989, 1996	National	1,000; 1,000
Tanzania	1992	City	1,004
Tunisia	1992	City	1,150
Uganda	1992	City	1,023
Ukraine	1997	City	1,000
United States	1989, 1992, 1996	National	1,996; 1,501; 1,003
Yugoslavia	1996	City	1,094

Source: Kangaspunta, Joutsen, and Ollus 1998, 194–195; Zvekić and Alvazzi del Frate 1995.

countries and countries in transition, samples are usually drawn only from one large city. Fourth, it is well known in survey research that nonrespondents usually vary from respondents in several important aspects, including levels of victimization (Block 1993). ICVS is likely to be especially prone to this problem because response rates vary from nation to nation. Finally, interviewing techniques vary across the sample of nations included. In developed nations, computer-assisted telephone interviewing (CATI) strategies are employed. Not only has CATI been shown to reveal a greater number of victimizations than other techniques (de Leeuw and van der Zouwen 1988), but it is also unable to reach potential respondents who do not have telephones (and who are likely to differ in their levels of victimization from those who do have telephones). Even given these problems, however, ICVS provides an alternative to official crime data and should provide valuable insight into our understanding of crime and victimization across cultures. Several scholars (see Block 1993; Lynch 1993) agree that, if used wisely, ICVS is capable of supporting cross-national comparisons.

Health data

Another source of limited victimization data is derived from health statistics. For instance, homicide is a popular topic among comparativists for several

Exhibit 2. Household and personal victimization in ICVS

Household victimizations	Personal victimizations
Theft of cars Theft from cars Vandalism to cars Theft of mopeds/motorcycles Theft of bicycles Burglary with entry Attempted burglary	Robbery Theft of personal property Pickpocketing Noncontact personal thefts Sexual incidents Sexual assaults Offensive behavior Assaults and threats Assaults with force Assaults without force (threats)
Source: Kangaspunta, Joutsen, and Ollus 1998, 190.	

reasons. First, the level and causes of lethal violence are salient issues for any nation. Second, many believe homicide rates to be representative of the level of criminal violence in general (Fox and Zawitz 1998). Third, measures of homicide are commonly accepted as the most valid of all crime indicators. Even so, measurement error in homicide counts from official criminal records varies extensively from country to country. One way researchers interested in homicide handle this issue is to use mortality data.

Vital statistics agencies in most countries employ International Classification of Diseases (ICD) codes to record the cause of death on each death certificate. One of ICD's external causes of death is "homicide and injury purposely inflicted by other persons" (World Health Organization 1996), with subcategories that include the manner of victimization, such as strangulation, poisoning, or the use of firearms or cutting instruments. If researchers wish to find this information at subnational levels, they must gain access to public health agencies within specific countries. At the national level, however, these figures are collected regularly from each country by the World Health Organization (WHO) and are available in its *World Health Statistics Annual*, which currently contains data for more than 80 nations. As shown in exhibit 3, the difference between police and mortality data may be relatively small in some countries. In the United States, for example, Rokaw, Mercy, and Smith (1990) reveal that the Federal Bureau of Investigation's Uniform Crime Reports annually report about 9 percent fewer homicides than mortality data from the National Center for Health Statistics; in the 1994 data presented in exhibit 3, the difference is only 5 percent. In other countries, however, discrepancies between police and mortality data may

Exhibit 3. Comparisons of selected nations' 1994 homicide counts from Interpol, the United Nations Crime Survey, and the World Health Organization

Nation	Interpol	UNCS	WHO
Australia	NA	196	323
Austria	88	88	94
Azerbaijan	496	499	4,620
Bulgaria	499	492	428
Canada	596	596	498
Chile	1,545	626	410
Colombia	NA	27,079	27,620
England and Wales	729	726	373
Greece	133	133	119
Israel	114	140	119
Japan	711	695	789
Kazakhstan	NA	2,664	2,985
Lithuania	NA	465	497
Nicaragua	733	549	241
Northern Ireland	104	83	100
Republic of Korea	653	577	720
Romania	NA	776	1,008
Russian Federation	29,897	29,913	47,870
Singapore	50	50	43
United States	23,310	23,330	24,547

Note: NA means data are not available. Interpol data are defined as "voluntary homicides." UNCS data are defined as "total intentional recorded homicides." WHO data are defined as "homicide and injury purposely inflicted by other persons."

Sources: Interpol 1995; United Nations Criminal Justice Information Network 1999; World Health Organization 1995, 1996.

be quite large. For example, Pridemore (1999) showed that during the 1990s, official crime data reported by the Russian Federation's Ministry of the Interior have annually recorded only about two-thirds the number of homicides reported by the Ministry of Public Health.¹² In the 1994 data presented in exhibit 3, this amounts to an absolute difference between the two of nearly 18,000 deaths.

One benefit of mortality data is that they likely provide a more valid representation of the level of lethal violence within a nation because they reflect medical

decisions about cause of death rather than police practices. Another advantage is that the age group¹³ and sex of the victims are provided. This allows the researcher to recognize unusually high or low rates of violent victimization among specific age and sex categories, for example, or to test elements of different theories of victimization.

These data, however, possess problems of their own. First, the level of medical expertise of those making the decision affects their ability to determine the exact cause of death, including homicide. This varies not only from country to country, but within nations as well. In the United States, for example, the coroner is a locally elected position in some jurisdictions; in these areas, the post is open to anyone who wishes to run for the office, regardless of medical training and experience (Baden 1998). As noted by Neapolitan (1997), a second problem is that developing nations, especially those in Africa and Asia, are underrepresented in WHO reports, making it difficult to include them in cross-national analyses.¹⁴ Finally, although not dependent on the recording practices of the police, the collection of mortality data is an official process and the agencies gathering this information face not only their own unique difficulties but also pressures common to any bureaucratic organization. However, overall mortality data probably provide a better measure of lethal violence than police data for most purposes, and they are commonly used for cross-national research on violence.

Self-report data

As shown in the previous section, victimization surveys present an alternative approach to measuring crime and provide the researcher with information that is not available from official data. However, these surveys are usually unable to capture much information about offenders, nor can they provide insight into victimless crimes. One remedy to these informational barriers is the self-report survey, which samples the population and asks respondents to provide information concerning their own offending behavior. Survey construction and testing, sampling, and interviewing are time- and cost-intensive procedures, however, and the self-report survey is a research methodology associated with individual studies rather than an instrument used to collect data on crime, delinquency, and offenders at the national level. This means that, although the self-report is a fairly common technique around the world (especially with juveniles), differences in samples and survey questions among different studies, as well as varying cultural definitions of crime and deviance from country to country, make it difficult to use the results gained from these surveys to make cross-national comparisons.

It was not until the late 1980s, in fact, that an attempt was made to construct a self-report survey that could be administered in several nations. The International Self-Report Delinquency (ISRD) study has now been through two sweeps, completing self-report surveys of juveniles in a limited number of nations¹⁵ (Junger-Tas, Terlouw, and Klein 1994; Klein 1989). The sampling strategy differs from one country to the next, but a standard questionnaire, modeled on the National Youth Survey (Elliott, Huizinga, and Ageton 1985), is employed. The survey obtains data about both the respondents—including sociodemographic attributes and information about family, school, and peers—and their delinquent acts, such as incidence and prevalence, contextual information about each event, and the reactions of others to the delinquent acts (Junger-Tas, Terlouw, and Klein 1994). Although a welcome addition and useful in several respects, the limited number of and similarity among the nations involved and the varying sample frames and response rates from one nation to the next make the ISRD study of limited value for sophisticated cross-national examinations of delinquent behavior. However, several nations, including a few Central and Eastern European countries in transition, have adopted the ISRD's survey instrument for their own use (Neapolitan 1997). Thus, self-report surveys at the cross-national level may eventually prove to be a valuable resource on offenders and patterns of delinquency.

Explanatory data

There is no room here for a thorough evaluation of all the different types of explanatory data available to criminologists. There is, however, one pressing issue that must be addressed concerning these data. Perhaps *the* major detriment to the systematic gathering of knowledge about the etiology of crime across nations is the lack of careful operationalization of theoretical models and consistency in measurement models. It seems that, once criminologists have what we consider to be a valid indicator of crime, we collect any available data to employ as representations of our theoretical concepts. It is vital that all of the effort we have placed in being careful about our measurement of crime must also be directed toward the definition, operationalization, and measurement of our independent variables. This is especially the case at the cross-national level, where different national governmental and international data collection agencies are likely to calculate specific socioeconomic statistics (e.g., poverty level, unemployment, inequality) in different manners. This presents a serious threat to

Perhaps the major detriment to the systematic gathering of knowledge about the etiology of crime across nations is the lack of careful operationalization of theoretical models and consistency in measurement models.

our models and must be taken seriously. We must also strive to agree on the best measures of theoretical elements and use these measures in each study if possible.¹⁶ If we do not, then we cannot truly compare the findings from each study nor treat them as replications of earlier work; this represents a major obstacle to a systematic understanding of crime causation across nations. Thus, stronger efforts must be made to match theoretical and measurement models when embarking on cross-national studies of crime causation.

The good news is that the accessibility, reliability, and validity of explanatory data at the cross-national level are increasing, with several sources available with which to cross-check the reliability of the information. This creates another risk, however, in that it is easy to become variable oriented in our study of crime instead of allowing theory to drive our research. Having introduced these caveats, the rest of this section briefly discusses several sources of cross-national explanatory data.

Census Bureau

The International Programs Center of the U.S. Census Bureau maintains an International Data Base (IDB) that contains useful information for criminologists on up to 227 countries throughout the world (U.S. Bureau of the Census 1999). IDB is available online and is easy to access. Given the nature of the Census Bureau's mission, most of the data is demographic in nature, and the extent of the data varies for each nation, but it contains measures relevant to different theoretical approaches, especially social disorganization. Major categories of information available from IDB, together with specific variables that might be helpful in testing criminological theories or using as controls, are listed in exhibit 4.

As mentioned, this information is not only especially helpful as a source for demographic variables often used as controls, it also is helpful for possible measures of structure-level social disorganization. For example, information on migration might be used as an indicator of mobility at the national level. Percentage urban might be employed as a proxy for population density. Information on marital status and households—such as proportion of the population that is married, single, or separated or divorced, as well as single heads of households—are amenable to use as measures of family structure. Poverty rates, income measures, and employment data are available to represent the poverty element. Heterogeneity indicators are available in terms of ethnicity, religion, and language.

Exhibit 4. Major categories and specific information relevant to criminological research provided by the U.S. Census Bureau's International Data Base

Data category	Information available
Population	<ol style="list-style-type: none"> 1. Total population 2. Percentage urban 3. Population by age, sex, and urban/rural residence
Vital rates, infant mortality, and life tables	<ol style="list-style-type: none"> 1. Net migration rate 2. Infant mortality rates by sex 3. Life expectancies at birth by sex
Migration	<ol style="list-style-type: none"> 1. Estimated net number of migrants by sex, age, and urban/rural residence 2. Migration rate by sex, age, and urban/rural residence
Marital status	<ol style="list-style-type: none"> 1. Population by marital status, age, sex, and urban/rural residence 2. Population that is single, married, and separated or divorced, in both absolute and percentage terms
Ethnicity, religion, and language	<ol style="list-style-type: none"> 1. Population by ethnic group and sex 2. Population by religious group and sex 3. Population by language and sex
Literacy	<ol style="list-style-type: none"> 1. Population by literacy, age, sex, and urban/rural residence 2. Female literate population in absolute and percentage terms
Labor force, employment, and income	<ol style="list-style-type: none"> 1. Economically active population by age, sex, and urban/rural residence 2. Economically active population by industry and occupation 3. Relative net income measures and poverty rates
Households	<ol style="list-style-type: none"> 1. Heads of households by age, sex, and urban/rural residence 2. Female heads of households; male heads of households

Source: U.S. Bureau of the Census 1999.

The World Health Organization

Health data (usually in the form of mortality statistics) can be important to criminologists as both dependent and independent variables. WHO's *World Health Statistics Annual* reports mortality data for more than 80 nations,¹⁷ while its electronic "WHO Mortality Database" contains data only for those nations that WHO believes provide data of good quality (WHO 1999).¹⁸ These mortality data are based on cause-of-death information provided by each nation's death registration system. Cause of death is reported for each deceased person on a death certificate, according to ICD specifications, and these data are collected and aggregated to the national level before being transferred to WHO. Of most import to criminologists are homicides, infant mortality, life expectancy, and alcohol use.

First, as discussed earlier, homicide counts based on victimization data from mortality statistics are thought to be a better measure of homicidal (though not necessarily criminal) violence in a nation than police data. These data can also be disaggregated by sex and age categories, which is helpful for testing certain theories of violence.¹⁹ Second, both infant mortality rates and life expectancy are considered by some to be indicators of modernization, so those interested in this approach may wish to use one or both of these in their measurement model. Third, the sociological, public health, and epidemiological literatures have found a strong and consistent correlation between measures of health (e.g., levels of infant mortality) and poverty. This has led some researchers to use this indicator as a proxy for poverty. Finally, though rarely used in criminology research, the public health and epidemiological literature often employ the rate of deaths due to cirrhosis of the liver as a measure of alcohol consumption. (These data are available from WHO.) Given the role played by alcohol in violence (both socially and psychologically and in both offenders and victims), it seems that this is a theoretically important concept and that future empirical work should at least employ this measure as a control in causal models, if not explore the relationship directly.

The World Bank and the International Monetary Fund

Both the World Bank and the International Monetary Fund (IMF) provide economic data that may be useful to criminologists employing several different theoretical perspectives. First, the World Bank has an extended list of indicators of economic structure (World Bank 1999) that are available in either its *World Development Indicators* or at its Web site. Researchers interested in the modernization/development thesis may wish to consult the World Bank's measure of long-term structural change, which contains indicators of several relevant concepts, including agricultural output, economic growth and structure,

government finance, labor force and employment, money and prices, and urbanization. Other indicators of development are to be found with energy production and use, GNP per capita, and the distribution of the labor force by occupation or economic activity. Overall quality of life plays a role in several theoretical approaches, and the World Bank employs indicators in several areas to measure it, including education, energy, health, life expectancy, and mortality.

Dependency theorists might find useful theoretical elements in the World Bank's measures of aid dependency, growth consumption, investment, structure of consumption, integration with the global economy, and the "Heavily Indebted Poor Countries Debt Initiative." The World Bank also provides data that could be helpful to those researching gender issues. The structure of the labor force and gender and education measures can be helpful when considering routine activities theories and theories based on the status of women within societies. Data on economic distress can be found in World Bank measures such as the distribution of both income and consumption, unemployment, the magnitude and depth of poverty, and purchasing power parity. Finally, basic demographic information is also provided in the form of population parameters, population dynamics, the absolute size of the urban population, and the percentage of the population living in all urban areas and in urban agglomerations of more than 1 million people.

IMF has recently taken steps to make access to its data more manageable. Their Dissemination Standards Bulletin Board has a Web site that contains economic and financial data reported to IMF from 47 countries. It is important to note that (1) data are not posted unless they subscribe to the Special Data Dissemination Standards set by IMF, and (2) statistical methodology must be documented and sources of reliability cross-checks must be identified (IMF 1999). Although much of these data are of limited use in criminological research, there are a total of 17 data categories that cover 4 sectors of the economy: the real, fiscal, financial, and external sectors. Exhibit 5 includes a list of information available from IMF that could be relevant for criminologists.

Some of these data might be helpful to criminologists, especially to those interested in world system/dependency theory, which is an approach that to this point has been difficult to test empirically. For example, the import and export of goods and services, as well as information on interest payments as a proportion of all expenditures and as the component of debt owed to foreign creditors, might provide dependency theorists with measures of capital penetration into a nation and its placement in the world economic hierarchy relative to other nations. Unfortunately, data are not available for several nations that are probably of most interest to world system theorists. However, data are available from countries in North, South, and Central America; Western, Central, and Eastern Europe; and Asia.

Exhibit 5. Economic data useful to criminologists available from the International Monetary Fund

Real sector	Fiscal sector	Financial sector	External sector
1. GDP in terms of major expenditure categories and productive sectors 2. Employment, unemployment, and earnings 3. Consumer price indexes	1. Domestic and foreign financing 2. Interest payments as a proportion of expenditures 3. Debt by foreign and domestic components	1. Monetary and credit conditions of the banking sector and central bank (including domestic credit broken down by general public and private categories)	1. Imports and exports of goods and services (according to IMF's <i>Balance of Payments</i> manual)

Source: International Monetary Fund 1999.

Conclusion

The goal of this chapter has been to survey the field of comparative criminology to highlight the issues of theory, method, and data that characterize this type of study. We began with a consideration of basic philosophical and definitional issues as we sought to specify the boundaries of the comparative perspective as well as its main concerns. Because firm theory is essential to quality investigations, we next turned our attention to the theoretical explanations for crime that underscore comparative work, offering suggestions for improvement and elaboration along the way. To further scientific studies of crime, researchers must be prepared to translate theoretical concepts into propositions that can be tested against actual observations. This movement from theory to observation requires the criminologist to attend to issues of method and data collection. Accordingly, we considered a variety of methodological approaches to comparative studies and explored the range of dependent variables brought under scrutiny. As we have discussed, there are many available sources of data with which theoretical propositions can be assessed. Still, some theoretical constructs will not find useful indicators in available databases; this means that scholars will have to develop their own data-gathering projects or find a way to make existing data

sources more responsive to theoretical needs. We now end our review of comparative criminology with a few observations that we hope might further the pursuit of comparative work.

Theoretical development and testing

The challenge for comparative criminologists is to develop theories with increased specificity while managing to construct them in such a way that they can be applied across more than one culture or nation-state. This eventually must demand that theories be developed to conceptualize societies as totalities (rather like the work only briefly begun by Parsons 1977) and that theories that manage to provide a world context in which total societies behave be further constructed. The only theory reviewed in this chapter that adopts this approach is world system theory. This theory is not well formulated to explain crime either within or across societies, largely because it was constructed by political scientists who had other interests in mind. As noted earlier, however, there is much in the research and theory of other social science disciplines that could be imported profitably into cross-national research on crime and criminal justice. We believe that theory must be taken more seriously in the context of comparative criminology. With sound theory, the construction of measurement models and the data-gathering process can be carried out in a more valid manner. While expanded and more systematic data-gathering strategies tell us how to collect observations about the world properly, these strategies tell us nothing about what to observe. We need good theory for this.

The testing of theories also depends essentially on the ability of the researcher to specify theoretical concepts and translate them into variables that can be measured. Doing this in the cross-national setting is indeed a challenge. We have noted how difficult it is to construct definitions of crime that can be applied across nations. But recent developments in cross-national data collection have helped to stem these difficulties and provide new opportunities for assessing the reliability and validity of crime data. Turning to the data solutions that have emerged in mainstream criminology in recent decades, comparative criminologists have fielded victimization surveys and self-report surveys to good effect. These approaches have allowed for the construction of behavioral definitions that bypass the legal definitions of each country. This solution must be seen as partial, however, because it does little to remedy shortcomings in the dissemination of official crime data. Improvements on this front might follow if the United Nations strengthened its diplomatic role of coordinating the collection of crime and justice information from nations. (Perhaps other international bodies could adopt the dissemination of crime data as one of their missions.)

But comparative scholars need not rely exclusively on international organizations for information of value to their investigations. We could use more primary observations by students of crime and criminal justice and more efforts to arrange secondary observations into utilitarian forms. Transparency and the World Wide Web permit construction of official databases directly by the researcher. Indeed, entire projects of official data collection could be conceived, then supported by enlightened funding agencies. Certainly, the field could stand an expansion of victimization studies and self-report studies at the international level. Besides making more systematic observations of crime and increasing the variety of crime data, we need to expand the methods of recording and representing this information. The study of criminology would benefit if observations were represented not only as numbers but also as images, sounds, and words (both printed and spoken). This would permit room for filmmaking, videography, oral history, and storytelling in the study of crime. And, with the World Wide Web and its potential to organize the dissemination of information, we have a medium that can traffic in all of these methods of representation at once.

Advancing comparative analysis

An important advance made in recent years in comparative criminology has been a growing sensitivity to the intricate details of the workings of the criminal justice systems in every country. Maybe this seems obvious, but comparative analysis was conducted mainly from a legal perspective for most of this century, preoccupied with particular nuances in legal definitions and procedures. The significance of this change in focus is considerable. The reason is that we now know that it is possible to identify the specific procedures that police and bureaucrats follow that will finally produce crime and justice statistics. The work of Langan and Farrington (1998) has demonstrated that sense can be made of official crime statistics of nations and meaningful comparisons made. This work confined itself to the comparison of two countries, but there is no substantive reason why this methodology could not be extended. The possibility for this research, however, depended on the existence of social scientists who were highly trained in social science methodology and had an intimate knowledge of their respective criminal justice systems. We should expect that there will be more and more researchers, from a wider variety of countries, who are so trained. The increased globalization we have noted in several places in this chapter will surely increase the possibility of more capable scholars coming together. Again, we suggest that an existing international body could advance the study of comparative criminology by ensuring regular meetings of scholars and researchers interested in the subject.

If asked how schools of criminology and criminal justice could advance the practice of comparative work, we offer the most obvious answer: Expand curricular offerings to support such comparative practices. But such offerings must extend beyond survey courses or a smattering of specialty seminars. Those activities will surely help the current state of affairs, but for comparative criminology to truly bear fruit and achieve its potential, we must deepen the repertoire of skills, talent, and knowledge that its practitioners command. To expand students' repertoire, schools could stress the acquisition of foreign language skills and encourage work that employs such skills. Some schools permit students to count language fluency toward methods requirements; this course of study ought to be encouraged for those with comparative interests. Multilingualism permits the comparative student to study crime and criminal justice more richly (see Moore and Fields 1996). A full palette of research methods should certainly be taught, but this course of study must be carefully balanced with a complement of theory-related classes. As a field of study, comparative criminology will be well served if its students are discouraged from grand theory and methodological inhibition and rather pointed in the direction of Mills' (1959) promise of the sociological imagination. Finally, more programs of international exchange between universities should be developed; perhaps the international divisions of the American Society of Criminology and the Academy of Criminal Justice Sciences could facilitate the creation of such programs.

Data explosion

Another significant development in the field of comparative criminology is the incredible explosion of crime and criminal justice information in the international arena. In the space of just 5 years, an enormous amount of information concerning the crime and justice situations in many countries has become readily available. Most major countries of the world now host Web sites that provide a range of statistics and criminal justice information. There are also many international bodies that provide crime and justice statistics on the Internet. Many sites also make available lists of new publications concerning crime and justice in their respective countries. Thus, it is possible not only to obtain more information than ever before, but also to obtain it more quickly, often without regard to crossing national borders.

The ready availability of data will do much to spawn new research. The United Nations database, for example, now contains more than 20 years of crime statistics, which invites researchers to undertake meta-analytical studies. However, with abundance also comes risk, and one primary risk is that these data may be used without regard to their well-known shortcomings (see Newman and

Howard 1999). It is essential that comparative studies are theoretically driven, that theoretical elements are rigorously operationalized, and that the researcher has a thorough understanding of the limitations of the specific data with which he or she is working. Just as important, it will be necessary that such studies be informed, as far as possible, by the various contexts from which these data have been extracted: by the countries and criminal justice systems that have produced them. The work of HEUNI (European Institute for Crime Prevention and Control, affiliated with the United Nations) (Kangaspunta, Joutsen, and Ollus 1998), in which empirical work is backed up by case studies of those countries included in the analysis, appears to be one valuable solution to this perennial methodological problem.

Policy development

In the modern nation-state, policies concerning crime and justice emerge after a highly complex—often adversarial—process of political exchange and bureaucratic procedure. Most such policies originate from issues or social problems that occur within nations; exceptions may be drug trafficking and other kinds of transnational crime. We are inclined to think that there is one more factor that may push or, more accurately, condition policy development: the enormous and rapidly growing availability of criminal justice information. Globalization will make it increasingly difficult for nation-states to ignore the criminal justice information of other countries. Politicians and influential bureaucrats increasingly will be forced to answer as to why their country displays crime rates, prosecution rates, incarceration rates, or rates of violence or gun ownership that are strikingly different from similar countries. Criminologists no doubt will have the opportunity to contribute to policy debates and implementation, provided their work demonstrates relevance and scientific merit.

Globalization and comparative studies of crime and criminal justice

As we enter the 21st century, the field of comparative criminology is in a position to expand and become more rigorous and relevant. Globalization, which has shrunk the world and made social interdependence all the more obvious, will continue to influence the ways in which we lead and understand our lives. Moreover, the increasing mobility of the world's citizens will contribute to a greater desire on the part of these individuals to come to terms with crime and responses aimed at containing it around the globe. Building on established bodies of theory, methodologies, and data such as those described in this chapter, comparative work in crime and justice offers significant promise for the future. Armed with more refined and sophisticated theory that will be capable of

informing better data collection efforts, comparative scholars will place themselves in a position to offer meaningful critiques of international criminal justice policy. While advances in comparative criminological theory and method will not come easily or cheaply, the promise of the endeavor justifies our efforts.

Notes

1. To conserve space, the theoretical discussion is highly abridged. We strongly suggest that readers refer to the original sources for a thorough description of each theory.
2. We employ the categorical designations grand theories and “structural theories” as generic terms used to describe the subtypes of theories discussed here.
3. It should be noted that Shelley (1981), while retaining the notion that social and economic change have a definite impact on crime, later recanted her earlier contention that all nations experience the same phases of development in the same order and at the same pace. Further, she stated that industrialization and urbanization are not always and necessarily accompanied by political and social modernization.
4. An alternative theoretical explanation is Rushton’s (1990, 1995) biological image of crime causation. He contends that constitutional differences between races explain the variation in levels of violence among nations. That is, Rushton argues that variations in genetic predisposition toward violent behavior account for presumably higher rates of violence in African countries, relatively intermediate rates in Caucasian nations, and low rates in Asian countries.
5. Studies that attempt either to include cultural measures or explain empirical differences via cultural variation include Parker’s (1998) examination of how cultural drinking patterns may affect homicide rates and Neapolitan’s (1994) work, which suggests a cultural explanation—due to its colonial past—for the high homicide rates in Latin America. However, the former uses only per capita alcohol consumption, and the latter regional location, as measures of shared culture.
6. This is especially the case because people of different cultures seem to have similar basic definitions of behaviors that are commonly accepted as criminal (Kick and LaFree 1985; Newman 1976) and because research has shown general agreement on the perceived seriousness of harmful behavior across cultures (Evans and Scott 1984; Scott and Al-Thokeb 1977).
7. Another measure of inequality is the Robin Hood index (also referred to as the Pietra ratio). This index is defined as that part of the overall income that would need to be transferred from those with above-average incomes to those with below-average incomes to achieve an equal distribution. Although found in the economic (see Atkinson and Micklewright 1992) and epidemiological literature (see Kennedy, Kawachi, and Prothrow-Stith 1996; Walberg et al. 1998), criminologists have yet to employ this measure of inequality in their research on cross-national crime.

8. Race/ethnic status is also a widely accepted correlate. In cross-national work, however, it is commonly used as a measure of heterogeneity—thus representing a hypothesized causal relationship, not simply an atheoretical correlate—and thus is not discussed in this section.

9. See Gelles (1987) and Korbin (1987) for a discussion of the methodological impediments encountered in cross-national and cross-cultural research on family violence.

10. For groundbreaking work in this area, see Johnson's (1999) survey of domestic violence around the world.

11. Each nation is also asked to provide the age range of its definition of juveniles. Because these definitions differ from country to country, and because each uses its own definition of juvenile when responding to the Interpol questionnaire, this category is not directly comparable across countries.

12. It should also be noted here that Russia, along with several other countries, includes attempts in the homicide category of official crime data.

13. These age groups are: 0, 1–4, 5–14, 15–24, 25–34, 35–44, 45–54, 65–74, and 75 and older.

14. This is not solely the fault of WHO, but also of those nations that fail to respond to the questionnaire or reliably collect this information. Even where data for these nations are available, their inclusion presents problems for cross-national comparisons. For example, variation in the level and availability of emergency medical services among nations might mean the difference between a homicide and an assault. Thus, the same violent event can easily produce two entirely divergent outcomes depending on where it occurs, meaning that the number of homicides in part may be the result not simply of the level of potentially lethal violence but of the level and availability of emergency medical services as well.

15. Countries participating in the first round of surveys included Belgium, Finland, Germany, Greece, Italy, The Netherlands, New Zealand, Portugal, Spain, Switzerland, the United Kingdom (including England, Wales, and Northern Ireland), and the United States (Junger-Tas, Terlouw, and Klein 1994).

16. To highlight only one example, the Gini coefficient employed more often than any other measure of inequality, although justification is rarely supplied for its use over a number of other possible indicators. Also, several of the studies published in the 1970s and 1980s relied on inequality data that were 10 to 20 years old. Further, the theoretical concept of inequality is based on the assumption that an unequal distribution of resources not only is perceived by individuals, but they also feel it to be an injustice and are angered enough by these inequities to resort to crime as an expression of their frustration. This social-psychological process in no way is captured by a measure of the distribution of income. Finally, because income inequality and poverty (or average income) are in actuality two functions of the same distribution, they are usually highly correlated.

If both measures are included in multivariate analysis, the resulting multicollinearity is likely to produce an unstable model, which can bias the coefficients as well as make it virtually impossible to separate the effects of inequality from those of poverty.

17. Much of these data are also available from the United States' National Center for Health Statistics (U.S. Department of Health and Human Services, Centers for Disease Control and Prevention 1999).

18. According to WHO, "Good quality information requires that death registration be near universal, that the cause of death be reported routinely on the death record, and that it be determined by a qualified observer according to the *International Classification of Diseases*" (1999, 1).

19. Although not provided by WHO, if researchers approach agencies within specific nations, they may be able to disaggregate even further on the basis of area, ethnicity, and type of victimization (e.g., gunshot, stabbing, strangulation, poisoning).

References

Abel, Ernest L. 1987. *Homicide: A bibliography*. Westport, Connecticut: Greenwood Press.

Adams, Ingrid, and Joyce Hickson. 1993. Wife-beating among coloureds in South Africa and its impact on the marital relationship. *International Journal of Sociology of the Family* 23:117-137.

Ade, Catherine M. 1995. Juvenile delinquency as reflected by offenders of the Borstal Institute of Cameroon. *International Journal of Comparative and Applied Criminal Justice* 19:257-268.

Adler, Freda, Gerhard O. Mueller, and William S. Laufer. 1994. *Criminal justice*. New York: McGraw-Hill.

Ali Badr, El-Din. 1986. Methodological problems in international criminal justice research. *International Journal of Comparative and Applied Criminal Justice* 10:163-176.

Alvazzi del Frate, Anna. 1998. *Victims of crime in the developing world*. Rome: United Nations Interregional Crime and Justice Research Institute.

American Sociological Association. 1999. Kenneth Prewitt leads Census Bureau toward 2000. *Footnotes* (April): 3.

Anderson, Tammy L., and Richard R. Bennett. 1996. Development, gender, and crime: The scope of the routine activities approach. *Justice Quarterly* 13:31-56.

- Archer, Dane, and Rosemary Gartner. 1984. *Violence and crime in cross-national perspective*. New Haven: Yale University Press.
- Aromaa, Kauko, and Andri Ahven. 1993. *Victims of crime in two Baltic countries: Finnish and Estonian data from the 1992/1993 International Crime Victimization Survey*. Helsinki: National Research Institute of Legal Policy.
- Arthur, John A. 1992. Social change and crime rates in Puerto Rico. *International Journal of Offender Therapy and Comparative Criminology* 36:103–119.
- . 1991. Development and crime in Africa: A test of modernization theory. *Journal of Criminal Justice* 19:499–513.
- Arthur, John A., and Otwin Marenin. 1995. Explaining crime in developing countries: The need for a case study approach. *Crime, Law, and Social Change* 23:191–214.
- Atkinson, Anthony B., and John Micklewright. 1992. *Economic transformation in Eastern Europe and the distribution of income*. Cambridge: Cambridge University Press.
- Avison, William R., and Pamela L. Loring. 1986. Population diversity and cross-national homicide: The effects of inequality and heterogeneity. *Criminology* 24:733–749.
- Bachman, Ronet. 1992. *Death and violence on the reservation: Homicide, family violence, and suicide in American Indian populations*. New York: Auburn House.
- Baden, Michael. 1998. Conversation with William Alex Pridemore, 14 April.
- Balakina, Galina F. 1994. Sovremennaya Tuva: Sotsiokulturnii aspekt razvitiia (The modern Tuva: Sociocultural aspects of development). *Sotsiologicheskaya Issledovaniya* 21:159–161.
- Barak, Gregg, ed. 2000. *Crime and crime control: A global view*. Westport, Connecticut: Greenwood Press.
- . 1991. *Crimes by the capitalist state: An introduction to state criminality*. Albany: State University of New York Press.
- Bayley, David H. 1976. *Forces of order: Police behavior in Japan and the United States*. Berkeley: University of California Press.
- Beirne, Piers. 1999. For a nonspeciesist criminology: Animal abuse as an object of study. *Criminology* 37:117–147.
- . 1997. *Issues in comparative criminology*. Brookfield, Vermont: Dartmouth Publishing Company.
- . 1983. Cultural relativism and comparative criminology. *Contemporary Crises* 7:371–391.

Beirne, Piers, and Joan Hill, comps. 1991. *Comparative criminology: An annotated bibliography*. Westport, Connecticut: Greenwood Press.

Bennett, Richard R. 1991a. Development and crime: A cross-national time-series analysis of competing models. *Sociological Quarterly* 32:343–363.

———. 1991b. Routine activities: A cross-national assessment of a criminological perspective. *Social Forces* 70:147–163.

Bennett, Richard R., and Peter P. Basiotis. 1991. Structural correlates of juvenile property crime: A cross-national time-series analysis. *Journal of Research in Crime and Delinquency* 28:262–287.

Bennett, Richard R., and James P. Lynch. 1990. Does a difference make a difference? Comparing cross-national crime indicators. *Criminology* 28:153–182.

Biderman, Albert D., and James P. Lynch. 1991. *Understanding crime incidence statistics: Why the UCR diverges from the NCS*. New York: Springer-Verlag.

Blau, Joan R., and Peter M. Blau. 1982. The cost of inequality: Metropolitan structure and violent crime. *American Sociological Review* 47:114–129.

Blau, Peter M. 1977. *Inequality and heterogeneity: A primitive theory of social structure*. New York: Free Press.

Block, Richard. 1993. Measuring victimisation: The effects of methodology, sampling, and fielding. In *Understanding crime: Experiences of crime and crime control*, edited by Anna Alvazzi del Frate, Ugljesa Zvekic, and Jan J.M. van Dijk. Rome: United Nations Interregional Crime and Justice Research Institute.

———, ed. 1984. *Victimization and fear of crime: World perspectives*. NCJ 93872. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Bouley, Eugene E., and Michael S. Vaughn. 1995. Violent crime and modernization in Colombia. *Crime, Law, and Social Change* 23:17–40.

Bouloukos, Adam C. 1999. Human rights conformity: An ideal type analysis of prisoners' rights in the United States and England. Ph.D. diss., State University of New York, University at Albany.

Bowker, Lee H. 1985. The effects of national development on the position of married women in the Third World: The case of wife-beating. *International Journal of Comparative and Applied Criminal Justice* 9:1–13.

Braithwaite, John, and Valerie Braithwaite. 1980. The effect of income inequality and social democracy on homicide. *British Journal of Criminology* 20:45–57.

Briggs, Carl M., and Phillips Cutright. 1994. Structural and cultural determinants of child homicide: A cross-national analysis. *Violence and Victims* 9:3–16.

- Bursik, Robert J., Jr. 1988. Social disorganization and theories of crime and delinquency: Problems and prospects. *Criminology* 26:519–552.
- Butler, Samuel. [1872] 1985. *Erewhon*. London: Penguin.
- Castells, Manuel. 1977. *The urban question: A Marxist approach*. Cambridge, Massachusetts: MIT Press.
- Chambliss, William J. 1993. State-organized crime. In *Making law: The state, the law, and structural contradictions*, edited by William J. Chambliss and Marjorie S. Zatz. Bloomington: Indiana University Press.
- Chimbos, Peter D. 1993. A study of patterns in criminal homicide in Greece. *International Journal of Comparative Sociology* 34:260–271.
- Chin, Ko-lin. 1994. Out-of-town brides: International marriages and wife abuse among Chinese immigrants. *Journal of Comparative Family Studies* 25:53–69.
- Chirot, Daniel, and Thomas D. Hall. 1982. World system theory. *Annual Review of Sociology* 8:81–106.
- Clinard, Marshall B., and Daniel J. Abbott. 1973. *Crime in developing countries: A comparative perspective*. New York: John Wiley & Sons.
- Cohen, Lawrence, and Marcus Felson. 1979. Social change and crime rate trends: A routine activities approach. *American Sociological Review* 44:588–608.
- Cornell, Claire P., and Richard J. Gelles. 1982. International perspectives on spouse abuse. *Focus on Women* 3:171–184.
- Corzine, Jay, Lin Huff-Corzine, and Hugh P. Whitt. 1998. Cultural and subcultural theories of homicide. In *Homicide: A sourcebook of social research*, edited by M. Dwayne Smith and Margaret A. Zahn. Thousand Oaks, California: Sage Publications.
- Curtis, Lynn A. 1975. *Violence, race, and culture*. Lexington, Massachusetts: D.C. Heath.
- Dahrendorf, Ralf. 1970. The intellectual and society: The social function of the fool in the twentieth century. In *On intellectuals: Theoretical studies, case studies*, edited by Philip Rieff. New York: Anchor Doubleday.
- David, Rene, and John E.C. Brierley. 1985. *Major legal systems of the world today: An introduction to the comparative study of law*. 3d ed. London: Stevens and Sons.
- De Leeuw, Edith D., and Johannes van der Zouwen. 1988. Data quality in telephone and face-to-face surveys: Comparative meta-analysis. In *Telephone survey methodology*, edited by Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, W.L. Nichols II, and Joseph Waksberg. New York: John Wiley & Sons.

DeKeseredy, Walter S., and Ronald Hinch. 1991. Wife abuse in Canada: The current state of sociological knowledge. *International Criminal Justice Review* 1:35–52.

DeVos, George. 1980. Delinquency and minority status: A psychocultural approach. In *Crime and deviance: A comparative perspective*, edited by Graeme R. Newman. Beverly Hills: Sage Publications.

DiCristina, Bruce. 1995. *Method in criminology: A philosophic primer*. Albany, New York: Harrow and Heston.

Downes, David. 1988. *Comparison of incarceration rates between The Netherlands and England and Wales: Contrast in tolerance*. Oxford: Clarendon Press.

Durkheim, Emile. [1893] 1964. *The division of labor in society*. Translated by George Simpson. New York: Free Press.

———. [1895] 1938. *The rules of sociological method*. Chicago: University of Chicago Press.

Easteal, Patricia W. 1994. Survivors of sexual assault: An Australian survey. *International Journal of the Sociology of Law* 22:329–354.

Ebbe, Obi N.I., ed. 1996. *Comparative and international criminal justice systems: Policing, judiciary, and corrections*. Boston: Butterworth-Heinemann.

Elias, Norbert. 1982. *The civilizing process, vol. 2: Power and incivility*. New York: Pantheon Books.

Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton. 1985. *Explaining delinquency and drug use*. Beverly Hills: Sage Publications.

Ermann, M. David, and Richard J. Lundman. 1992. *Corporate and governmental deviance: Problems of organizational behavior in contemporary society*. New York: Oxford University Press.

Evans, Peter B., and Michael Timberlake. 1980. Dependence, inequality, and the growth of the tertiary: A comparative analysis of less developed countries. *American Sociological Review* 45:531–554.

Evans, Sandra S., and Joseph E. Scott. 1984. The seriousness of crime cross-culturally: The impact of religiosity. *Criminology* 22:39–59.

Farrell, Graham, Kashfia Mansur, and Melissa Tullis. 1996. Cocaine and heroin in Europe, 1983–1993: A cross-national comparison of trafficking and prices. *British Journal of Criminology* 36:255–281.

Felson, Marcus, and Lawrence Cohen. 1980. Human ecology and crime: A routine activity approach. *Human Ecology* 8:389–406.

- Fiala, Robert, and Gary LaFree. 1988. Cross-national determinants of child homicide. *American Sociological Review* 53:432–445.
- Finckenauer, James O. 1995. *Russian youth: Law, deviance, and the pursuit of freedom*. New Brunswick, New Jersey: Transaction Publishers.
- Fish, Mary. 1984. Deterring sex sales to international tourists: A case study of Thailand, South Korea, and the Philippines. *International Journal of Comparative and Applied Criminal Justice* 8:175–186.
- Fowles, Richard, and Mary Merva. 1996. Wage inequality and criminal activity: An extreme bounds analysis for the United States, 1975–1990. *Criminology* 34:163–182.
- Fox, James A., and Marianne W. Zawitz. 1998. Homicide trends in the United States. NCJ 174433. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics. Retrieved 3 January 2000 from the World Wide Web: <http://www.ojp.usdoj.gov/bjs/homicide/homtrnd.htm>.
- Freud, Sigmund. 1962. *Civilization and its discontents*. Translated by James Strachey. New York: W.W. Norton.
- Friday, Paul C., Akira Yamagami, and John P.J. Dussich. 1999. The threshold of violence: A comparative study of USA and Japan. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.
- Fromm, Erich. 1976. *To have or to be?* Toronto: Bantam Books.
- Gartner, Rosemary. 1990. The victims of homicide: A temporal and cross-national comparison. *American Sociological Review* 55:92–106.
- Gartner, Rosemary, Kathryn Baker, and Fred C. Pampel. 1990. Gender stratification and the gender gap in homicide victimization. *Social Problems* 37:593–612.
- Gelles, Richard J. 1987. What to learn from cross-cultural and historical research on child abuse and neglect: An overview. In *Child abuse and neglect: Biosocial dimensions*, edited by Richard J. Gelles and Jane B. Lancaster. New York: Aldine de Gruyter.
- Gelles, Richard J., and Jane B. Lancaster, eds. 1987. *Child abuse and neglect: Biosocial dimensions*. New York: Aldine de Gruyter.
- Gilbert, Alan, and Josef Gugler. 1982. *Cities, poverty, and development: Urbanization in the Third World*. New York: Oxford University Press.
- Gillis, A.R. 1994. Literacy and the civilization of violence in 19th-century France. *Sociological Forum* 9:371–401.
- Gondolf, Edward W., and Dmitri Shestakov. 1997. Spousal homicide in Russia: Gender inequality in a multifactor model. *Violence Against Women* 3:533–546.

Gottlieb, Peter, and Gorm Gabrielsen. 1992. Alcohol-intoxicated homicides in Copenhagen, 1959–1983. *International Journal of Law and Psychiatry* 15:77–87.

Granucci, A.F. 1969. Nor cruel and unusual punishments inflicted: The original meaning. *California Law Review* 57 (October): 839–865.

Groves, W. Byron, Richard McCleary, and Graeme R. Newman. 1985. Religion, modernization, and world crime. *Law & Society Review* 25:287–312.

Gurr, Ted R., Peter N. Grabosky, and Richard C. Hula. 1977. *Politics of crime and conflict: A comparative history of four cities*. Beverly Hills: Sage Publications.

Hansmann, Henry B., and John M. Quigley. 1982. Population heterogeneity and the sociogenesis of homicide. *Social Forces* 61:206–224.

Hartjen, Clayton A., and S. Priyadarsini. 1984. *Delinquency in India: A comparative analysis*. New Brunswick, New Jersey: Rutgers University Press.

Hartnagel, Timothy F. 1982. Modernization, female social roles, and female crime: A cross-national investigation. *Sociological Quarterly* 23:477–490.

Hatalak, O., Anna Alvazzi del Frate, and Ugljesa Zvekic, eds. 1998. *The International Crime Victim Survey in countries in transition: National reports*. Rome: United Nations Interregional Crime and Justice Research Institute.

Heiland, Hans-Gunther, and Louise I. Shelley. 1992. Civilization, modernization, and the development of crime and control. In *Crime and control in comparative perspectives*, edited by Hans-Gunther Heiland, Louise I. Shelley, and Hisao Katoh. Hawthorne, New York: Walter de Gruyter.

Heiland, Hans-Gunther, Louise I. Shelley, and Hisao Katoh, eds. 1992. *Crime and control in comparative perspectives*. Hawthorne, New York: Walter de Gruyter.

Helal, Adel A., and Charisse T.M. Coston. 1991. Low crime rates in Bahrain: Islamic social control; Testing the theory of synnomie. *International Journal of Comparative and Applied Criminal Justice* 15:125–144.

Hobsbawm, Eric J. 1969. *Bandits*. New York: Delacorte Press.

———. 1959. *Primitive rebels: Studies in archaic forms of social movement in the 19th and 20th centuries*. New York: W.W. Norton.

Huang, W.S. Wilson. 1995. A cross-national analysis on the effect of moral individualism on murder rates. *International Journal of Offender Therapy and Comparative Criminology* 39:63–75.

Huang, W.S. Wilson, and Charles F. Wellford. 1989. Assessing indicators of crime among international crime dataset series. *Sociological Commentator* (Spring): 41–50.

- Inciardi, James A. 1984. *Criminal justice*. Orlando, Florida: Academic Press.
- International Criminal Police Organization. 1995. *International crime statistics*. Lyon, France: International Criminal Police Organization.
- International Monetary Fund. 1999. IMF dissemination standards bulletin board. Retrieved 3 January 2000 from the World Wide Web: <http://dsbb.imf.org>.
- Ivkovic, Sonja K., and Carl Klockars. 1996. A comparative cross cultural study of police corruption. Paper presented at the 41st Annual Meeting of the American Society of Criminology, 20–23 November, Chicago.
- Johnson, Elmer H. 1999. Domestic violence in particular countries. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.
- Johnson, Eric A. 1990. The Industrial Revolution and crime in Germany: Quantitative and qualitative observations. *Historical Social Research* 15:41–50.
- Joutsen, Matti. 1998. Introduction. In *Crime and criminal justice in Europe and North America, 1990–1994*, edited by Kristiina Kangaspunta, Matti Joutsen, and Natalia Ollus. Helsinki: European Institute for Crime Prevention and Control, affiliated with the United Nations.
- Junger-Tas, Josine. 1996. Youth and violence in Europe: A quantitative review. *Studies on Crime and Crime Prevention* 5:31–58.
- Junger-Tas, Josine, Gert-Jan Terlouw, and Malcolm W. Klein, eds. 1994. *Delinquent behavior among young people in the Western world: First results of the International Self-Report Delinquency study*. New York: Kugler Publications.
- Kaiser, Gunther. 1984. *Prison systems and correctional laws: Europe, the United States, and Japan: A comparative analysis*. Dobbs Ferry, New York: Transnational Publishers.
- Kangaspunta, Kristiina. 1999. Reduction of the prison population in Finland. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.
- Kangaspunta, Kristiina, Matti Joutsen, and Natalia Ollus, eds. 1998. *Crime and criminal justice in Europe and North America, 1990–1994*. Helsinki: European Institute for Crime Prevention and Control, affiliated with the United Nations.
- Kennedy, Bruce P., Ichiro Kawachi, and Deborah Prothrow-Stith. 1996. Income distribution and mortality: Cross-sectional ecological study of the Robin Hood index in the United States. *British Medical Journal* 312:1004–1007.
- Khan, Naira. 1995. Patterns of child sexual abuse in Zimbabwe: An overview. *Zimbabwe Journal of Educational Research* 7:181–208.

- Kick, Edward, and Gary LaFree. 1985. Development and the social context of murder and theft. *Comparative Social Research* 8:37–57.
- Killias, Martin. 1993. International correlations between gun ownership and rates of homicide and suicide. *Canadian Medical Association Journal* 148:1721–1776.
- . 1992. Gun ownership, suicide, and homicide: An international perspective. In *Understanding crime: Experiences of crime and crime control*, edited by Anna Alvazzi del Frate, Ugljesa Zvekic, and Jan J.M. van Dijk. Rome: United Nations Interregional Crime and Justice Research Institute.
- . 1990. New methodological perspectives for victimization surveys: The potentials of computer-assisted telephone surveys and some related innovations. *International Review of Victimology* 1:153–167.
- Klein, Malcolm W., ed. 1989. *Cross-national research in self-reported crime and delinquency*. Norwell, Massachusetts: Kluwer Academic Publishers.
- Kohn, Melvin L. 1989. Cross-national research as an analytic strategy. In *Cross-national research in sociology*, edited by Melvin L. Kohn. Beverly Hills: Sage Publications.
- Korbin, Jill E. 1987. Child maltreatment in cross-cultural perspective: Vulnerable children and circumstances. In *Child abuse and neglect: Biosocial dimensions*, edited by Richard J. Gelles and Jane B. Lancaster. New York: Aldine de Gruyter.
- Krahn, Harvey, Timothy F. Hartnagel, and John W. Gartrell. 1986. Income inequality and homicide rates: Cross-national data and criminological theories. *Criminology* 24:269–295.
- Krohn, Marvin, and Charles F. Wellford. 1977. A static and dynamic analysis of crime and the primary dimensions of nations. *International Journal of Criminology and Penology* 5:1–16.
- Krug, Etienne G., Kenneth E. Powell, and Linda L. Dahlberg. 1988. Firearm-related deaths in the United States and 95 other high- and upper-middle-income countries. *International Journal of Epidemiology* 27:214–221.
- LaFree, Gary. 1998. A summary and review of cross-national comparative studies of homicide. In *Homicide: A sourcebook of social research*, edited by M. Dwayne Smith and Margaret A. Zahn. Thousand Oaks, California: Sage Publications.
- LaFree, Gary, Kriss A. Drass, and Patrick O'Day. 1992. Race and crime in postwar America: Determinants of African-American and white rates, 1957–1988. *Criminology* 30:157–188.
- LaFree, Gary, and Edward Kick. 1986. Cross-national effects of developmental, distributional, and demographic variables on crime: A review and analysis. *International Annals of Criminology* 24:213–235.

- Landau, Simha F., and Danny Pfeffermann. 1988. A time series analysis of violent crime and its relation to prolonged states of warfare: The Israeli case. *Criminology* 26:489–504.
- Langan, Patrick A., and David Farrington. 1998. *Crime and justice in the United States and in England and Wales, 1981–96*. Report, NCJ 169284. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Leavitt, Gregory C. 1990. Relativism and cross cultural criminology: A critical analysis. *Journal of Research in Crime and Delinquency* 27 (February): 5–29.
- Lee, Yoon H. 1990. An integration and elaboration of theory to explain the delinquency of Korean boys. *International Journal of Comparative and Applied Criminal Justice* 14:25–40.
- Lester, David. 1994. Bellicose nations and rates of personal violence. *International Social Science Review* 69:13–16.
- . 1991. Crime as opportunity: A test of the hypothesis with European homicide rates. *British Journal of Criminology* 31:186–191.
- Lindstrom, Peter. 1997. Patterns of school crime: A replication and empirical extension. *British Journal of Criminology* 37:121–130.
- Loftin, Colin, and Robert N. Parker. 1985. An errors-in-variable model of the effect of poverty on urban homicide rates. *Criminology* 23:269–287.
- Los, Maria. 1990. *The second economy in Marxist states*. New York: St. Martin's Press.
- Luckenbill, D.F., and D.P. Doyle. 1989. Structural position and violence: Developing a cultural explanation. *Criminology* 27:419–436.
- Luikkonen, M. 1997. *Motor vehicle theft in Europe*. Helsinki: European Institute for Crime Prevention and Control, affiliated with the United Nations.
- Lunden, W.A. 1976. Violent crimes in Japan in war and peace. *International Journal of Criminology and Penology* 4:349–363.
- Lynch, James P. 1995a. Building data systems for cross-national comparisons of crime and criminal justice policy: A retrospective. *ICPSR Bulletin* 15:1–6.
- . 1995b. Crime in international perspective. In *Crime*, edited by James Q. Wilson and Joan Petersilia. San Francisco: Institute for Contemporary Studies.
- . 1993. Secondary analysis of international crime survey data. In *Understanding crime: Experiences of crime and crime control*, edited by Anna Alvazzi del Frate, Ugljesa Zvekic, and Jan J.M. van Dijk. Rome: United Nations Interregional Crime and Justice Research Institute.

Maguire, Edward R., Gregory J. Howard, and Graeme Newman. 1998. Measuring the performance of national criminal justice systems. *International Journal of Comparative and Applied Criminal Justice* 22:31–59.

Mahabir, C. 1988. Crime in the Caribbean: Robbers, hustlers, and warriors. *International Journal of the Sociology of Law* 16:315–338.

Marsh, Robert Mortimer. 1967. *Comparative sociology: A codification of cross-societal analysis*. New York: Harcourt Brace Jovanovich.

Martin, John M., and Anne T. Romano. 1992. *Multinational crime: Terrorism, espionage, drug and arms trafficking*. Newbury Park, California: Sage Publications.

Matsuoka, J., and T. Kelly. 1988. The environmental, economic, and social impacts of resort development and tourism on Native Hawaiians. *Journal of Sociology and Social Welfare* 15:29–44.

McClintock, F.H., and Per-Olaf Wikstrom. 1992. The comparative study of urban violence: Criminal violence in Edinburgh and Stockholm. *British Journal of Criminology* 32:505–520.

McDonald, Lynn. 1976. *The sociology of law and order*. Boulder, Colorado: Westview.

McDowall, David. 1986. Poverty and homicide in Detroit, 1926–1978. *Violence and Victims* 1:23–34.

Merton, Robert K. 1938. Social structure and anomie. *American Sociological Review* 3:672–682.

Messerschmidt, James D. 1986. *Capitalism, patriarchy, and crime: Toward a Socialist feminist criminology*. Totowa, New Jersey: Rowman & Littlefield.

Messner, Steven F. 1989. Economic discrimination and societal homicide rates: Further evidence on the cost of inequality. *American Sociological Review* 54:597–611.

———. 1986. Modernization, structural characteristics, and societal rates of crime: An application of Blau's macrosociological theory. *Sociological Quarterly* 27:27–41.

———. 1983a. Regional and racial effects on the urban homicide rate: The subculture of violence revisited. *American Journal of Sociology* 88:997–1007.

———. 1983b. Regional differences in the economic correlates of the urban homicide rate: Some evidence on the importance of cultural context. *Criminology* 21:477–488.

———. 1982. Societal development, social equality, and homicide: A cross-national test of a Durkheimian model. *Social Forces* 61:225–240.

———. 1980. Income inequality and murder rates: Some cross-national findings. *Comparative Social Research* 3:185–198.

- Messner, Steven F., and Richard Rosenfeld. 1997. Political restraint of the market and levels of criminal homicide: A cross-national application of institutional-anomie theory. *Social Forces* 75:1393–1416.
- Mills, C. Wright. 1959. *The sociological imagination*. London: Oxford University Press.
- Miyazawa, K., and S.E. Cook. 1990. Juvenile delinquency and juvenile justice in Japan. *International Annals of Criminology* 28:55–65.
- Moore, Richter H., Jr., and Charles B. Fields. 1996. Comparative criminal justice: Why study? In *Comparative criminal justice: Traditional and nontraditional systems of law and control*, edited by Charles B. Fields and Richter H. Moore, Jr. Prospect Heights, Illinois: Waveland Press.
- Morales, E. 1986. Coca and cocaine economy and social change in the Andes of Peru. *Economic Development and Cultural Change* 35:143–161.
- Mueller, Gerhard O.W., and Freda Adler. 1996. Globalization and criminal justice: A prologue. In *Comparative criminal justice: Traditional and nontraditional systems of law and control*, edited by Charles B. Fields and Richter H. Moore, Jr. Prospect Heights, Illinois: Waveland Press.
- Murdock, G.P., and D.R. White. 1969. Standard cross-cultural sample. *Ethnology* 8:326–369.
- Murty, K.S., A.M. Al-Lanqawi, and J.B. Roebuck. 1990. Self-reported juvenile delinquency in Kuwait. *International Journal of Comparative and Applied Criminal Justice* 14:49–63.
- Mutzell, S. 1995. Human violence in Stockholm County, Sweden. *International Journal of Adolescence and Youth* 6:75–88.
- Natarajan, M. 1995. Victimization of women: A theoretical perspective on dowry deaths in India. *International Review of Victimology* 3:297–308.
- Neapolitan, Jerome L. 1998. Cross-national variation in homicide: Is race a factor? *Criminology* 36:139–156.
- . 1997. *Cross-national crime: A research review and sourcebook*. Westport, Connecticut: Greenwood Press.
- . 1996. Cross-national crime data: Some unaddressed problems. *Journal of Criminal Justice* 19:95–112.
- . 1995. Differing theoretical perspectives and cross-national variation in thefts in less-developed nations. *International Criminal Justice Review* 5:17–31.
- . 1994. Cross-national variation in homicides: The case of Latin America. *International Criminal Justice Review* 4:4–22.

- Neuman, W.L., and R.J. Berger. 1988. Competing perspectives on cross-national crime: An evaluation of theory and evidence. *Sociological Quarterly* 29:281–313.
- Newman, Graeme. 1999. *Global report on crime and justice*. New York: Oxford University Press.
- . 1976. *Comparative deviance: Perception and law in six cultures*. New York: Elsevier North-Holland.
- Newman, Graeme, and Gregory J. Howard. 1999. Criminal justice resources. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.
- Niggli, M.A., and F. Pfister. 1997. Paradise lost? Paradise ever? On crime development in Switzerland. *Studies on Crime and Crime Prevention* 6:73–99.
- Nisbett, Richard E., and Dov Cohen. 1996. *Culture of honor: The psychology of violence in the South*. Boulder, Colorado: Westview Press.
- O'Malley, Patrick. 1980. The class production of crime: Banditries and class strategies in England and Australia. *Research in Law and Society* 3:181–199.
- Ogunlesi, A.O. 1990. Nigerian juvenile offenders: A case-controlled study. *Forensic Science International* 44:187–192.
- Ortega, S.T., J. Corzine, C. Burnett, and T. Poyer. 1992. Modernization, age structure, and regional context: A cross-national study of crime. *Sociological Spectrum* 12:257–277.
- Park, Robert E., and Ernest Burgess. 1924. *Introduction to the science of sociology*. Chicago: University of Chicago Press.
- Parker, Robert J. 1998. Alcohol, homicide, and cultural context: A cross-national analysis of gender-specific homicide victimization. *Homicide Studies* 2:6–30.
- Parsons, Talcott. 1977. *The evolution of societies*. Englewood Cliffs, New Jersey: Prentice Hall.
- Passos, Nikos. 1995. The mirror of global evils: A review essay on the BCCI affair. *Justice Quarterly* 12:801–829.
- Patrignani, A., and R. Ville. 1995. *Violence in the family: An international bibliography with literature review*. Rome: United Nations Interregional Crime and Justice Research Institute.
- Pfeiffer, C. 1998. *Trends in juvenile violence in European countries*. Research in Progress Preview, FS 000202. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

- President's Commission on Law Enforcement and Administration of Justice. 1967. *The challenge of crime in a free society*. Washington, D.C.: U.S. Government Printing Office.
- Pridemore, William A. 1999. Spatial, temporal, and demographic variation of homicide rates in Russian regions. Paper presented at the 44th Annual Meeting of the American Society of Criminology, 17–21 November, Toronto.
- Pridemore, William A., and Vitali I. Kvashis. 1999. Social problems and juvenile delinquency in post-Soviet Russia. Paper presented at the 44th Annual Meeting of the American Society of Criminology, 17–21 November, Toronto.
- Rattner, R.S. 1991. Multi-tiered terrorism in Peru. In *Crimes by the capitalist state: An introduction to state criminality*, edited by Gregg Barak. Albany: State University of New York Press.
- Rechea, C., Rosemary Barberet, and J. Montanes. 1995. *Juvenile delinquency in Spain*. Madrid: Spanish Ministry of Justice and Interior.
- Redfield, H.V. 1880. *Homicide, North and South: Being a crime against the person in several parts of the United States*. Philadelphia: J.B. Lippincott.
- Reichel, Philip L. 1994. *Comparative criminal justice systems: A topical approach*. Englewood Cliffs, New Jersey: Prentice Hall.
- Reiman, Jeffrey H. 1998. *The rich get richer and the poor get prison: Ideology, class, and criminal justice*. 3d ed. Boston: Allyn and Bacon.
- Rokaw, W.M., J.A. Mercy, and J.C. Smith. 1990. Comparing death certificate data with FBI crime reporting statistics on U.S. homicides. *Public Health Reports* 105:447–455.
- Rosenfeld, Richard, and Steven F. Messner. 1991. The social sources of homicide in different types of societies. *Sociological Forum* 6:51–70.
- Rummel, R.J. 1994. Power, genocide, and mass murder. *Journal of Peace Research* 31:1–10.
- Rushton, Jean P. 1995. Race and crime: International data for 1989–1990. *Psychological Reports* 76:307–312.
- . 1990. Race and crime: A reply to Roberts and Gabor. *Canadian Journal of Criminology* 32:315–334.
- Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 277:918–924.
- Schichor, David. 1990. Crime patterns and socioeconomic development: A cross-national analysis. *Criminal Justice Review* 15:64–78.

- . 1985. Effects of development on official crime rates, 1967–1978: Homicide rates and larceny patterns differ greatly. *Sociology and Social Research* 70:96–97.
- Schwartz, B. 1971. *The Bill of Rights: A documentary history*. Broomall, Pennsylvania: Chelsea House Publishers.
- Scott, J.E., and F. Al-Thokeb. 1977. The public's perception of crime. In *Crime and deviance: A comparative perspective*, edited by Graeme Newman. Beverly Hills: Sage Publications.
- Shelley, Louise I. 1981. *Crime and modernization: The impact of industrialization and urbanization on crime*. Carbondale: Southern Illinois University Press.
- Shoham, Shlomo G. 1995. *Walhalla, Golgotha, Auschwitz: Über die interdependenz von Deutschen und Juden*. Vienna: Edition S.
- Skinner, W.F. 1986. Delinquency, crime, and development: A case study of Iceland. *Journal of Research in Crime and Delinquency* 23:268–294.
- Skogan, Wesley. 1986. Methodological issues in the study of victimization. In *From crime policy to victim policy: Reorienting the justice system*, edited by Ezzat A. Fattah. Basingstoke, England: Macmillan.
- . 1981. *Issues in the measurement of victimization*. NCJ 74682. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Smith, N. 1984. *Uneven development: Nature, capital, and the production of space*. New York: Basil Blackwell.
- Stack, Steven. 1984. Income inequality and property crime: A cross-national analysis of relative deprivation theory. *Criminology* 22:229–257.
- Stojsavljevic, J. 1995. Women, conflict, and culture in former Yugoslavia. *Gender and Development* 3:36–41.
- Strasser, Herman, and Susan C. Randall. 1981. *An introduction to theories of social change*. London: Routledge & Kegan Paul.
- Sutherland, Edwin H. 1947. *Criminology*. 4th ed. Philadelphia: J.B. Lippincott.
- Swidler, A. 1986. Culture in action. *American Sociological Review* 51:273–286.
- United Nations. 1995. *Ninth United Nations Congress on the Prevention of Crime and the Treatment of Offenders*. Cairo: United Nations.
- United Nations Criminal Justice Information Network. 1999. *Fifth United Nations Survey of crime trends and operations of criminal justice systems*. Retrieved 5 January 2000 from the World Wide Web: <http://www.uncjin.org/5wcs/5intro.txt>.

United Nations Development Programme. 1998. *Human development report 1998*. New York: Oxford University Press.

United Nations International Drug Control Programme. 1997. *Statistics on illicit trafficking in and supply of narcotic drugs and psychotropic substances in 1995*. New York: United Nations International Drug Control Programme.

Unnithan, N.P. 1997. Child homicide in developed countries. *International Review of Victimology* 4:313–326.

U.S. Bureau of the Census. 1999. *International Data Base (IDB)*. Retrieved 5 January 2000 from the World Wide Web: <http://www.census.gov/ipc/www/idbnew.html>.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. 1999. *National Center for Health Statistics*. Retrieved 3 January 2000 from the World Wide Web: <http://www.cdc.gov/nchswww>.

Van Dijk, Jan J.M. 1999. The experience of crime and justice. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.

Van Dijk, Jan J.M., and Pat Mayhew. 1993. Criminal victimisation in the industrialized world: Key findings of the 1989 and 1992 International Crime Surveys. In *Understanding crime: Experiences of crime and crime control*, edited by Anna Alvazzi del Frate, Ugljesa Zvekic, and Jan J.M. van Dijk. Rome: United Nations Interregional Crime and Justice Research Institute.

Van Dijk, Jan J.M., Pat Mayhew, and Martin Killias. 1990. *Experiences of crime across the world: Key findings of the 1989 International Crime Survey*. Deventer, The Netherlands: Kluwer Academic Publishers.

Van Dijk, Jan J.M., and J. van Kesteren. 1996. Criminal victimisation in European cities: Some results of the International Crime Victim Survey. *European Journal on Criminal Policy and Research* 4:9–21.

Vigderhous, G. 1978. Methodological problems confronting cross-cultural criminological research using official data. *Human Relations* 31:229–247.

Walberg, P., M. McKee, Vladimir Shkolnikov, L. Chenet, and D.A. Leon. 1998. Economic change, crime, and mortality crisis in Russia: Regional analysis. *British Medical Journal* 317:312–318.

Waller, Irvin, and Brandon Welsh. 1999. International trends in crime prevention: Cost-effective ways to reduce victimization. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.

Walton, J. 1982. The international economy and peripheral urbanization. In *Urban policy under capitalism*, edited by Norman I. Fainstein and S.S. Fainstein. Beverly Hills: Sage Publications.

Widom, Cathy S., and A.J. Stewart. 1986. Female criminality and the status of women. *International Annals of Criminology* 24:137–162.

Wigmore, John H. 1936. *A panorama of the world's legal systems*. Washington, D.C.: Washington Law Book Company.

Williams, Kirk R., and R.L. Flewelling. 1988. The social production of homicide: A comparative study of the disaggregated rates in American cities. *American Sociological Review* 53:421–431.

Williams, Phil. 1999. Emerging issues: Transnational crime and control. In *Global report on crime and justice*, edited by Graeme Newman. New York: Oxford University Press.

Wolfgang, Marvin E., and Franco Ferracuti. 1967. *The subculture of violence*. London: Tavistock.

World Bank. 1999. *The World Bank group development data: Data by topic*. Retrieved 5 January 2000 from the World Wide Web: <http://www.worldbank.org/data/databytopic/databytopic.html>.

World Health Organization. 1999. *WHO mortality statistics*. Retrieved 5 January 2000 from the World Wide Web: <http://www.who.int/whosis/mort/index.html>.

———. 1996. *World health statistics annual*. Geneva: World Health Organization.

———. 1995. *World health statistics annual*. Geneva: World Health Organization.

Wu, B. 1996. The impact of an antiamphetamine law on juvenile delinquency in Taiwan. *International Journal of Offender Therapy and Comparative Criminology* 40:234–242.

———. 1995. The declining gender difference in crime: The Taiwanese case. *International Journal of Offender Therapy and Comparative Criminology* 39:359–366.

Zaluar, A., and A.I. Ribeiro. 1995. The drug trade, crime, and policies of repression in Brazil. *Dialectical Anthropology* 20:95–108.

Zehr, Howard. 1976. *Crime and development in modern society: Patterns of criminality in 19th century Germany and France*. Totowa, New Jersey: Rowman & Littlefield.

Zvekic, Ugljesa, and Anna Alvazzi del Frate. 1995. *Criminal victimisation in the developing world*. Rome: United Nations Interregional Crime and Justice Research Institute.

Zvekic, Ugljesa, ed. 1990. *Essays on crime and development*. Rome: United Nations Interregional Crime and Justice Research Institute.

Spatial Analyses of Crime

by *Luc Anselin, Jacqueline Cohen, David Cook,
Wilpen Gorr, and George Tita*

The new century brings with it growing interest in crime *places*. This interest spans theory from the perspective of understanding the etiology of crime, and practice from the perspective of developing effective criminal justice interventions to reduce crime. We do not attempt a comprehensive treatment of the substantial body of theoretical and empirical research on place and crime but focus instead on methodological issues in spatial statistical analyses of crime data. Special attention is given to some practical and accessible methods of exploratory data analysis that arguably should be the starting place of any empirical analyses of the relationship of place to crime. Many of the capabilities to support computerized mapping and spatial statistical analyses emerged only recently during the 1990s. The promise of using spatial data and analyses for crime control still remains to be demonstrated and depends on the nature of the relationship between crime and place. If spatial features serve as actuating factors for

Luc Anselin is a Professor in the Department of Agricultural and Consumer Economics of the University of Illinois at Urbana-Champaign and is a member of the National Consortium on Violence Research. Jacqueline Cohen is a Principal Research Scientist at the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University and is a member of the National Consortium on Violence Research. David Cook works in crime analysis for the Greensboro, North Carolina, Police Department. Wilpen Gorr is Professor of Public Policy and Management at the H. John Heinz III School of Public Policy and Management at Carnegie Mellon University. George Tita is Associate Policy Analyst for the Criminal Justice Program at The RAND Corporation and is a member of the National Consortium on Violence Research.

A
B
S
T
R
A
C
T

213

**A
B
S
T
R
A
C
T**

crime, either because of the people who or facilities that are located there, then interventions designed to alter those persons and activities might well affect crime. Alternatively, if the spatial distribution of crime is essentially random, then targeting specific places is not likely to be an effective crime control strategy. Sorting out the place/crime relationship requires analytical methods that are best suited to isolating the impacts of place on crime.

As we near the end of the 20th century, interest in crime *places* continues to grow. The identification of crime hot spots (Sherman, Gartin, and Buerger 1989) was perhaps a watershed in refocusing attention on spatial/locational features of crime. This interest spans theory from the perspective of understanding the etiology of crime, and practice from the perspective of developing effective criminal justice interventions to reduce crime. Theoretical concerns focus on how place might be a factor in crime, either by influencing or shaping the types and levels of criminal behavior by the people who frequent an area, or by attracting to an area people who already share similar criminal inclinations. These theoretical concerns, which are ably addressed in a growing published literature on the criminology of place, are only briefly discussed in this paper.¹ We focus instead on the analytical methods best suited to isolating the impact of place on crime.

Technological advances, primarily in computer capabilities, are fundamental to recent analytical advances in the methods available for analyzing place-based crime data. The advent of computer mapping applications and accompanying geographic information systems (GIS) are crucial to being able to measure and represent the spatial relationships in data. Perhaps the most powerful analytical tools emerging from GIS technologies are (1) flexible spatial aggregation capabilities to facilitate the measurement of place-based crime and (2) simple contiguity matrices for representing neighbor relationships between different areal units. In addition to these analytical advances, computerized police records management systems and computer aided dispatch (CAD) systems of citizen calls to police make it possible to systematically quantify varying levels of criminal activity at different places within a city.

The paper that follows begins with a brief overview of some conceptual links between place and crime. We do not attempt a comprehensive treatment of the substantial body of theoretical and empirical research on this topic. Our intent is merely to provide an illustrative context for the main focus of the paper—spatial statistical analyses of crime data with special emphasis on pragmatic concerns about how these analyses are best implemented. The text guides readers through a variety of methodological concerns relating to the analysis of spatial data and space/time data. Perhaps the most valuable service is to direct analysts to relevant parts of a growing research literature, with many sources published only recently. Thorny issues are raised, not to warn analysts off altogether, but rather to encourage the exercise of due caution in the conduct and interpretation of empirical analyses. Special attention is given to some practical and accessible methods of exploratory data analysis that arguably should be the starting place of any empirical analyses of the relationship of place to crime.

Routine activities that bring together potential offenders and criminal opportunities are especially effective in explaining the role of place in encouraging or inhibiting crime. The resulting crime locales often take the form of facilities—places that people frequent for a specific purpose—that are attractive to offenders or conducive to offending.

Crime and Place

In this section, we briefly review some theoretical and empirical developments in research on crime and place. These trace back to the work of the early social ecologists in France during the middle of the 19th century, through the sociological tradition emerging from the Chicago School in the early 20th century, and finally to the recent revival of this tradition in contemporary ecological studies of crime. The social ecology perspective evolved into more specifically focused, place-based theories of crime, particularly the routine activities theory. Routine activities that bring together potential offenders and criminal opportunities are especially effective in explaining the role of place in encouraging or inhibiting crime. The resulting crime locales often take the form of facilities—places that people frequent for a specific purpose—that are attractive to offenders or conducive to offending. Facilities might provide an abundance of criminal opportunities (e.g., either a target-rich environment for thefts or abandoned or otherwise unguarded properties that could be used for illicit activities like drug dealing). Or they might be the sites of licit behaviors that are associated with increased risk of crime (e.g., heavy alcohol consumption in crowds

where disputes can easily turn violent). The relationship between specific types of facilities and observed crime hot spots is an important question, and these chronic crime places are particularly well suited for further empirical investigations of the distinctive criminogenic features associated with place.

Social ecology theories of crime

Early social ecologists

Invariably, research articles that focus on the concentration of crime in distinct types of communities cite the work of the early French social ecologists Guerry ([1833a] 1984, [1833b] 1974) and Quetelet (1833, 1842).² As in Durkheim's classic studies of suicide ([1897] 1966) and crime ([1901] 1950) a half-century later, Guerry and Quetelet were interested in explaining differences in community crime levels in terms of the varying social conditions of the resident populations. It is humbling to see the level of analytical sophistication displayed in

their early maps of population-based rates of crime, suicide, alcoholism, population age structure, family structure, educational levels, and population diversity in 19th-century French “Departments” (i.e., geopolitical areas analogous to contemporary States or provinces). These historical works are among the earliest examples of a type of empirical social research that falls within the tradition of ecological studies of crime—that is, studies in which the units of analysis are spatially defined population aggregates.

The next flourishing of ecological research on crime was in the early 20th century. More than any other academic body of work, the Chicago School of the early 1920s is responsible for the emergence of ecological studies in sociological research (for example, Park, Burgess, and McKenzie 1925). The Chicago School represents a sociological paradigm that encourages a synthesis of qualitative and quantitative methods. While many view it as atheoretical and primarily empirical, it is difficult to deny its importance in theoretical developments in community studies and criminology.³ As Abbott (1997, 1152) writes:

[T]he Chicago School thought—and thinks—that one cannot understand social life without understanding the arrangements of particular social actors in particular social times and places. . . . [N]o social fact makes any sense abstracted from its context in social (and often geographic) space and social time. Social facts are *located facts*. (emphasis in original)

The original data of the Chicago School were records obtained from the Cook County (Illinois) Juvenile Court, Boys’ Court and Jail. They included basic demographic measures like age and sex of each offender, along with the offender’s home address. The following passage from Bursik and Grasmick (1993, 31) describes the procedure used by Shaw and colleagues to map these data:

The residential address of each individual . . . was plotted (by hand!) on a base map of the city of Chicago (see Shaw et al. 1929:24) [sic] for a full description of the process) and then copied into outline maps of Chicago by means of a reflector and glass-top table. . . . The rates of delinquency (defined in terms of the number of boys referred to juvenile court) were then computed on the basis of census tracts, the official local community areas of Chicago, and one-square-mile areas of the city, which was their most common operational definition of the neighborhood.

Though not causally related, recent developments of widely accessible computerized mapping and spatial analysis techniques have accompanied the resurgence in popularity of ecological explanations of crime.

Relying only on “visual inspection . . . and rudimentary statistical tests” (Bursik and Grasmick 1993) of the resulting spatial distribution of offenders, Shaw and McKay (1942)⁴ emerged with their seminal findings regarding the stability of delinquency over time within certain neighborhoods and the negative relationship between crime and distance from the central business districts. The social disorganization theory of crime was born from these observations.

Other important work on crime and place emerging from the original Chicago School includes Thrasher’s (1927) census of urban street gangs. Mapping the locations where gangs formed, Thrasher found “gangland” in the “interstitial” areas of Chicago, and not in areas that could easily be labeled as “commercial” or “residential.” Gangs form where “better residential districts recede before the encroachment of business and industry” (p. 23). Understandably, Thrasher did not undertake what was then a formidable task of cataloging all of the features that distinguish “gangland” from nongang areas. With the advent of computers, and perhaps more importantly, the accessibility of computerized census data, this task is much more easily managed today.

The “new” Chicago School

A featured plenary session at the 1996 annual meeting of the American Society of Criminology held in Chicago addressed the question, “Whither the Chicago School?” As an esteemed panel of former Chicago School students and mentors discussed the past, present, and future of Chicago-style ecological studies, it became clear that we are currently in the midst of a Chicago School revival.⁵ Over the past two decades, a number of excellent studies have resurrected and advanced the methodological and theoretical traditions of the original Chicago School.⁶ Though not causally related, recent developments of widely accessible computerized mapping and spatial analysis techniques have accompanied the resurgence in popularity of ecological explanations of crime. The new GIS capabilities that permit flexible measurements at various levels of spatial aggregation have facilitated many recent analyses of ecological features of crime.

For instance, relying on their ability to map the location of homicides, aggregate these point locations to census tracts, and then examine the distribution of gang homicides controlling for “social disorganization,” Curry and Spergel (1988) find crime to be correlated with poverty and a lack of social control, but violence (e.g., homicide) to be correlated with their measure of social disorganization. Tita, Engberg, and Cohen (1999) provide another contemporary ecological study of gangs. They find that the areas where gangs form are low on a variety of measures of informal social control and share features associated with the “underclass.” Furthermore, once racial composition is accounted for,

their measure of social disorganization is not predictive of gang location. Gangs form in high-crime neighborhoods, but the arrival of gangs in an area does not alter local crime levels. The notable exception is a significant increase in shots fired after gangs form in an area (Tita 1999).

An excellent example of the blending of quantitative spatial measures with qualitative observational studies is Bernard Cohen's (1980) ecological study of street-level prostitution in New York City. Cohen finds that streetwalking spans all levels of income across neighborhoods and census tracts of Manhattan. However, he noticed important similarities in the block faces and street corners frequented by prostitutes and johns. Using hand-drawn maps, Cohen identified "hot spots" of prostitution activity. Relying on participant observation, he recorded and quantified the amount of deviance in the study areas, as well as the age, race, and gender of nearby residents. He extrapolated family structure from census tract data.

Areas with a high incidence of prostitution were notable in their absence of young children and young women. Not surprisingly, households in these areas were much more likely to be made up of single adults and unrelated roommates. Cohen also noted several important crime-enhancing features of the built environment, such as wide streets (to provide inconspicuous traffic flow of johns through the area), the types of business establishments in the area (to attract the "right" clientele), and the spatial proximity of unlit alleys, parks, or lots (to provide locations for sex acts). Although not widely recognized as such, Cohen's (1980) book, *Deviant Street Networks*, may be one of the first empirical studies to document the spatial and temporal intersection of "motivated" offenders and the crime-facilitating properties of place proposed by the routine activities theory (Cohen and Felson 1979).

Bernard Cohen's work underscores the importance of specifying the correct areal unit of analysis in ecological studies. When examining the presence of streetwalkers as a function of various socioeconomic measures aggregated to the level of neighborhoods or census tracts, there were few differences between areas with prostitution and those without. It is only when Cohen examined sub-census tract variation that important differences emerged. Modern GIS capabilities, combined with point data on the locations of individual crimes, make it feasible to routinely obtain measures of crime variables at these nontraditional and smaller levels of aggregation.

Place-based theories of crime

Ecological theories look for explanations of individual actions in general features of the social structure in which an individual is embedded. Place-based

theories fall squarely within the theoretical tradition of social ecology, but are more specific about the mechanisms by which structural context is translated into individual action. The dominant theoretical perspectives derive from the routine activities theory (Cohen and Felson 1979) and rational choice theory (Cornish and Clarke 1986). In both cases, the distribution of crime is determined by the intersection in time and space of suitable targets and motivated offenders. This spatial and temporal intersection is determined by the organization of certain types of activities at specific places, ranging from highly structured environments like work and school to less structured environments in the home and leisure places.

Routine activities

The routine activities theory was first introduced in Cohen and Felson (1979), later refined in Felson (1986, 1994), and extended to crime pattern theory in Brantingham and Brantingham (1993). Place is central to this perspective, serving as the locus where *motivated offenders* come together with *desirable targets* in the absence of *crime suppressors* (who include guardians, intimate handlers [Felson 1986], and place managers [Eck 1994]). This convergence of crime opportunities in space and time is facilitated by various situational features, of both the physical and social variety, that provide a context or setting that is more or less conducive to crime (Clarke 1992).

Place can facilitate (or inhibit) crime in two ways. First, the physical or built features of a place can decrease the social control capacities of various crime suppressors. Such concerns motivate interest in the design of “defensible space” (Jeffrey 1971; Newman 1972). For example, Newman’s study of public housing suggests that highrise housing increases population density, but because residents live vertically, they are physically removed from monitoring activities in public spaces, especially those at street level.⁷ These conditions leave this type of housing with relatively few place managers who will monitor and control public behavior and seriously limit the levels of informal social control exercised over all forms of disruptive behavior from minor incivilities to more serious illicit activities. Roncek and Francik (1981) find elevated crime levels in and near public housing even after including controls for the composition of the resident population on a variety of attributes. This provides support for a criminogenic role of the facility itself that is independent of the types of people who are found there.

Second, aside from physical features, crime at places is apparently influenced by the routine activities that occur there. Crime is not distributed evenly or randomly over space. Instead, higher levels of crime plague places with some types of facilities and not others. In some cases, crimes seem to be elevated by

a target-rich environment—for example, thefts of 24-hour convenience stores, auto thefts from large parking lots, or robberies of shoppers in heavily frequented commercial areas (e.g., Engstad 1975; Duffala 1976; Brantingham and Brantingham 1982). In others, certain activities such as alcohol consumption seem to contribute to increased levels of violence (Roncek and Bell 1981; Roncek and Pravatiner 1989; Roncek and Maier 1991; Homel and Clark 1995; Block and Block 1995). Still other places seem to be prone to higher levels of crime because of the types of people they attract and repel. Places with abandoned buildings or rundown housing with absentee owners are attractive to illicit drug dealers who are looking for places where they can establish stable marketing locations without fear of owner or neighbor complaints (Eck 1994).

Crime hot spots

The concentration of crime in identifiable places was noted in Brantingham and Brantingham (1982). These crime hot spots are prime exemplars of the potential value of place in the analysis of crime. Sherman, Gartin, and Buerger (1989) published one of the first studies to quantify what many qualitative studies had suggested—namely, that crime in a city is highly concentrated in relatively few small areas. The study found that 3.3 percent of street addresses and intersections in Minneapolis generated 50.4 percent of all dispatched police calls for service. Similar patterns emerged in other cities (Pierce, Spaar, and Briggs 1988; Sherman 1992; and Weisburd and Green 1994). While often motivated by pragmatic concerns about what interventions are likely to be effective in reducing crime, results like these also serve to sharply focus crime theory on developing satisfactory accounts of these apparently strong relationships between crime and place.

Crime studies that examine the spatial distribution of crime clearly demonstrate that certain land uses and population characteristics are associated with crime hot spots. Roncek and Maier (1991) found a positive relationship between levels of crime and the number of taverns and lounges located in city blocks in Cleveland. The influence of taverns on crime was compounded when the taverns were located in areas with more anonymity and lower guardianship. Five of the top ten hot spots identified in Sherman, Gartin, and Buerger (1989) included bars. Cohen, Gorr, and Olligschlaeger (1993) found that drug hot spots tended to be in areas with nuisance bars, rundown commercial establishments, or areas with poverty and low family cohesion as measured by female-headed households.

Skogan and Maxfield (1981) reported that environmental conditions such as abandoned buildings, public incivilities such as fights and other minor assaults, disorderly youths, broken windows or other forms of vandalism, public drug use or drinking, prostitution, loitering, noise, litter, and obscene behavior

increase community fear of crime. "Broken windows" and other public signs of disorder may also contribute to actual increases in more serious crime as visible signs of urban disorder signal that a community has lost its ability to exercise social control, further encouraging and perpetuating crime (Wilson and Kelling 1982; Greenberg, Rohe, and Williams 1985). Likewise, vigorous law enforcement strategies directed against various forms of public disorder and nuisance violations may actually inhibit more serious crimes by establishing visible signs of a vigilant and self-protective community (Boydstrun 1975; Wilson and Boland 1978; Pate et al. 1985; Sherman 1986; Sampson and Cohen 1988; Kelling and Coles 1996). This suggests that crime hot spots may arise first as concentrations of "soft" crimes that later harden to more serious crimes.

Whether or not hot spots contribute to crime in a causal way depends on whether or not the elevated levels of crime observed at hot spots are systematic (regular and predictable) and not just random occurrences. If hot spots are random and can occur anywhere, then crime in these locations does not depend on distinctive features found in the observed hot spots; and crime reduction efforts that target these features are likely to fail. Thus, careful identification of hot spots and methodologically sound analyses to establish whether they have meaningful links to crime are crucial.

Spatial Data Analysis Tools

The spatial concentration of crime in hot spots leads naturally to their representation on crime maps. Maps of crime incidents permit rapid identification of the geographic location of crime hot spots, but by themselves they contribute little to understanding why crime is concentrated in certain locations. A crucial aspect of pattern recognition techniques such as hot spot analysis is the determination of the extent to which patterns on the map reflect "true" clusters or outliers or whether they are spurious. As is well known, simple visual interpretation of the map is inadequate in this respect because the human mind is conditioned to find meaning and identify patterns and clusters, even when the data represented may be purely random. The use of sound cartographic principles alone does not ensure that a proper interpretation is obtained (Rheingans and Landreth 1995; Gahegan and O'Brien 1997; MacEachren and Kraak 1997). What is needed is a careful structuring of the visualization strategy while supplementing the visual aspects with quantitative information (Cleveland 1993).

Hot spot representation

A crime hot spot is a location, or small area within an identifiable boundary, with a concentration of criminal incidents. These chronic crime places where

crime is concentrated at high rates over extended periods of time may be analogous to the small percentage of chronic offenders who are responsible for a large percentage of crime.⁸ To date, little is known about the actual life cycle of crime hot spots. Sherman (1995) and Spelman (1995) were first to invoke many features of offender criminal careers to describe careers of hot spots in terms of processes like initiation, growth, crime-type hardening or escalation in crime seriousness, persistence, decline, displacement, and termination. Their research also broke new ground by beginning to empirically explore the merits of this characterization, looking for evidence of sustained offending over time in some locations.

Minimally, crime hot spots share the key features of a boundary and criminal events within that boundary (e.g., 911 calls, offense reports). Perhaps the easiest means of identifying hot spots is to partition a jurisdiction into a fixed set of boundaries (e.g., square grid cells, census block groups, or some other boundary set) and to develop a set of rules (a “rule base”) using threshold values. Sherman and Weisburd (1995) objectively defined hot spots in terms of location, time interval, crime types, and number of events.

Suppose that the boundaries are square grid cells of a fixed size and origin. Then a rule for hot spot *initiation* at any grid cell might be the following: If the cell were not a hot spot in the previous time period but the number of crimes of a designated type now exceeds a specified threshold value, then the cell becomes a hot spot during the current period. A rule base will incorporate life cycle states, time intervals, threshold crime counts, and changes in crime counts. Gorr, Olligschlaeger, and Szczyplula (1998) are designing such a rule base to empirically explore the numbers, durations, branching probabilities, crime mixes, and concentrations of hot spots.

The choice of boundaries—fixed or ad hoc—is of particular interest in representing hot spots. Fixed boundaries (e.g., census tracts, police precincts, or uniform grid cells) have the advantage of giving rise to the space/time series data commonly used for crime reporting and spatial modeling. Their disadvantage is that hot spots may cross the fixed boundaries or vary in size. One example of ad hoc clustering of observed crime point data are the ellipses created in STAC (Spatial and Temporal Analysis of Crime) software (Block 1994). Such boundaries have the advantage of yielding sizes and shapes tuned specifically to individual hot spots. They, however, do not yield a consistent collection of space and time series data on crimes and enforcement activities. As modelers, we prefer fixed boundaries.

Hot spots are, by definition, small in area. Using visual inspection of pin maps and threshold counts for “hard” and “soft” crimes, Sherman and Weisburd

(1995) identified hot spots in Minneapolis, Minnesota, of no more than one linear block of a street—an area in which a police officer can easily see and be seen. Hot times were between 7:00 p.m. and 3:00 a.m. For drug markets in Jersey City, New Jersey, hot spots were defined by intersections and the four connected street blocks, and hot times were from noon to midnight (Weisburd and Green 1994). This scale may be too small for most practical purposes. For example, hot spots may move short distances over short time periods (e.g., displacement to nearby locations in response to enforcement activities). Hot spot areas of a few blocks in size or even larger may better accommodate such sporadic short-term moves within what is essentially the same activity space.

The presence of variations in the estimated effects of models arising from differences in the areal units that are selected for analysis is known as the modifiable areal unit problem (MAUP) in geography (Holt et al. 1996). Widely varying parameter estimates can result from reaggregating data by areal units of different sizes. For example, Gehlke and Biehl (1934) found that correlation coefficients tended to increase with the level of aggregation of census tracts. Fotheringham and Wong (1991) found that changes in the parameter estimates of multiple linear regression models were complex and unpredictable when changing the scale at which data were collected and aggregated. Though important in theory, MAUP is likely to be of less concern in analyses of hot spots, because size is often dictated by the need to represent crime hot spots for enforcement purposes, and this function constrains the range of relevant sizes.

Hot spot modeling and analysis

Understanding the relationship between place and crime requires knowledge of the dynamics of hot spot development over space and time, with special attention to the ways that a location's facilities and utilization contribute to criminal behavior. This sort of knowledge can be derived from combining theory with exploratory and confirmatory empirical research. Several kinds of spatial models and analyses are appropriate for hot spots. Preliminary to actual causal models, these include descriptive models and predictive models.

Descriptive models

The life cycle of hot spots includes various stages of development, the duration of time spent in each stage, and branching probabilities of transitions between the stages. A better understanding of hot spots requires space and time data of crime and its covariates for a sample of cities. Those data should include a consistent rule base for classifying fixed areas into non-hot spots and hot spots at different stages of development. Then analysts will have a better basis for distinguishing random stochastic phenomena, such as regression to the mean

(some hot spots fade on their own accord), from systematic hardening of soft-crime hot spots to more serious crimes.

Predictive models

After description, the next step in understanding hot spots is building successful predictive models. For example, the “broken windows” hypothesis posits that a variety of soft crimes (e.g., vandalism and public order disturbances) serve as leading indicators of serious crimes like assault and robbery. Leading indicator models require multivariate data that include the dependent variable (e.g., number of robberies per month) along with precursor leading indicator variables that are lagged one or more time periods (e.g., number of gang- or drug-related 911 calls from prior months). Lags may also be over space, such as a simple total or weighted average of 911 calls at contiguous (nearby) locations in prior months.⁹

The Vector Autoregression (VAR) model is a common time series model for estimating and testing leading indicators. Researchers have used VAR models extensively for applied modeling and forecasting since the work of Sims (1980). These are simple multivariate models in which a variable is explained by its own past values and past values of all other variables (leading indicators) in the system (Holden 1995). The Bayesian Vector Autoregression (BVAR) model is a restricted form of VAR.

Introduced by Litterman (1980, 1986), BVAR relies on Bayes’ estimates of priors to overcome collinearity and degrees of freedom problems that typically arise in applications of vector autoregressive models. Doan, Litterman, and Sims (1984) introduced the so-called Minnesota priors for BVAR. LeSage and Pan (1995) introduced spatial contiguity to further specify the priors in regional studies. BVAR models have been successful in time series analysis and forecasting models for regional data, especially in exploratory analyses of the appropriate time- and space-lagged model specifications (LeSage 1989, 1990; LeSage and Pan 1995).

Granger and Newbold (1977, 224–226) introduced rules and tests for a weak form of causality testing based on VAR and relative to the limited information set of variables used. Now known as “Granger causality,” Factor A “Granger causes” B if Lag A is a significant predictor of B, but Lag B is not a significant predictor of A. Enders (1995, 315) presents a standard F-test to determine Granger causality.

Exploratory spatial data analysis

Recently, the set of methods for structuring the visualization of spatial data has been referred to as *exploratory spatial data analysis*, or ESDA. As defined by Anselin (1994, 1998, 1999a), ESDA is a collection of techniques to describe and visualize spatial distributions; identify atypical locations or spatial outliers; discover patterns of spatial association, clusters, or hot spots; and suggest spatial regimes or other forms of spatial heterogeneity (changing structure or changing association across space). As such, ESDA forms a subset of exploratory data analysis or EDA (Tukey 1977), but with an explicit focus on the distinguishing characteristics of geographical data (Anselin 1989). In this section, we outline how principles from ESDA are relevant in the analysis of spatial patterns in crime. Specifically, we start by reviewing the concept of spatial autocorrelation and how it can be applied to both point data (e.g., location of burglaries) and areal data (e.g., number of homicides or homicide rate per census tract). We next outline some recently developed approaches that focus on “local” indicators of spatial association (or LISA) and discuss how these may be used to detect hot spots and spatial outliers. Finally, we review the integration of these techniques in an interactive computing environment.

The interest in quantification of patterns in maps has led to a large number of spatial statistics and other map summaries, reviewed in the classic treatments of spatial autocorrelation by Cliff and Ord (1973, 1981). Similarly, detection of clusters and outliers in maps is a major concern in epidemiology and medical statistics, and a large body of literature is devoted to the topic (e.g., as reviewed in Marshall 1991). Formally, the presence or absence of pattern is indicated by the concept of *spatial autocorrelation*, or the co-occurrence of similarity in value with similarity in location. In other words, when high values in a place tend to be associated with high values at nearby locations, or low values with low values for the neighbors, positive spatial autocorrelation or *spatial clustering* is said to occur. In contrast, when high values at a location are surrounded by nearby low values, or vice versa, negative spatial autocorrelation is present in the form of *spatial outliers*. The point of reference in the analysis of spatial autocorrelation is spatial randomness, or the lack of any structure. For example, under spatial randomness, the particular arrangement of crimes on a given map would be just as likely as any other arrangement, and any grouping of high or low values in a particular area would be totally spurious.

Point pattern analysis

The formal assessment of the presence and extent of spatial autocorrelation depends on the type of data under consideration. The simplest situation is when only the location of a given phenomenon is known (for example, the street

addresses where burglaries occurred). In this situation, the primary interest lies in assessing whether these locations, abstracted as points on a map, are seemingly randomly scattered across space, or instead, show systematic patterns in the form of clusters (more points are systematically closer together than they would be in a purely random case) or dispersion (more points are systematically further away from each other than under randomness). Point pattern analysis is concerned with detecting when “significant” deviations from spatial randomness occur.¹⁰

Quadrat count method. The construction of tests for point patterns may be approached in a number of different ways. A popular technique that is easily carried out in a GIS environment is the quadrat count method, in which a square grid is overlaid on the points. The number of points in each grid cell is counted and compared with “expected” number under spatial randomness by means of a chi-squared test of goodness-of-fit. While intuitive and readily implemented, this approach suffers from a number of conceptual problems, such as arbitrariness in the choice of the grid cell size and the possibility of correlation between counts in nearby cells (spatial autocorrelation).

Kernel estimation. A natural extension of the quadrat approach is kernel estimation, in which a smooth estimate of the intensity of the point process is derived by means of a moving window over the data. In other words, the number of points within the moving window (sometimes transformed to improve interpretation and visualization) is taken as an indicator of the intensity of the event at that location (e.g., how many burglaries per square mile). Rather than the points themselves, this intensity measure can be visualized in a map and assessed for systematic deviations from randomness. A particular implementation of this technique consists of drawing many overlapping circles of variable sizes and assessing the extent to which “clusters” may be present. For example, this is implemented in the “geographical analysis machine” of Openshaw and associates (Openshaw et al. 1987, 1988; for a recent review, see Openshaw and Alvanides 1999).¹¹

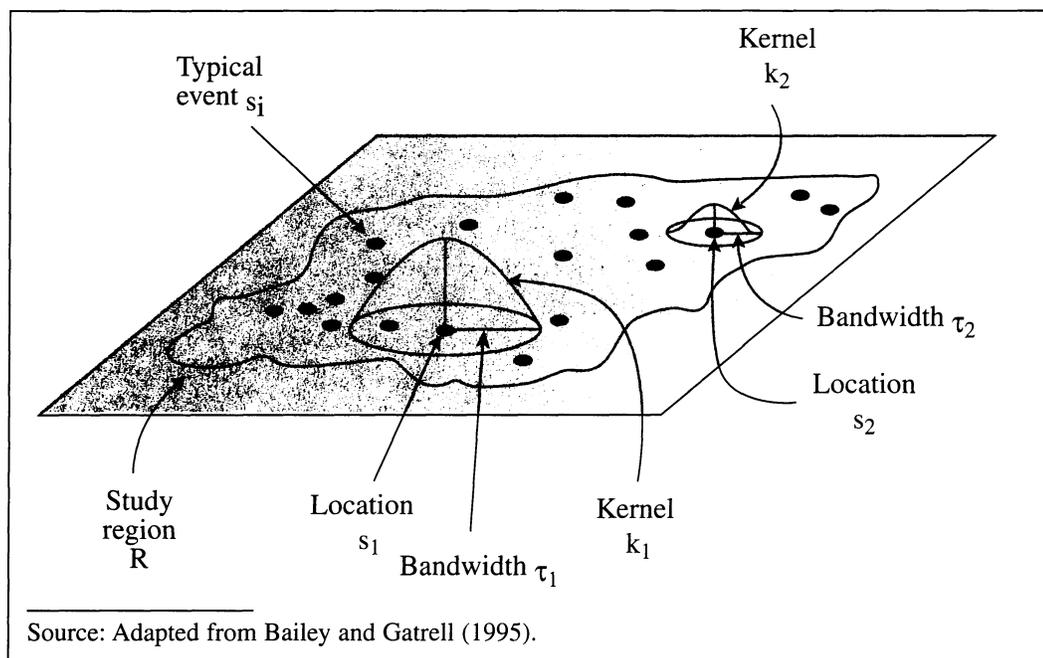
Kernel estimation or kernel smoothing is one method for examining large-scale global trends in point data. The goal of kernel estimation is to estimate how event levels vary continuously across a study area based on an observed point pattern for a sample of points (Bailey and Gatrell 1995; Williamson et al. 1998). Kernel estimation creates a smooth map of values using spatial data. The smoothed map appears like a spatially based histogram, with the level at each location along the map reflecting the point pattern intensity for the surrounding area.

In kernel estimation, a moving three-dimensional function (k_1) of a given radius or “bandwidth” visits every cell of a fine grid that has been overlaid on the study region or area. As the kernel visits each cell, distances are measured from the center of the grid cell (s_1) to each observation (s_i) falling within the bandwidth (τ_1). Each distance contributes to the intensity level of that grid cell, with greater weight given to observations lying closer to the center of the cell (see exhibit 1).

The choice of an appropriate bandwidth is crucial when applying kernel estimation to point data, and can prove a significant weakness if selected arbitrarily (see Silverman 1986). Bandwidth is crucial because it determines the amount of smoothing applied to a point pattern. In general, a large bandwidth will result in a large amount of smoothing, producing a fluid map with low intensity levels. A smaller bandwidth results in less smoothing, producing a spiky map with local variations in intensity levels. Ideally, bandwidth should represent the actual distance between the points in the distribution. However, there is no steadfast rule for determining bandwidth.¹²

Kernel estimation has been applied across a number of different fields, particularly epidemiology. In epidemiological applications, a distribution of discrete points, each of which represents the incidence of disease among the population, is transformed into a continuous smoothed surface map indicating disease risk (see Sabel 1998).¹³ By transforming spatial point patterns of criminal incidents

Exhibit 1. Kernel estimation



into a smooth image, kernel estimation can be equally effective in visualizing areas of criminal activity and risk.

Kernel estimation offers several practical benefits in the spatial analysis of crime. The first benefit is accessibility. Kernel estimation allows analysts to visually simplify and examine complex point patterns of criminal incidents. The greater accessibility of point data on crime incidents can easily result in data overload (Block 1998). Displaying even modest amounts of point data on a map can quickly become confusing and uninformative. Kernel estimation does not diminish the import of point-based spatial data. Instead, a smooth image captures and displays hot spots and potential hot spots as areas of high density. These areas of high density can then be verified by examining the level of statistical significance of estimated hot areas to determine the likelihood of observing levels this high if incidents are in fact distributed randomly over space and time.

Kernel estimation also allows greater flexibility in defining the borders of hot spots and in analyzing hot spot areas. Hot spot areas are often influenced by natural boundaries that break up population areas, such as gullies and highways. These boundaries make the areas irregular in shape. In addition, concentrations of crime often flow across police beats and jurisdictions rather than being confined to predefined administrative boundaries. Therefore, whenever the distribution of crime is not uniform, the contours that define hot spot areas are unlikely to be the well-behaved circles or ellipses required in some crime-clustering methods (e.g., STAC method in Block 1998). Kernel estimation allows for flexible boundaries and the display of the intensity of criminal incidents across an entire region.

Finally, kernel estimation can be important in analyzing incident patterns over time. Density images can be compared for consecutive or corresponding time periods (e.g., the same month or year-to-date comparisons in successive years). These provide a context for interpreting short-term changes in relation to long-term trends and seasonal patterns. Kernel smoothed maps also reveal the larger spatial context of changes over time.

Distance statistics

Other point pattern techniques are based on the distance between the points, either between each point and its nearest neighbor (nearest neighbor statistics) or between all the points (second order statistics). The underlying rationale is that when events are clustered in space, small interpoint distances should be more prevalent than under spatial randomness. A large number of nearest neighbor statistics have been suggested in the literature.¹⁴ Their properties

are either derived or approximated analytically, or, more interestingly, based on a computational approach. The latter consists of simulating the location of the same number of points as in the dataset (e.g., the total number of burglaries in a given year) by randomly assigning locations, thus mimicking the null hypothesis of spatial randomness. For each of the simulated patterns, the value of the statistic (or statistics) can be computed, thus yielding a reference distribution to which the statistic for the observed pattern can be compared. This provides an intuitive and highly visual way to assess the degree of nonrandomness in a point pattern. For example, this can be applied to the empirical cumulative distribution function for the nearest neighbor distances for each point, or to all the distances between points.¹⁵

Nearest neighbor statistics have been extended to test for clusters in space and time. For example, the Knox statistic (Knox 1964) consists of counting how many pairs of events are closer in space and time than would be the case under randomness.¹⁶ Although initially developed to detect clusters of disease incidence, the application of these methods to criminal activity is straightforward.

The techniques discussed so far address so-called “general” levels of clustering (or, global spatial autocorrelation) in the sense of assessing the extent to which spatial randomness can be rejected. In many instances, it is interesting to locate “where” the clusters may be present. For example, one may be interested in finding out if the clusters center around particular locations of crime-inducing facilities, such as liquor stores or 24-hour convenience stores. Such tests are

A fundamental concept in the analysis of spatial autocorrelation for areal data is the spatial weights matrix. The spatial weights matrix is used to formalize a notion of locational similarity and is central to every test statistic.

referred to as “focused” tests (Besag and Newell 1991) and relate the number of points in a cell (or counts of events) to the distance from a “putative source.” Again, the general principle underlying these tests is that deviations from spatial randomness would yield a higher frequency of points close to the putative source.¹⁷

Estimation of spatial point processes

Tests for point patterns or clusters of events are implemented in a number of software packages, several of which are readily available commercially or as freeware/shareware. Most of these can also easily be integrated in a GIS environment. For example, nearest neighbor statistics and the K function are included in the S+SpatialStats add-on to the S-Plus statistical package (MathSoft 1996), which can be integrated

with the ArcView GIS through the S+ArcView link (Bao et al. forthcoming).¹⁸ Infomap (Bailey and Gatrell 1995) contains a number of quadrat count methods, as well as nearest neighbor and second order statistics, together with a set of basic mapping functions. A wide range of cluster and scan statistics are also included in Stat! (BioMedware 1994), which, although developed with health events in mind, can be readily applied to crime statistics. A specific focus on pattern detection in the locations of crime incidents is implemented in the CrimeStat package. This software tool, which was developed with the support of the National Institute of Justice, can be linked to a variety of commercial GIS software and spatial data formats (Levine 1999).

Areal analysis

So far, the discussion of spatial autocorrelation has dealt with situations where the data come in the form of points, and their location is the primary focus of interest. An equally important setting is that in which the data are collected for areal units or “regions,” such as homicide counts or rates by county or census tract.¹⁹ A large number of spatial autocorrelation tests have been developed to assess the extent to which the spatial arrangement of values on a map shows deviations from a null hypothesis of spatial randomness, as reviewed in Cliff and Ord (1973, 1981), Upton and Fingleton (1985), and Griffith (1987), among others.

A fundamental concept in the analysis of spatial autocorrelation for areal data is the spatial weights matrix. This is a square matrix of dimension equal to the number of observations, with each row and column corresponding to an observation. Typically, an element w_{ij} of the weights matrix W is non-zero if locations i and j are neighbors, and zero otherwise (by convention, the diagonal elements w_{ii} equal zero). A wide range of criteria may be used to define neighbors, such as binary contiguity (common boundary) or distance bands (locations within a given distance of each other), or even general “social” distance.²⁰ The spatial weights matrix is used to formalize a notion of locational similarity and is central to every test statistic. In practice, spatial weights are typically derived from the boundary files or coordinate data in a geographic information system (see Can 1996).

Viewed from a more technical standpoint, almost every test for “global” spatial autocorrelation can be expressed as a special case of a general cross-product or “gamma” statistic (Hubert 1985, 1987; Hubert, Golledge, and Costanzo 1981). This statistic consists of a sum of cross-products between two sets of terms, one related to the similarity in value between two observations, the other to their similarity in location, or, $\Gamma = \sum_i \sum_j a_{ij} \cdot w_{ij}$. In this expression, the a_{ij} term corresponds to value similarity, such as a cross product, $x_i \cdot x_j$, or a squared difference

$(x_i - x_j)^2$, while the w_{ij} are elements in a spatial weights matrix. Inference for this general class of statistics is based on permutation. Specifically, a reference distribution is constructed that simulates spatial randomness by arbitrarily rearranging the values observed in a given map over the available locations and recomputing the statistic for each of these random arrangements.

Classic statistics

Classic test statistics for spatial autocorrelation are the join count statistic, Moran's I and Geary's c (Cliff and Ord 1973). The join count statistic is appropriate when the data are binary, for example, the presence (coded B for black) or absence (coded W for white) of arson fires by city block. The number of times neighboring spatial units also have B in common is called a BB join count.²¹ The tests are based on the extent to which the observed number of BB joins (or, WW, BW) is compatible with a null hypothesis of spatial randomness. Similarly, when the data are variables measured on a continuous scale (such as crime rates or counts of homicides), Moran's I and Geary's c statistics measure the deviation from spatial randomness. Moran's I is a cross-product coefficient similar to a Pearson correlation coefficient and scaled to be less than one in absolute value. Positive values for Moran's I indicate positive spatial autocorrelation (clustering), while negative values suggest spatial outliers.²² In contrast to Moran's I, Geary's c coefficient is based on squared deviations. Values of Geary's c less than one indicate positive spatial autocorrelation, while values larger than one suggest negative spatial autocorrelation.²³ Adjustments to Moran's I to account for the variance instability in rates have been suggested in the epidemiological literature, for example, the Ipop statistic of Oden (1995). Extensions of Moran's I to a multivariate setting are outlined in Wartenberg (1985).

Moran scatterplot

When variables are used in standardized form (that is, their mean is zero and standard deviation one), the degree of spatial autocorrelation in a dataset can be readily visualized by means of a special scatterplot, termed Moran scatterplot in Anselin (1995, 1996). The Moran scatterplot is centered on the mean and shows the value of a variable (z) on the horizontal axis against its spatial lag (Wz , or $\sum_j w_{ij}z_j$; i.e., a weighted average of the neighboring values) on the vertical axis. The four quadrants in the scatterplot correspond to locations where high values are surrounded by high values in the upper right (an above mean z with an above mean Wz), or low values are surrounded by low values in the lower left, both indicating positive spatial autocorrelation. The two other quadrants correspond with negative spatial autocorrelation, or high values surrounded

by low values (high z , low Wz) and low values surrounded by high values (low z , high Wz). The slope of the linear regression line through the Moran scatterplot is Moran's I coefficient. Moreover, a map showing the locations that correspond to the four quadrants provides a summary view of the overall patterns in the data. Hence, this device provides an intuitive means to visualize the degree of spatial autocorrelation, not only in a traditional cross-sectional setting, but also across variables and over time (Anselin 1998). Recent illustrative examples of the application of these concepts in homicide studies can be found in Sampson, Morenoff, and Earls (1999) and Cohen and Tita (1999).

Distance-based statistics

An alternative perspective on spatial autocorrelation for data available at discrete locations (points, areas) is to consider these as sampling points for an underlying continuous surface in a geostatistical approach. For example, crime statistics by police station would be used to estimate a continuous crime surface for the whole city. The primary interest in this paradigm lies in spatial interpolation, or kriging. The measure of spatial autocorrelation is taken to be a function of the squared difference between the values for each pair of observations compared with the distance that separates them. Formally, this is carried out in a variogram (or, more precisely, a semi-variogram).²⁴ One visualization of the variogram consists of a scatterplot of the squared differences organized by distance band, possibly with a box plot for each distance band—a variogram cloud plot or variogram box plot (see Cressie 1993; Haslett et al. 1991). Another visualization focuses on each distance lag separately, in a spatially lagged scatterplot (Cressie 1984). The mean or median in the variogram cloud plot for each distance band suggests an overall pattern for the change in spatial autocorrelation with distance, and a focus on outliers indicates pairs of observations that may unduly influence this central tendency (see also Majure and Cressie 1997; Anselin 1998, 1999a).

Local indicators of spatial association—LISA statistics

The measures of spatial autocorrelation reviewed so far are general, or global, in the sense that the overall pattern in the data is summarized in a single statistic. Paralleling the focused tests of point pattern analysis, *local* indicators of spatial association (LISA) provide a measure of the extent to which the arrangement of values around a specific location deviates from

Modern computational implementations of exploratory data analysis are based on the paradigm of dynamically linked windows, in which the user interacts with different "views" of the data on a computer screen.

spatial randomness. Closely related to the focused tests, the G_i and G_i^* statistics of Getis and Ord (1992; Ord and Getis 1995) measure the extent to which the concentration of high or low values within a given distance band around a location deviates from spatial randomness. These statistics are designed to find clusters of high or low values. They can be applied to each location in turn or to using increasing distance bands away from a given location. A general framework for LISA is outlined in Anselin (1995), where local forms are derived for several global statistics, such as the local Moran and local Geary statistics.²⁵ The local Moran is closely related to the Moran scatterplot and indicates the presence of local clusters or local spatial outliers. LISA statistics lend themselves well to visualization by means of a GIS, for example, in symbol maps that show the locations with significant local statistics. In addition, when combined with a Moran scatterplot, the locations with significant local Moran can be classified in terms of the type of association they represent.

Estimation of spatial autocorrelation

Routines to test for spatial autocorrelation are found in a wide range of special purpose as well as commercial software. An extensive listing is given on the AI-GEOSTAT's Web site (<http://curie.ei.jrc.it:80/software>). Other recent reviews can be found in Legendre (1993) and Levine (1996). Most of these software implementations are specialized and contain one or a few statistic(s). Comprehensive treatments are the S+SpatialStats add-on for the S-Plus statistical system (MathSoft 1996) and the SpaceStat™ package (Anselin 1992). The latter is the only system to date that contains both global and local spatial statistics. It is integrated with the ArcView GIS by means of an extension, which, among other things, allows for the visualization of Moran scatterplot maps and locations with significant LISA.²⁶

Modern computational implementations of exploratory data analysis are based on the paradigm of dynamically linked windows, in which the user interacts with different "views" of the data on a computer screen. The views typically consist of standard statistical graphics such as histograms, box plots, and scatterplots, but increasingly include a map as well. The dynamic linking consists of allowing an analyst who uses a pointing device (mouse) to establish connections between data points in different graphs, highlight (brush) subsets of the data and rotate, cut through, and project high dimensional data (for a recent review, see Buja, Cook, and Swayne 1996). Geographical data can easily be included in this framework when viewed as x, y points in a standard scatterplot. A more extensive framework that also includes choropleth maps was originally proposed in the Spider software of Haslett, Unwin, and associates (Haslett, Wills, and Unwin 1990; Haslett et al. 1991; Unwin 1996; Unwin et al. 1996).

Recent efforts in this regard also incorporate explicitly spatial statistics, such as a variogram cloud or box plot, Moran scatterplot, and LISA maps in a spatial association visualizer (Anselin 1998).

A practical example of a link between a GIS and various exploratory data analysis tools is represented by the work of Symanzik and colleagues (1998, 1999). Here, a form of software integration is obtained between the ArcView GIS and the XGobi (Buja, Cook, and Swayne 1996) and XploRe (Härdle, Klinke, and Turlach 1995) EDA software packages. This link is based on point data and allows for the brushing of a variogram cloud plot. Each of the exploration tools can be used in isolation or linked with the other. As long as the data are represented by points, powerful visualization can be obtained, including space-time dynamics and complex multivariate linkages.²⁷ For example, this method can be used to track the location and frequency of a given type of crime across space as well as over time and to suggest potentially useful correlates. A similar approach is taken in the implementation of dynamically linked windows in the DynESDA extension for ArcView (Anselin and Smirnov 1998). Here, a view in ArcView is augmented with a series of statistical graphs, including histograms, boxplots, scatterplots, and Moran scatterplots that are all dynamically linked. Using brushing and linking techniques, both multivariate as well as spatial association between a number of variables can be assessed and visualized. A recent application of this technique to the study of the spatial diffusion of homicides is given by Messner and colleagues (1999).

Until recently, the role of space (and space-time) was not explicitly acknowledged in the methodology used in these studies, but it is central in a number of respects. Spatial concentration will tend to result in spatial autocorrelation, which runs counter to the usual assumption of independence in regression analysis. In addition, strong spatial variation in criminal activity suggests the need for an explicit spatial perspective and the consideration of spatial heterogeneity (spatial structural change).

Spatial Modeling

The techniques of exploratory analysis reviewed in the previous section are extremely useful in assessing the existence and location of nonrandom local patterns in spatial data. However, they are also limited by the lack of mechanisms to “explain” the observed patterns. EDA and ESDA are exploratory by nature. They “suggest” potential associations between variables and elicit hypotheses, but the formal testing of these hypotheses is left for confirmatory

As is now increasingly recognized, in this instance specialized methods of spatial regression analysis (spatial econometrics) must be used to avoid potentially biased results and faulty inference due to the presence of so-called spatial effects, consisting of spatial dependence and spatial heterogeneity.

analysis, typically carried out by means of multivariate regression modeling (Anselin and Getis 1992).

In the specific context of criminal justice, regression analysis plays a crucial role in the attempts to explain the causes of criminal activity (e.g., Land, McCall, and Cohen 1990; Kposowa and Breault 1993; DeFronzo and Hannon 1998). Until recently, the role of space (and space-time) was not explicitly acknowledged in the methodology used in these studies, but it is central in a number of respects. For example, it is well known that urban crimes such as theft and burglary, as well as most categories of violent crimes, are likely to be spatially concentrated in low-income urban areas that have relatively high proportions of unemployed persons and racial minorities. This spatial concentration will tend to result in spatial autocorrelation, which runs counter to the usual assumption of independence in regression analysis. In addition, law enforcement efforts (Chambliss 1994) and gang activity (Cohen et al. 1998) vary spatially, strongly suggesting the need for an explicit spatial perspective (Roncek 1993) and the consideration of spatial heterogeneity (spatial structural change). A spatial perspective

is further motivated by the findings of large-scale spatial differences for various crimes (for example, urban, suburban, and rural as reported in the Federal Bureau of Investigation's Uniform Crime Reports as well as the Bureau of Justice Statistics' semiannual National Crime Victimization Survey). This in turn has prompted a search for spatial mechanisms such as proximity and diffusion to explain these phenomena (Tolnay, Deane, and Beck 1996; Morenoff and Sampson 1997; Sampson, Morenoff, and Earls 1999).

The challenge of spatial effects

In most of these studies, the regression analysis employs data for cross-sectional units, such as census tracts or counties. As is now increasingly recognized, in this instance specialized methods of spatial regression analysis (spatial econometrics) must be used to avoid potentially biased results and faulty inference (Anselin 1988; Anselin and Bera 1998). This is due to the presence of *spatial effects*, consisting of *spatial dependence* and *spatial heterogeneity*, which violate the basic assumptions underlying classical regression analysis. Statisticians have long been aware of the problems associated with analyzing spatial (geographical) data, but spatial statistical techniques did not disseminate into the

empirical practice of the mainstream social sciences until recently (for a review, see Anselin 1999b).²⁸

The motivation for the explicit incorporation of spatial effects in regression models that explain criminal activity is twofold. On the one hand, crime and enforcement data are readily geocoded, but the spatial scale of observation does not necessarily match the spatial scale of the process under study. For example, the occurrence of certain types of crimes, say dealing in illicit drugs, may be explained by socioeconomic variables and land use data collected at the block level. However, if the illicit drug trading zone for a given group covers multiple blocks, the data for several units of observation will be correlated. Similarly, if the unmodeled variables such as “social capital” or “sense of community” spill over across multiple units of observation, a spatial correlation of these “errors” will result. Hence, the concern with accounting for the presence of spatial autocorrelation in a regression model is driven by the fact that the analysis is based on *spatial data* for which the unit of observation is largely arbitrary (such as administrative units). The methodology focuses on making sure that the estimates and inference from the regression analysis (whether for spatial or a-spatial models) are correct in the presence of spatial autocorrelation.

On the other hand, much recent theoretical work in urban sociology, economics, and criminology has emphasized concepts related to the “interaction” of agents, such as copycatting, social norms, neighborhood effects, diffusion, and other peer group effects. These theories focus on questions of how individual interactions can lead to emergent collective behavior and aggregate patterns (e.g., Brock and Durlauf 1995; Akerlof 1997; Durlauf 1994; Borjas 1995; Glaeser, Sacerdote, and Scheinkman 1996). Here, the need for an explicit *spatial model* is driven by theoretical concerns and the interest lies in a correct specification of the form and range of interaction and the estimation of its strength.

The two different motivations for consideration of spatial effects in regression models lead to methods to handle spatial dependence as a nuisance (data problems) versus substantive spatial dependence (theory driven). Formally, this results in techniques to model spatial dependence in the error terms of the regression model or to transform the variables in the model to eliminate spatial correlation (spatial filtering), versus methods to explicitly add a spatial interaction variable as one of the regressors in the model.

Spatial statistical techniques

The two different motivations for consideration of spatial effects in regression models lead to methods to handle spatial dependence as a *nuisance* (data problems) versus *substantive* spatial dependence (theory driven) (Anselin 1989). Formally, this results in techniques to model spatial dependence in the error terms of the regression model or to transform the variables in the model to eliminate spatial correlation (spatial filtering), versus methods to explicitly add a spatial interaction variable as one of the regressors in the model. Common to all methodological approaches is the need to rigorously express the notion of “neighbor effects,” which is based on the concept of a spatial weights matrix, discussed previously. A spatially explicit variable takes the form of a “spatial lag” or spatially lagged dependent variable, which consists of a weighted average of the neighboring values. More precisely, the spatial lag of a dependent variable at location i , y_i , would be $\sum_j w_{ij}y_j$, where the weighted sum is over those “neighbors” j that have a nonzero value for element w_{ij} in the weights matrix (or, in general, the weight is w_{ij}). For practical purposes, the elements of the spatial weights matrix are often row-standardized, which facilitates interpretation and comparison across models (for technical details, see Anselin 1988, forthcoming; Anselin and Bera 1998).

A typical specification of a linear regression equation that expresses substantive spatial interaction (or spatial autocorrelation) is the *mixed regressive, spatial autoregressive model*, or *spatial lag* model. This includes, in addition to the usual set of regressors (say, x_i , the regressive part), a spatially lagged dependent variable $\sum_j w_{ij}y_j$, (the spatial autoregressive part), with a spatial autoregressive coefficient ρ .²⁹ The inclusion of a spatial lag term is similar to a temporal autoregressive term in time series analysis, although there are several important differences that require a specialized methodology for estimation and testing.³⁰

The interpretation of the spatial lag model is best illustrated with a simple example. Say we were interested in explaining the crime rate by the usual socioeconomic variables as well as by a police intervention measure, and assume that the data are collected at the census-tract level. The spatial lag would capture the average crime rate for neighboring tracts. This measure of “potential” crime is one way to formalize the spatial interaction in the model. Therefore, the significance and value of the autoregressive coefficient have a direct interpretation as an indication of the strength of the spatial interaction. In our example, the estimate for ρ would suggest to what extent the crime rate in each census tract is “explained” by the average of the neighbors.

There are two potential pitfalls in this interpretation. First, the spatial lag does not “explain” anything (similar to a time lag in time series), but instead is a

proxy for the simultaneity in the whole system. This is best seen in a formal way, but for the sake of simplicity can be thought of as a *spatial multiplier*. After transforming the model to reduced form, so only “exogenous” variables remain on the right-hand side of the equation, it follows that the value of y at each location (e.g., the crime rate) depends not only on the explanatory variables for that location (the x_i), but also on these variables at all other locations, suitably adjusted to reflect the effect of distance decay. In our example, the presence of a spatial multiplier implies that a change in police intervention at one location (census tract) not only affects the crime rate at that location, but at all other locations in the system as well (suitably decayed), hence the notion of a multiplier.³¹

The second problem is due to the use of aggregate entities, such as census tracts or counties, as observational units. The interpretation of the autoregressive term as an indication of “interaction” between units can easily lead to an “ecological fallacy.” This follows from the fact that these units are not social agents themselves, but only aggregates (averages) of individual behavioral units. Drawing inferences for individual behavior from relations observed at the aggregate level can only be carried out under a strict set of assumptions (essentially imposing extreme homogeneity), which is clearly unwarranted in the current context (for an extensive discussion, see King 1997).³² An alternative interpretation is that the spatial lag model allows for *filtering out* the potentially confounding effect of spatial autocorrelation in the variable under consideration. The main motivation for this is to obtain the proper inference on the coefficients of the other covariates in the model (the β). For example, spatial autocorrelation of the lag variety may result from a mismatch between the spatial extent of the criminal activity and the census tract as the spatial unit of observation.³³

From an estimation point of view, the problem with this model is that the spatial lag term contains the dependent variables for neighboring observations, which in turn contain the spatial lag for their neighbors, and so on, leading to simultaneity (the spatial multiplier effect mentioned previously). This simultaneity results in a nonzero correlation between the spatial lag and the error term, which violates a standard regression assumption. Consequently, ordinary least squares (OLS) estimation will yield inconsistent (and biased) estimates, and inference based on this method will be flawed. Instead of OLS, specialized estimation methods must be employed that properly account for the spatial simultaneity in the model. These methods are either based on the maximum likelihood (ML) principle, or on the application of instrumental variable (IV) estimation in a spatial two-stage, least-squares approach.³⁴

In contrast to the lag model, there are a number of ways to incorporate the spatial autocorrelation into the structure of the regression model error term. The

most commonly used models are based on spatial processes, such as a spatial autoregressive (SAR) or spatial moving average (SMA) process, in parallel to the time series convention.³⁵ The particular form for the process yields a nondiagonal covariance structure for the errors, with the value and sign of the off-diagonal elements corresponding to the “spatial correlation” (that is, the correlation between the error terms at two different locations).

An interesting aspect of this correlation structure is the range of interaction that is implied. For a SAR process, every error term is correlated with every other error term, but the magnitude of the correlation follows a distance decay effect. In other words, the implied interaction is global, as in the spatial multiplier of the spatial lag model. In contrast, the SMA process yields local interaction, where only first and second order neighbors have a nonzero correlation. Since this pertains to the error terms in a model, or the “ignored” or “unmeasurable” effects, the two specifications also have different policy implications. For example, if there were an unmeasurable “neighborhood” effect in our model of crime, the SAR specification would imply that change in this effect in one location affects all the locations in the system, whereas in an SMA specification this change would only affect the immediate neighbors. However, more precisely, these measurement errors only pertain to the precision of the estimates, and “on average” their impact is zero on the predicted crime, in contrast to the spatial multiplier in the lag model, in which shocks pertaining to the regressor (X) are transmitted throughout the system.

In space, the error variances are also heteroskedastic, which is not the case in the time domain (see Anselin and Bera 1998). The heteroskedasticity is induced by the spatial process and will complicate specification testing (i.e., distinguishing “true” heteroskedasticity from that induced by a spatial process). This is an important distinction between the spatial error processes and their covariance structure and the time series counterpart.

An alternative approach to handling spatial processes is to specify the magnitude of the spatial error covariance as a function of the distance that separates pairs of observations.³⁶ This “direct representation” approach is inspired by geostatistical modeling and lends itself well to spatial forecasting (or interpolation). In contrast to the spatial process models, there is no induced heteroskedasticity. However, for the direct representation approach to yield a valid covariance (e.g., to avoid negative variances), a number of restrictive assumptions must be satisfied (see e.g., Cressie 1993; Anselin forthcoming).

The estimation of spatial error models falls under the generic category of regression models with nonspherical error variance. Technically, a form of generalized least squares will be applied, although in contrast to the time domain,

there is no simple two-step estimation procedure. Instead, an explicit maximum likelihood approach or generalized moment technique must be followed.³⁷ In these methods, the coefficient of the spatial model is considered a “nuisance” parameter in the sense that it improves the precision of the estimates for the regressors (β), but in and of itself is of little interest.

Compared with spatial dependence, spatial effects in the form of spatial heterogeneity can be handled in a fairly straightforward way with standard econometric models. The resulting heteroskedasticity, varying coefficients, or structural instability is only distinct in the sense that the specification of the heterogeneity is in terms of *spatial* or regional differences (e.g., different crime rates in central city versus suburb). However, because spatial heterogeneity often occurs jointly with spatial dependence (or the two are observationally equivalent), explicit consideration of the latter is required in empirical applications. Examples of techniques that address spatial heterogeneity are spatial analysis of variance (Sokal et al. 1993), spatially varying coefficients as some form of hierarchical linear modeling in the spatial expansion method (Jones and Casetti 1992; Casetti 1997), locally different regression coefficients in the spatial adaptive filter (Foster and Gorr 1986; Gorr and Olligschlaeger 1994), geographically weighted regression (Brunsdon, Fotheringham, and Charlton 1996; McMillen and McDonald 1997), and the correction for spatial outliers by means of Bayesian techniques (LeSage 1997, 1999).

When observations are available for a cross-section at different points in time, in the form of panel data, it becomes possible to model complex combinations of spatial heterogeneity and spatial dependence.³⁸ For example, different model coefficients can be specified for different subregions and/or different time periods; the spatial autoregressive coefficients can be allowed to vary over time, etc. The types of methods appropriate for addressing such models consist of seemingly unrelated regressions, error components, and Bayesian approaches, in conjunction with a spatial lag or spatial error dependence. Overviews of the methodological issues are given in Anselin (1988, ch. 10; 1990b, 1999b) and LeSage (1995).

In practice, the most important aspect of spatial modeling may well be specification testing. In fact, even if discovering spatial interaction of some form is not of primary interest, ignoring spatial lag or spatial error dependence when it is present creates serious model misspecification. Of the two spatial effects, ignoring lag dependence is the more serious offense, since, as an omitted variable problem, it results in biased and inconsistent estimates for all the coefficients in the model; and the inference derived from these estimates is flawed. When spatial error dependence is ignored, the resulting OLS estimator remains unbiased, although it is no longer most efficient. Moreover, the estimates for

the OLS coefficient standard errors will be biased, and, consequently, t-tests and measures of fit will be misleading.

Spatial model estimation

Tests for the presence of potential spatial effects are complicated by a number of factors. First, as mentioned earlier, spatial processes yield heteroskedastic errors, so that it will be difficult to distinguish true heteroskedasticity from that induced by the spatial processes. The reverse is true as well, so that tests against spatial dependence will be sensitive to the presence of heteroskedasticity and may point to the wrong alternative. Second, the spatial lag and spatial error specifications are highly related, so that tests against one form of dependence will also have power against the other form, again complicating the specification search. Third, all tests for spatial effects are based on large sample properties (asymptotics) and their performance in small data sets may be suspect.

If spatial features serve as actuating factors for crime, either because of the people or facilities that are located there, then interventions designed to alter those persons and activities might well affect crime. Alternatively, if the spatial distribution of crime is essentially random, then targeting specific places is not likely to be an effective crime control strategy.

Despite these problems, however, there are a number of practical guidelines that can be followed in empirical applications. The most straightforward testing approach is to use Lagrange Multiplier tests that are based on the residuals of an OLS regression. Separate tests are available for a spatial lag and a spatial error alternative, and a simple rule of thumb exists to guide the researcher in the proper direction (the most significant test suggests the proper alternative).³⁹ Other tests with high power are based on the application of Moran's I to regression residuals, which is a valid misspecification test against a wide range of alternatives and applicable in various econometric specifications (Anselin and Kelejian 1997; Kelejian and Robinson 1998, 1999; Kelejian and Prucha 1999b; Pinkse 1999).

Spatial econometric methods are not routinely incorporated in commercial software packages. Hence, several authors have developed "tricks" to carry out estimation and specification testing using macro or script facilities in statistical computing software. Examples are routines in Limdep, Gauss, Shazam, and S-Plus in Anselin and Hudak (1992), and maximum likelihood estimation in SAS (Griffith 1993), Matlab (Pace and Barry 1998), or R (Bivand 1999). Estimation of spatial error models is included in the S+SpatialStats add-on to the S-Plus software

(MathSoft 1996), but the only comprehensive suite of routines to handle both specification testing and estimation is contained in SpaceStat (Anselin 1992).

Conclusion

As we near the end of the 20th century, spatial analyses of crime remain poised at the early stages of development. Many of the capabilities to support computerized mapping and spatial statistical analyses emerged only recently during the 1990s. The promise of using spatial data and analyses for crime control still remains to be demonstrated, and its usefulness depends on the nature of the relationship between crime and place. If spatial features serve as actuating factors for crime, either because of the people or facilities that are located there, then interventions designed to alter those persons and activities might well affect crime. Alternatively, if the spatial distribution of crime is essentially random, then targeting specific places is not likely to be an effective crime control strategy.

Research aimed at sorting out the nature of the relationship between place and crime is crucial and becoming increasingly feasible as spatial data capabilities proliferate. One of the first priorities is research on the nature of crime hot spots, especially the typical life course (or crime “career”) of areas with high concentrations of crime, to determine whether the unusually high levels persist for any length of time. While spatial analyses remain a promising tool, the very early stage of research on the relationship between crime and place is reason for a degree of caution. Considerably more research is needed before we look to location as a primary target for crime control efforts. Both basic social science research and well-designed applied research on specific police interventions will be of value.

Notes

1. An overview of the literature, including the theoretical underpinnings of the crime-place relationship and empirically based results, is available in Eck and Weisburd (1995). See also Weisburd (1997).
2. See, for example, Sampson (1986) and Eck and Weisburd (1995).
3. See Ritzer (1988) for a critique of the Chicago School.
4. See the 1969 University of Chicago Press edition with new chapters of updated data.
5. The panel consisted of James Short, Robert Sampson, Robert Bursik, Ruth Horowitz, Karen Heimer, and Ross Matsueda.

6. For examples of empirical and theoretical treatments of social disorganization and social control theories, see Bursik (1986, 1988), Bursik and Grasmick (1993), Sampson (1986, 1995), Sampson and Groves (1989), Abbott (1997), and Wikstrom (1998).

7. To those physical dimensions, we would add the general disenfranchisement of residents from having a stake in maintaining both the physical and social environment of this special type of housing.

8. Wolfgang and colleagues highlighted the impact of chronic offenders within birth cohorts of boys in Philadelphia (Wolfgang, Figlio, and Sellin 1972; Wolfgang, Thornberry, and Figlio 1987; Tracy, Wolfgang, and Figlio 1990). Similarly, skewed distributions of offending, in which a small number of offenders are responsible for a disproportionately large share of crimes, have been repeatedly confirmed both in official criminal records (Blumstein and Cohen 1979; Shannon 1988; Haapanen 1990), and in self-reports by offenders (Petersilia, Greenwood, and Lavin 1977; Peterson and Braiker 1981; Chaiken and Chaiken 1982, 1985; English 1990; Miranne and Geerken 1991; Horney and Marshall 1991). Such distributions are quite common for a wide array of behaviors (Greenberg 1991).

9. See discussion of spatial lags in the subsection titled "Areal analysis."

10. Highly readable introductions to point pattern analysis are given in Boots and Getis (1988), Upton and Fingleton (1985), and Bailey and Gatrell (1995). More advanced materials are contained in Cliff and Ord (1973), Ripley (1981), Diggle (1983), and Cressie (1993), among others.

11. Similar ideas, but with a more formal probabilistic basis, underlie the cluster tests of Besag and Newell (1991), Turnbull et al. (1990), and Kulldorff's scan statistic (Kulldorff and Nagarwalla 1995; Nagarwalla 1996; Hjalmars et al. 1996). The principle of using a spatial window to compute smoothed rates from a GIS consisting of the addresses of the occurrence of an event is also discussed in Rushton and Lolonis (1996).

12. There are several "rules of thumb" for determining bandwidth. For a discussion of these rules of thumb, see Bailey and Gatrell (1995) and Williamson et al. (1998).

13. Surface maps arise naturally for representing geographical variations in risk. They also can be derived for other attributes thought to be associated with an outcome. For example, results from a sample of point estimates of crime offending risk by age and birth cohort of offenders could be transformed to a smoothed surface of risk over all ages and cohorts.

14. For example, Cressie (1993, 604) lists no less than 17 nearest neighbor statistics. See also Ord (1990) for an extensive review.

15. This is the familiar K-function outlined in Ripley (1976); see also Bailey and Gatrell (1995, ch. 3).

16. For an extension and generalization, see Ederer, Myers, and Mantel (1964), Mantel (1967), and Bailey and Gatrell (1995, ch. 4).

17. See also Diggle (1991), Lawson (1993), Diggle and Rowlingson (1994), Waller et al. (1995), and Waller (1996) for reviews of the methodology.

18. A more comprehensive set of S-based software for point pattern analysis is contained in the Splancs (spatial point pattern analysis code in s-plus) routines of Rowlingson and Diggle (1993); see also Venables and Ripley (1998).

19. In Cressie's (1993) taxonomy of spatial data statistics, these types of data are referred to as lattice data.

20. For an extensive discussion of spatial weights, see Cliff and Ord (1981), Upton and Fingleton (1985), and Anselin (1988).

21. Formally, a BB joint count statistic is $(1/2) \sum_i \sum_j w_{ij} x_i x_j$, where w_{ij} are the elements of a binary spatial weights matrix (i.e., one for neighbors, zero for others) and x_i, x_j take on 1 for B, 0 for W. Similarly, WW joint counts are $(1/2) \sum_i \sum_j w_{ij} (1-x_i)(1-x_j)$. See Cliff and Ord (1973, 1981) and Upton and Fingleton (1985) for extensive treatments.

22. Moran's $I = (N/S_0) \sum_i \sum_j w_{ij} z_i z_j / \sum_i z_i^2$, where the z_i are variables in deviations from the mean, w_{ij} are elements of a possibly row-standardized spatial weights matrix and S_0 is a scaling factor equal to the sum of all the elements in the weights matrix. For details, see Cliff and Ord (1973, 1981), and Upton and Fingleton (1985).

23. Geary's $c = (N-1)/2S_0 [\sum_i \sum_j w_{ij} (x_i - x_j)^2 / \sum_i z_i^2]$, where the x_i are the original variables and z_i deviations from the mean; the other notation is as in footnote 22. For details, see Cliff and Ord (1973, 1981), and Upton and Fingleton (1985).

24. For a formal treatment, see Cressie (1993). A more introductory overview is offered in Isaaks and Srivastava (1989).

25. LISA are different from regional measures of spatial autocorrelation, which are global statistics applied to a subset of the data, as in Munasinghe and Morris (1996).

26. S+SpatialStats also contains a set of functions to carry out geostatistical analysis. Specialized routines for geostatistics and kriging can be found in Deutsch and Journel (1992) and Pannatier (1996).

27. For further details, see also Cook et al. (1996, 1997).

28. An extensive review of the statistical perspective is given in Cressie (1993).

29. In matrix notation, with y as an N by 1 vector of observations on the dependent variable, X as an N by K matrix of observations on the explanatory variables with regression coefficient vector β , Wy as a vector of spatially lagged dependent variables with spatial autoregressive coefficient ρ , and ϵ as a vector of random (independent,

identically distributed, or i.i.d) errors, the model can be expressed as $y = \rho Wy + X\beta + \varepsilon$. An extensive discussion of technical issues can be found in Anselin (1988, 1999b) and Anselin and Bera (1998).

30. See Anselin and Bera (1998) for an extensive discussion of the differences between dependence in spatial models and in time series models.

31. Formally, the reduced form is $y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$, where W is the N by N spatial weights matrix and I is an identity matrix (see Anselin 1988).

32. The ecological fallacy problem is known in spatial analysis as the “modifiable areal unit problem,” which in essence means that different results will be found when the size and arrangement of the spatial units of observation changes. The classic reference is Openshaw (1979).

33. For a more extensive discussion of the idea behind spatial filtering, see Getis (1995). Also note the difference between this concept of filtering and the spatial adaptive filter model of Foster and Gorr (1986), which offers an approach to deal with spatial heterogeneity.

34. The original ML estimator is due to Ord (1975) (see also Anselin 1988 and Cressie 1993 for technical details). IV estimation is outlined in Anselin (1988, 1990a), Land and Deane (1992), and Kelejian and Robinson (1993), among others.

35. In matrix notation, a spatial autoregressive error process (SAR) can be expressed as $\varepsilon = \lambda W\varepsilon + u$, whereas a spatial moving average process (SMA) is $\varepsilon = \lambda Wu + u$ (with ε as the regression error terms, λ as the spatial parameter, W as the weights matrix, and u as a vector of i.i.d. errors).

36. This type of model is commonly used in real estate analysis, originally in Dubin (1988). See also Olmo (1995) and Basu and Thibodeau (1998) for some recent examples.

37. Reviews of the relevant issues can be found in Ord (1975), Anselin (1988), and Kelejian and Prucha (1997, 1998, 1999a).

38. The panel data setting is different from true space-time dynamics, for example, as the basis of space-time forecasting.

39. Reviews are given in Anselin and Florax (1995), Anselin et al. (1996), and Anselin (forthcoming).

References

Abbott, Andrew. 1997. Of time and space: The contemporary relevance of the Chicago School. *Social Forces* 75:1149–1182.

Akerlof, George A. 1997. Social distance and social decision. *Econometrica* 65:1005–1027.

Anselin, Luc. Forthcoming. Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*.

———. 1999a. Interactive techniques and exploratory spatial data analysis. In *Geographical information systems*. 2d ed. Edited by P. Longley, M. Goodchild, D. Maguire, and D. Rhind. New York: John Wiley & Sons.

———. 1999b. Spatial econometrics. In *Companion in theoretical econometrics*, edited by B. Baltagi. Oxford: Basil Blackwell.

———. 1998. Exploratory spatial data analysis in a geocomputational environment. In *Geocomputation, a primer*, edited by P. Longley, S. Brooks, R. McDonnell, and B. Macmillan. New York: John Wiley & Sons.

———. 1996. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial analytical perspectives on GIS*, edited by M. Fischer, H. Scholten, and D. Unwin. London: Taylor and Francis.

———. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27:93–115.

———. 1994. Exploratory spatial data analysis and geographic information systems. In *New tools for spatial analysis*, edited by M. Painho. Luxembourg: Eurostat.

———. 1992. SpaceStat™, a software package for the analysis of spatial data. Santa Barbara: University of California, Santa Barbara, National Center for Geographic Information and Analysis. Retrieved 30 April 2000 from the World Wide Web: <http://www.spacestat.com>.

———. 1990a. Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics* 20:141–163.

———. 1990b. Spatial dependence and spatial structural instability in applied regression analysis. *Journal of Regional Science* 30:185–207.

———. 1989. What is special about spatial data. Technical Report 89–4. Santa Barbara: University of California, Santa Barbara, National Center for Geographic Information and Analysis. [Contained on CD-ROM. Fundamental research in geographic information and analysis. Santa Barbara: National Center for Geographic Information and Analysis.]

———. 1988. *Spatial econometrics: Methods and models*. Boston: Kluwer Academic Publishers.

Anselin, Luc, and Anil Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of applied economic statistics*, edited by Amman Ullah and David E.A. Giles. New York: Marcel Dekker.

- Anselin, L., A. Bera, R. Florax, and M. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26:77–104.
- Anselin, Luc, and Raymond Florax. 1995. Small sample properties of tests for spatial dependence in regression models: Some further results. In *New directions in spatial econometrics*, edited by L. Anselin and R. Florax. Berlin: Springer-Verlag.
- Anselin, Luc, and Arthur Getis. 1992. Spatial statistical analysis and geographic information systems. *Annals of Regional Science* 26:19–33.
- Anselin, Luc, and Sheri Hudak. 1992. Spatial econometrics in practice, a review of software options. *Regional Science and Urban Economics* 22:509–536.
- Anselin, Luc, and Harry H. Kelejian. 1997. Testing for spatial autocorrelation in the presence of endogenous regressors. *International Regional Science Review* 20:153–182.
- Anselin, L., and O. Smirnov. 1998. The *DynESDA extension for ArcView*. Retrieved 30 April 2000 from the World Wide Web: <http://www.spacestat.com/downloads.htm>.
- Bailey, Trevor C., and Anthony Gatrell. 1995. *Interactive spatial data analysis*. Harlow, England: Longman Scientific & Technical.
- Bao, S., L. Anselin, D. Martin, and D. Stralberg. Forthcoming. Seamless integration of spatial statistics and GIS: the S-Plus for ArcView and the S+Grassland Links. *Journal of Geographical Systems*.
- Basu, Sabyasachi, and Thomas G. Thibodeau. 1998. Analysis of spatial autocorrelation in housing prices. *Journal of Real Estate Finance and Economics* 17:61–85.
- Besag, J., and J. Newell. 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A* 154:143–155.
- BioMedware. 1994. *Stat! Statistical software for the clustering of health events*. Ann Arbor, Michigan: BioMedware.
- Bivand, Roger. Forthcoming. Implementing functions for spatial statistical analysis using the R language. *Journal of Geographical Systems*.
- Block, Carolyn R. 1998. The GeoArchive: Information foundation for community policing. In *Crime mapping and crime prevention*, edited by D. Weisburd and T. McEwen. Monsey, New York: Criminal Justice Press.
- . 1994. *STAC hot spot areas: A statistical tool for law enforcement decisions*. Chicago: Illinois Criminal Justice and Information Authority. Retrieved 30 April 2000 from the World Wide Web: <http://www.icjia.state.il.us/public/pdf/stac/Hotspot.pdf>.

Block, Richard, and Carolyn Block. 1995. Space, place, and crime: Hot-spot areas and hot places of liquor-related crime. In *Crime and place*, edited by J.E. Eck and D. Weisburd. Vol. 4 of *Crime prevention studies*. Monsey, New York: Criminal Justice Press.

Blumstein, Alfred, and Jacqueline Cohen. 1979. Estimation of individual crime rates from arrest records. *Journal of Criminal Law and Criminology* 70:561–585.

Boots, Barry, and Arthur Getis. 1988. *Point pattern analysis*. Newbury Park, California: Sage Publications.

Borjas, George. 1995. Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review* 85:365–390.

Boydstrun, J.E. 1975. *San Diego field interrogation: Final report*. Washington, D.C.: Police Foundation.

Brantingham, P.L., and P.J. Brantingham. 1993. Environment, routine, and situation: Toward a pattern theory of crime. In *Routine activity and rational choice: Advances in criminological theory*, edited by Ronald V. Clarke and Marcus Felson. Vol. 5. New Brunswick, New Jersey: Transaction Publishers.

———. 1982. Mobility, notoriety, and crime: A study of crime patterns in urban nodal points. *Journal of Environmental Systems* 11:89–99.

Brock, William A., and Steven N. Durlauf. 1995. Discrete choice with social interactions. I: Theory. Working paper no. W5291. National Bureau of Economic Research, Cambridge, Massachusetts.

Brunsdon, C., A. Fotheringham, and M. Charlton. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28:281–298.

Buja, A., D. Cook, and D.F. Swayne. 1996. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics* 5:78–99.

Bursik, Robert J. 1988. Social disorganization and theories of crime and delinquency: Problems and prospects. *Criminology* 26:519–551.

———. 1986. Ecological stability and the dynamics of delinquency. In *Communities and crime*, edited by A.J. Reiss, Jr., and M. Tonry. Vol. 8 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

Bursik, Robert J., and Harold G. Grasmick. 1993. *Neighborhoods and crime: The dimensions of effective community control*. New York: Lexington Books.

Can, A. 1996. Weight matrices and spatial autocorrelation statistics using a topological vector data model. *International Journal of Geographical Information Systems* 10:1009–1017.

- Casetti, Emilio. 1997. The expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review* 20:9–33.
- Chaiken, Jan M., and Marcia Chaiken. 1985. Identifying high-rate serious offenders. Draft report. Santa Monica, California: RAND.
- . 1982. *Varieties of criminal behavior*. Santa Monica, California: RAND.
- Chambliss, W.J. 1994. Policing the ghetto underclass: The politics of law and law enforcement. *Social Problems* 41 (May): 177–194.
- Clarke, Ronald V., ed. 1992. *Situational crime prevention: Successful case studies*. Albany, New York: Harrow & Heston.
- Cleveland, W.S. 1993. *Visualizing data*. Summit, New Jersey: Hobart Press.
- Cliff, A.D., and J.K. Ord. 1981. *Spatial processes, models, and applications*. London: Pion.
- . 1973. *Spatial autocorrelation*. London: Pion.
- Cohen, Bernard. 1980. *Deviant street networks*. Lexington, Massachusetts: Lexington Books.
- Cohen, Jacqueline, Wilpen Gorr, and Andreas Olligschlaeger. 1993. Modeling street-level illicit drug markets. Working paper 93–64, The H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh.
- Cohen, J., D. Cork, J. Engberg, and G. Tita. 1998. The role of drug markets and gangs in local homicide rates. *Homicide Studies* 2:241–262.
- Cohen, J., and G. Tita. 1999. Spatial diffusion in homicide: An exploratory analysis. *Journal of Quantitative Criminology* 15 (4): 451–493.
- Cohen, Lawrence E., and Marcus Felson. 1979. Social change and crime rate trends: A routine activity approach. *American Sociological Review* 44:588–605.
- Cook, D., J. Majure, J. Symanzik, and N. Cressie. 1996. Dynamic graphics in a GIS: Exploring and analyzing multivariate spatial data using linked software. *Computational Statistics* 11:467–480.
- Cook, D., J. Symanzik, J. Majure, and N. Cressie. 1997. Dynamic graphics in a GIS: More examples using linked software. *Computers and Geosciences* 23:371–385.
- Cornish, L.E., and R.V. Clarke. 1986. *The reasoning criminal: Rational choice perspectives on offending*. New York: Springer-Verlag.

Cressie, N. 1984. Towards resistant geostatistics. In *Geostatistics for natural resources characterization*, Part I, edited by G. Verly, M. Davic, A.G. Journel, and A. Marechal. Dordrecht, The Netherlands: D. Reidel.

Cressie, Noel A.C. 1993. *Statistics for spatial data*. New York: John Wiley & Sons.

Curry, David, and Irving Spergel. 1988. Gang homicide, delinquency, and community. *Criminology* 26:381–405.

DeFronzo, J., and L. Hannon. 1998. Welfare assistance levels and homicide rates. *Homicide Studies* 1:31–45.

Deutsch, C.V., and A.G. Journel. 1992. *GSLIB geostatistical software library and user's guide*. New York: Oxford University Press.

Diggle, Peter. 1991. A point process modeling approach to raised incidence of a rare phenomenon in the vicinity of a pre-specified point. *Journal of the Royal Statistical Society A* 153:349–362.

———. 1983. *Statistical analysis of spatial point patterns*. London: Academic Press.

Diggle, P., and B.S. Rowlingson. 1994. A conditional approach to point process modeling of elevated risk. *Journal of the Royal Statistical Society A* 157:433–440.

Doan, T., R.B. Litterman, and C. Sims. 1984. Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews* 3:1–100.

Dubin, Robin. 1988. Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Review of Economics and Statistics* 70:466–474.

Duffala, D.C. 1976. Convenience stores, armed robbery, and physical environmental features. *American Behavioral Scientist* 20:227–246.

Durkheim, Emile. [1897] 1966. *Suicide*. Translated by John A. Spaulding and George Simpson. New York: Free Press.

———. [1901] 1950. *The rules of the sociological method*. Translated by S.A. Solovay and J.H. Mueller. Glencoe, Illinois: Free Press.

Durlauf, Steven N. 1994. Spillovers, stratification, and inequality. *European Economic Review* 38:836–845.

Eck, John E. 1994. Drug markets and drug places: A case control study of the spatial structure of illicit drug dealing. Ph.D. diss., University of Maryland.

Eck, John E., and David Weisburd. 1995. Crime places in crime theory. In *Crime and place*, edited by J.E. Eck and D. Weisburd. Vol. 4 of *Crime prevention studies*. Monsey, New York: Criminal Justice Press.

- Ederer, F., H.M. Myers, and N. Mantel. 1964. A statistical problem in space and time: Do leukemia cases come in clusters? *Biometrics* 20:626–638.
- Enders, Walter. 1995. *Applied econometric time series*. New York: John Wiley & Sons.
- English, Kim. 1990. Measuring crime rates in Colorado: 1988–89. Paper presented at the Annual Meeting of the American Society of Criminology, November, Baltimore.
- Engstad, P.A. 1975. Environmental opportunities and the ecology of crime. In *Crime in Canadian society*, edited by R.A. Silverman and J.J. Teevan. Toronto: Butterworths.
- Felson, Marcus. 1994. *Crime and everyday life: Impact and implications for society*. Thousand Oaks, California: Pine Forge Press.
- . 1986. Linking criminal choices, routine activities, informal crime control, and criminal outcomes. In *The reasoning criminal: Rational choice perspectives on offending*, edited by D. Cornish and R.V. Clarke. New York: Springer-Verlag.
- Foster, S., and W. Gorr. 1986. An adaptive filter for estimating spatially-varying parameters: Application to modeling police hours spent in response to calls for service. *Management Science* 32:878–889.
- Fotheringham, A.S., and D.W.S. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 32:1025–1044.
- Gahegan, Mark, and David O'Brien. 1997. A strategy and architecture for the visualization of complex geographical datasets. *International Journal of Pattern Recognition and Artificial Intelligence* 11:239–261.
- Gehlke, C., and Biehl, K. 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association* 29:169–170.
- Getis, A. 1995. Spatial filtering in a regression framework: Examples using data on urban crime, regional inequality, and government expenditures. In *New directions in spatial econometrics*, edited by L. Anselin and R. Florax. Berlin: Springer-Verlag.
- Getis, A., and J.K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24:189–206.
- Glaeser, E., B. Sacerdote, and J. Scheinkman. 1996. Crime and social interactions. *Quarterly Journal of Economics* 111:507–548.
- Gorr, W., and A. Olligschlaeger. 1994. Weighted spatial adaptive filtering: Monte Carlo studies and application to illicit drug market modeling. *Geographical Analysis* 26:67–87.

Gorr, Wilpen L., Andreas M. Olligschlaeger, and Janusz Szczypula. 1998. Crime hot spot forecasting: Modeling and comparative evaluation. Paper prepared under NIJ grant no. 98-IJ-CX-K005, U.S. Department of Justice, National Institute of Justice, Washington, D.C.

Granger, C.W.J., and Paul Newbold. 1977. *Forecasting economic time series*. New York: Academic Press.

Greenberg, David F. 1991. Modeling criminal careers. *Criminology* 29:17-46.

Greenberg, S., W. Rohe, and J. Williams. 1985. *Informal citizen action and crime prevention at the neighborhood level: Synthesis and assessment of the research*. Executive Summary, NCJ 97977. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Griffith, Daniel. 1993. *Spatial regression analysis on the PC: Spatial statistics using SAS*. Washington, D.C.: Association of American Geographers.

———. 1987. *Spatial autocorrelation, a primer*. Washington, D.C.: Association of American Geographers.

Guerry, A. [1833a] 1984. Note. In *Research on the propensity for crime at different ages*, by A. Quetelet. Translated by Sawyer F. Sylvester. Cincinnati: Anderson Publishing Company.

———. [1833b] 1974. *Essai sur la Statistique Morale de la France*. New York: Clearwater.

Haapanen, Rudy A. 1990. *Selective incapacitation and the serious offender*. New York: Springer-Verlag.

Härdle, W., S. Klinke, and B.A. Turlach. 1995. *XploRe: An interactive statistical computing environment*. Berlin: Springer-Verlag.

Haslett, J., R. Bradley, P. Craig, A. Unwin, and G. Wills. 1991. Dynamic graphics for exploring spatial data with applications to global and local anomalies. *American Statistician* 45:234-242.

Haslett, J., G. Wills, and A. Unwin. 1990. Spider, an interactive statistical tool for the analysis of spatially distributed data. *International Journal of Geographical Information Systems* 4:285-296.

Hjalmar, Ulf, Martin Kulldorff, Goran Gustafsson, and Neville Nagarwalla. 1996. Childhood leukaemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine* 15:707-715.

Holden, Ken. 1995. Vector autoregression modelling and forecasting. *Journal of Forecasting* 14:159-166.

- Holt, D., D.G. Steel, M. Tranmer, and N. Wrigley. 1996. Aggregation and ecological effects in geographically based data. *Geographical Analysis* 28:224–262.
- Homel, R., and J. Clark. 1995. The prediction and prevention of violence in pubs and clubs. In *Crime prevention studies*, edited by R.V. Clarke. Vol. 3. Monsey, New York: Criminal Justice Press.
- Horney, Julie, and Ineke H. Marshall. 1991. Measuring lambda through self reports. *Criminology* 29:471–496.
- Hubert, Lawrence. 1987. *Assignment methods in combinatorial data analysis*. New York: Marcel Dekker.
- . 1985. Combinatorial data analysis: Association and partial association. *Psychometrika* 50:449–467.
- Hubert, L.J., R. Golledge, and M. Costanzo. 1981. Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis* 13:224–233.
- Isaaks, E.H., and R.M. Srivastava. 1989. *An introduction to applied geostatistics*. Oxford: Oxford University Press.
- Jeffrey, C.R. 1971. *Crime prevention through environmental design*. Beverly Hills: Sage Publications.
- Jones, J.P., and E. Casetti. 1992. *Applications of the expansion method*. London: Routledge.
- Kelejian, Harry H., and Ingmar Prucha. 1999a. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40 (2): 509–533.
- . 1999b. On the asymptotic distribution of the Moran I Test statistic with applications. Working paper, Department of Economics, University of Maryland, College Park.
- . 1998. A generalized two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17:99–121.
- . 1997. Estimation of spatial regression models with autoregressive errors by two stage least squares procedures: A serious problem. *International Regional Science Review* 20:103–111.
- Kelejian, Harry H., and Dennis P. Robinson. 1999. The influence of spatially correlated heteroskedasticity on tests for spatial correlation. In *Advances in spatial econometrics*, edited by L. Anselin and R. Florax. Heidelberg: Springer-Verlag.

———. 1998. A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics* 28:389–417.

———. 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in Regional Science* 72:297–312.

Kelling, G.L., and C.M. Coles. 1996. *Fixing broken windows: Restoring order and reducing crime in our communities*. New York: Free Press.

King, Gary. 1997. *A solution to the ecological inference problem*. Princeton, New Jersey: Princeton University Press.

Knox, E.G. 1964. The detection of space-time interactions. *Applied Statistics* 13:25–29.

Kposowa, A., and K.D. Breault. 1993. Reassessing the structural covariates for U.S. homicide rates: A county level study. *Sociological Forces* 26:27–46.

Kulldorff, Martin, and Neville Nagarwalla. 1995. Spatial disease clusters: Detection and inference. *Statistics in Medicine* 14:799–810.

Land, K., and G. Deane. 1992. On the large-sample estimation of regression models with spatial- or network-effects terms: A two stage least squares approach. In *Sociological methodology*, edited by P. Marsden. San Francisco: Jossey-Bass.

Land, K., P. McCall, and L. Cohen. 1990. Structural covariates of homicide rates: Are there invariances across time and social space? *American Journal of Sociology* 95:922–963.

Lawson, A.B. 1993. On the analysis of mortality events associated with a prespecified fixed point. *Journal of the Royal Statistical Society A* 156:363–377.

Legendre, Pierre. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology* 74:1659–1673.

LeSage, James. 1999. A family of geographically weighted regression models. In *Advances in spatial econometrics*, edited by L. Anselin and R. Florax. Heidelberg: Springer-Verlag.

———. 1997. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20:113–129.

———. 1995. A multiprocess mixture model to estimate space-time dimensions of weekly pricing of certificates of deposit. In *New directions in spatial econometrics*, edited by L. Anselin and R. Florax. Berlin: Springer-Verlag.

- . 1990. Forecasting turning points in metropolitan employment growth rates using Bayesian techniques. *Journal of Regional Science* 30:533–548.
- . 1989. Incorporating regional wage relations in local forecasting models with a Bayesian prior. *International Journal of Forecasting* 5:37–47.
- LeSage, J.P., and Z. Pan. 1995. Using spatial contiguity as Bayesian prior information in regional forecasting models. *International Regional Science Review* 18:33–53.
- Levine, N. 1999. *CrimeStat: A spatial statistics program for the analysis of crime incident locations*. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.
- . 1996. Spatial statistics and GIS: Software tools to quantify spatial patterns. *Journal of the American Planning Association* 62:381–391.
- Litterman, R.B. 1986. Forecasting with Bayesian vector autoregressions—Five years of experience. *Journal of Business & Economic Statistics* 4:25–38.
- . 1980. Techniques for forecasting with vector autoregressions. Ph.D. diss., University of Minnesota.
- MacEachren, A.M., and M.-J. Kraak. 1997. Exploratory cartographic visualization: Advancing the agenda. *Computers and Geosciences* 23:335–343.
- Majure, J., and N. Cressie. 1997. Dynamic graphics for exploring spatial dependence in multivariate spatial data. *Geographical Systems* 4:131–158.
- Mantel, N. 1967. The detection of disease clustering and a generalised regression approach. *Cancer Research* 27:209–220.
- Marshall, Roger J. 1991. A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society A* 154:421–441.
- MathSoft. 1996. *S+SpatialStats*. Seattle: MathSoft.
- McMillen, Daniel, and John F. McDonald. 1997. A nonparametric analysis of employment density in a polycentric city. *Journal of Regional Science* 37:591–612.
- Messner, S., L. Anselin, R. Baller, D. Hawkins, G. Deane, and S. Tolnay. 1999. The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *Journal of Quantitative Criminology* 15 (4): 423–450.
- Miranne, Alfred C., and Michael Geerken. 1991. The New Orleans inmate survey: A test of Greenwood's predictive scale. *Criminology* 29:497–518.
- Morenoff, J., and R.J. Sampson. 1997. Violent crime and the spatial dynamics of neighborhood transition: Chicago 1970–1990. *Social Forces* 76:31–64.

Munasinghe, R., and R. Morris. 1996. Localization of disease clusters using regional measures of spatial autocorrelation. *Statistics in Medicine* 15:893–905.

Nagarwalla, Neville. 1996. A scan statistic with a variable window. *Statistics in Medicine* 15:845–850.

Newman, Oscar. 1972. *Defensible space*. New York: Macmillan.

Oden, N. 1995. Adjusting Moran's I for population density. *Statistics in Medicine* 12:1915–1929.

Olmo, J.C. 1995. Spatial estimation of housing prices and locational rents. *Urban Studies* 32:1331–1344.

Openshaw, Stan. 1979. A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In *Statistical methods in the spatial sciences*, edited by N. Wrigley. London: Pion.

Openshaw, S., and S. Albanides. 1999. Applying geocomputation to the analysis of spatial distributions. In *Geographical Information Systems*. 2d ed. Edited by P. Longley, M. Goodchild, D. Maguire, and D. Rhind. New York: John Wiley & Sons.

Openshaw, S., M. Charlton, C. Wymer, and A. Craft. 1987. A Mark 1 geographical analysis machine for the automated analysis of point datasets. *International Journal of Geographical Information Systems* 1:335–358.

Openshaw, S., A. Craft, M. Charlton, and J. Birch. 1988. Investigation of leukemia clusters by use of a geographical analysis machine. *Lancet* 1:272–273.

Ord, J. Keith. 1990. Statistical methods for point pattern data. In *Spatial statistics: Past, present, and future*, edited by D. Griffith. Ann Arbor: ImaGE.

———. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70:120–126.

Ord, J.K., and A. Getis. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27:286–306.

Pace, R.K., and R. Barry. 1998. *Spatial Statistics Toolbox 1.0*. Baton Rouge: Louisiana State University, Real Estate Institute.

Pannatier, Yvan. 1996. *Variowin, software for spatial data analysis in 2D*. Heidelberg: Springer-Verlag.

Park, R.E., E.W. Burgess, and R.D. McKenzie, eds. 1925. *The city*. Chicago: University of Chicago Press.

Pate, A., W.G. Skogan, M.A. Wycoff, and L.W. Sherman. 1985. *Reducing the signs of crime: Executive summary*. Washington, D.C.: Police Foundation.

Petersilia, Joan, Peter W. Greenwood, and M. Lavin. 1977. *Criminal careers of habitual offenders*. Santa Monica, California: RAND.

Peterson, Mark A., and Harriet B. Braiker. 1981. *Doing crime: A survey of California prison inmates*. Cambridge, Massachusetts: Oelgeschlager, Gunn, and Hain.

Pierce, G., S.A. Spaar, and L. Briggs IV. 1988. The character of police work: Strategic and tactical implications. M.S. thesis, Northeastern University, Center for Applied Social Research.

Pinkse, Joris. 1999. Moran-flavored test with nuisance parameters. In *Advances in spatial econometrics*, edited by L. Anselin and R. Florax. Heidelberg: Springer-Verlag.

Quetelet, Adolphe J. [1833] 1984. *Research on the propensity for crime at different ages*. Translated by Sawyer F. Sylvester. Cincinnati: Anderson Publishing Company.

———. [1842] 1969. *A treatise on man and the development of his faculties. A facsim. reproduction of the English translation of 1842*. Gainesville, Florida: Scholar's Facsimiles and Reprints.

Rheingans, P., and C. Landreth. 1995. Perceptual principles for effective visualizations. In *Perceptual issues in visualization*, edited by G. Grinstein and H. Levkowitz. Berlin: Springer-Verlag.

Ripley, Brian. 1981. *Spatial statistics*. New York: John Wiley & Sons.

———. 1976. The second order analysis of stationary point processes. *Journal of Applied Probability* 13:255–266.

Ritzer, George. 1988. *Contemporary sociological theory*. New York: Alfred A. Knopf.

Roncek, D. 1993. Mapping crime: An inescapable but valuable task for intracommunity analysis. In *Questions and answers in lethal and non-lethal violence: Proceedings of the second annual workshop of the Homicide Research Working Group*, edited Carolyn Block and Richard L. Block. NCJ 147480. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Roncek, Dennis W., and R. Bell. 1981. Bars, blocks, and crime. *Journal of Environmental Systems* 11:35–47.

Roncek, Dennis W., and J.M.A. Francik. 1981. Housing projects and crime: Testing the proximity hypothesis. *Social Problems* 29:151–166.

Roncek, Dennis W., and P.A. Maier. 1991. Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of hot spots. *Criminology* 29:725–755.

Roncek, Dennis W., and M.A. Pravatiner. 1989. Additional evidence that taverns enhance nearby crime. *Sociology and Social Research* 73:185–188.

Rowlingson, B.S., and P.J. Diggle. 1993. Splancs: Spatial point pattern analysis code in S-Plus. *Computers and Geosciences* 19:627–655.

Rushton, Gerard, and Panos Lolonis. 1996. Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine* 15:717–726.

Sabel, Clive. 1998. Peaks and troughs: Space-time cluster detection in rare diseases. Working paper presented at the Environmental Systems Research Institute International User Conference, 27–31 July. Retrieved 30 March 2000 from the World Wide Web: <http://www.esri.com/library/userconf/proc98/PROCEED.HTM>.

Sampson, Robert J. 1995. The community. In *Crime*, edited by James Q. Wilson and Joan Petersilia. San Francisco: ICS Press.

———. 1986. Crime in cities: The effects of formal and informal social control. In *Communities and crime*, edited by A.J. Reiss, Jr., and M. Tonry. Vol. 8 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

Sampson, Robert J., and Jacqueline Cohen. 1988. Deterrent effects of the police on crime: A replication and theoretical extension. *Law & Society Review* 22:165–189.

Sampson, Robert J., and W. Byron Groves. 1989. Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology* 94:774–802.

Sampson, Robert J., Jeffrey Morenoff, and Felton Earls. 1999. Beyond social capital: Neighborhood mechanism and structural sources of collective efficacy for children. *American Sociological Review* 64:633–660.

Shannon, Lyle W. 1988. *Criminal career continuity*. New York: Human Sciences Press.

Shaw, Clifford R., and Henry D. McKay. 1942. *Juvenile delinquency and urban areas: A study of rates of delinquents in relation to differential characteristics of local communities in American cities*. Chicago: University of Chicago Press.

Shaw, Clifford R., Frederick M. Zorbaugh, Henry D. McKay, and Leonard S. Cottrell. 1929. *Delinquency areas*. Chicago: University of Chicago Press.

Sherman, Lawrence W. 1995. Hot spots of crime and criminal careers of places. In *Crime and place*, edited by J.E. Eck and D. Weisburd. Vol. 4 of *Crime prevention studies*. Monsey, New York: Criminal Justice Press.

———. 1992. Attacking crime: Police and crime control. In *Modern policing*, edited by M. Tonry and N. Morris. Vol. 15 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

———. 1986. Policing communities: What works? In *Communities and crime*, edited by A.J. Reiss, Jr., and M. Tonry. Vol. 8 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

- Sherman, L.W., P.R. Gartin, and M.E. Buerger. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology* 27:27–55.
- Sherman, L.W., and D.A. Weisburd. 1995. General deterrent effects of police patrol in crime hot spots: A randomized, controlled trial. *Justice Quarterly* 12:625–648.
- Silverman, B.W. 1986. *Density estimation for statistics and data analysis*. London: Chapman & Hall.
- Sims, C.A. 1980. Macroeconomics and reality. *Econometrica* 48:1–48.
- Skogan, W.G., and M.G. Maxfield. 1981. *Coping with crime: Individual and neighborhood reactions*. Beverly Hills: Sage Publications.
- Sokal, R., N. Oden, B. Thomson, and J. Kim. 1993. Testing for regional differences in means: Distinguishing inherent from spurious spatial autocorrelation by restricted randomization. *Geographical Analysis* 25:199–210.
- Spelman, William. 1995. Criminal careers of public places. In *Crime and place*, edited by J.E. Eck and D. Weisburd. Vol. 4 of *Crime and prevention studies*. Monsey, New York: Criminal Justice Press.
- Symanzik J., S. Klinke, S. Schmelzer, D. Cook, and N. Lewin. 1998. The ArcView/XGobi/XploRe environment: Technical details and applications for spatial data analysis. ASA Proceedings of the Section on Statistical Graphics. Alexandria, Virginia: American Statistical Association.
- Symanzik J., T. Kötter, S. Schmelzer, S. Klinke, D. Cooke, D. Swayne. Forthcoming. Spatial data analysis in the dynamically linked ArcView/XGobi/XploRe environment. *Computing Science and Statistics*.
- Thrasher, Frederick. 1927. *The gang*. Chicago: University of Chicago Press.
- Tita, George E. 1999. An ecological study of urban street gangs and their impact on crime. Ph.D. diss., The H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh.
- Tita, George E., John Engberg, and Jacqueline Cohen. 1999. An ecological study of violent gangs: The social organization of set space. Working paper, National Consortium on Violence Research, Carnegie Mellon University, Pittsburgh.
- Tolnay, S.E., G. Deane, and E.M. Beck. 1996. Vicarious violence: Spatial effects on Southern lynchings, 1890–1919. *American Journal of Sociology* 102:788–815.
- Tracy, Paul E., Marvin E. Wolfgang, and Robert M. Figlio. 1990. *Delinquency careers in two birth cohorts*. New York: Plenum Press.

- Tukey, John W. 1977. *Exploratory data analysis*. Reading, Massachusetts: Addison-Wesley.
- Turnbull, B.W., E.J. Iwano, W.S. Burnett, H.L. Howe, and L.C. Clarke. 1990. Monitoring for clusters of disease: Application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 132:S136–S143.
- Unwin, A. 1996. Exploratory spatial analysis and local statistics. *Computational Statistics* 11:387–400.
- Unwin, A., G. Hawkins, H. Hofman, and B. Siegl. 1996. Interactive graphics for data sets with missing values—MANET. *Journal of Computational and Graphical Statistics* 5:113–122.
- Upton, Graham, and Bernard Fingleton. 1985. *Spatial data analysis by example*. New York: John Wiley & Sons.
- Venables, W., and B. Ripley. 1998. *Modern applied statistics with S-Plus*. 2d ed. New York: Springer-Verlag.
- Waller, Lance. 1996. Statistical power and design of focused clustering studies. *Statistics in Medicine* 15:765–782.
- Waller, Lance, Bruce Turnbull, Goran Gustafsson, Ulf Hjalmar, and Bertil Andersson. 1995. Detection and assessment of clusters of disease: An application to nuclear power plant facilities and childhood leukaemia in Sweden. *Statistics in Medicine* 14:3–16.
- Wartenberg, D. 1985. Multivariate spatial autocorrelation: A method for exploratory geographical analysis. *Geographical Analysis* 17:263–283.
- Weisburd, David. 1997. *Reorienting crime prevention research and policy: From the causes of criminality to the context of crime*. Research Report, NCJ 165041. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.
- Weisburd, D.L., and L. Green. 1994. Defining the street level drug market. In *Drugs and crime: Evaluating public policy initiatives*, edited by D.L. MacKenzie and C.D. Uchida. Beverly Hills: Sage Publications.
- Wikstrom, Per-Olaf. 1998. Communities and crime. In *Handbook of crime and punishment*, edited by Michael Tonry. New York: Oxford University Press.
- Williamson, Doug, Sara McLafferty, Victor Goldsmith, John Mollenkopf, and Phil McGuire. 1998. Smoothing crime incident data: New methods for determining the bandwidth in Kernel estimation. Working paper presented at the Environmental Systems Research Institute International User Conference, 27–31 July. Retrieved 30 March 2000 from the World Wide Web: <http://www.esri.com/library/userconf/proc98/PROCEED.HTM>.

Wilson, J.Q., and B. Boland. 1978. The effect of the police on crime. *Law & Society Review* 12:367–390.

Wilson, J.Q., and G.L. Kelling. 1982. Broken windows: The police and neighborhood safety. *Atlantic Monthly* 249:29–38.

Wolfgang, Marvin E., Robert M. Figlio, and Thorsten Sellin. 1972. *Delinquency in a birth cohort*. Chicago: University of Chicago Press.

Wolfgang, Marvin E., Terence P. Thornberry, and Robert M. Figlio. 1987. *From boy to man, from delinquency to crime*. Chicago: University of Chicago Press.

Measuring the Costs and Benefits of Crime and Justice

by Mark A. Cohen

Cost-effectiveness and benefit-cost analyses are tools that have been used by public policy analysts for years. Programs as diverse as environmental and land use regulations, welfare benefits, job training programs, and immunization policies have all been analyzed in this manner. Since the early 1980s, Federal regulatory agencies have been required to conduct benefit-cost analyses on major regulatory initiatives. Despite their widespread use, cost-effectiveness and benefit-cost analyses have not been staples of the criminal justice policy analyst's tool kit. This is rapidly changing in response to both increasing public demand for accountability of government agencies and the availability of new data and analysis techniques for identifying the costs of crime. This chapter reviews state-of-the-art techniques for estimating the costs and benefits of criminal justice and prevention programs. Although official government estimates of the cost of street crime have been available for many years, recent studies have attempted to go beyond government statistics by incorporating the monetary value of pain, suffering, and lost quality of life. Many of these studies use methodologies that are employed by environmental, health, and safety economists. Because these methodologies are new to the criminal justice research community, considerable attention is given to understanding their underlying assumptions, limitations, and alternatives. Cost-benefit analysis has arrived in the criminal justice policy arena, and it will not go away. Forcing analysts to quantify

Mark A. Cohen is an Associate Professor of Management at Owen Graduate School of Management, Vanderbilt University.

A
B
S
T
R
A
C
T

263

expected costs and benefits sheds new light on the merits of alternative programs and will undoubtedly change the focus of the debate in many criminal justice program areas. Whereas one could previously claim that “prevention is cheaper than prison” or “three strikes and you’re out pays for itself,” the benefit-cost framework allows decisionmakers to examine these claims more carefully and begin to make more rational, scientifically based judgments.

**A
B
S
T
R
A
C
T**

Benefit-cost analysis is a tool that has been in use by public policy analysts for many years. Its origins have been traced back to economists in the 19th century, and its use has been documented in the United States as early as the 1940s in evaluating alternative river development projects.¹ Programs as diverse as environmental and land use regulations, welfare benefits, job training programs, and immunization policies have all been analyzed in this manner. Schools of public policy and departments of economics teach courses devoted solely to the intricacies of benefit-cost analysis.

Since the early 1980s, Federal government regulatory agencies have been required to conduct benefit-cost analyses on major regulatory initiatives. These requirements have been adopted through Executive order and implemented by the Office of Management and Budget.² Recent proposals in Congress would legislatively mandate similar requirements.³ Thus, benefit-cost analysis has become a routine tool in the development of environmental, health, and safety regulations.

Cost-effectiveness analysis is a somewhat less ambitious but equally important policy tool. Unlike benefit-cost analysis, which requires all benefits and costs to be expressed in monetary terms, cost-effectiveness only requires that costs be monetized. Benefits still need to be expressed in some common denominator—such as comparable crimes, comparable injuries, lost years of life, and so forth. For example, one might compare the cost-per-life-year-saved of two different programs. Both of these methods require rigorous estimates of effectiveness—e.g., the number of crimes estimated to be averted by the policy under consideration. Such estimates are standard currency in most regulatory programs. For example, the Consumer Product Safety Commission estimates the number of injuries avoided by a recall, and the Environmental Protection Agency estimates the number of cancer cases averted. To date, criminal justice agencies have rarely been held to this high a standard. Researchers in sociology, psychology, and criminology are often content to show correlations or significance levels and seldom quantify “effectiveness” with measures such as the percentage reduction in recidivism or frequency of psychological trauma.

Because the academic literature now contains methodologies for doing benefit-cost analysis in the criminal justice arena, and there is growing literature applying benefit-cost analysis, the next generation of criminal justice students will soon be learning about these tools in courses on criminology and criminal justice policy.

When used properly, cost-effectiveness and benefit-cost analyses can be valuable tools that help inform the public policy debate. However, when used improperly, they can become nothing but rhetorical ammunition in an ideological debate.

Despite its widespread use elsewhere, cost-effectiveness and benefit-cost analysis have not been staples of the criminal justice policy analyst's tool kit. This is rapidly changing in response to both increasing public demand for accountability of government agencies and the availability of new data and analysis techniques for identifying the costs and benefits of criminal justice policies. Criminal justice researchers and policymakers will increasingly be confronted with cost-effectiveness and benefit-cost analyses—whether they like it or not. Ultimately, benefit-cost analysis might be required for newly proposed criminal justice policies. Because the academic literature now contains methodologies for doing benefit-cost analysis in the criminal justice arena, and there is growing literature applying benefit-cost analysis, the next generation of criminal justice students will soon be learning about these tools in courses on criminology and criminal justice policy.

Benefit-cost analysis is an art that is built on many important assumptions. It is important to understand some of these assumptions before attempting to either conduct such a study or to interpret a study that has been done by others. The purpose of this chapter is to assist researchers and policymakers in understanding the value and pitfalls of both cost-effectiveness and benefit-cost analyses. It also provides a vision for future research in this area by highlighting the key issues that have yet to be resolved. When used properly, cost-effectiveness and benefit-cost analyses can be valuable tools that help inform the public policy debate. However, when used improperly, they can become nothing but rhetorical ammunition in an ideological debate. My goal in writing this chapter is to promote the former while discouraging the latter.

The chapter is organized as follows: The first section discusses the conceptual underpinnings of this line of research, asking questions such as: Why put dollar values on crime? Whose costs and whose benefits are relevant? What criticisms have been offered against the economic approach to measuring the impact of crime and the use of benefit-cost analysis? The second section reviews alternative methodologies to measure the costs of crime and society's response to crime. The third section reviews the existing empirical literature on estimating the cost of crime and criminal justice programs, and the fourth section reviews the application of cost-effectiveness and benefit-cost analysis to

criminal justice, crime prevention, and offender treatment programs. The fifth section considers issues related to implementation that must be considered before using any cost or benefit estimates for policy analysis. A final section contains concluding remarks.

Conceptual Issues in Costs and Benefits of Criminal Justice Policy

Despite the fact that benefit-cost analysis has been used for many years, there are still many unresolved issues—both in terms of theory and application. This section reviews the most important theoretical issues raised concerning the methodology. I first consider why measurement of the costs and benefits of crime and criminal justice policies is worthwhile. Second, I discuss the difficulty of defining social costs and introduce the notion of external costs imposed by crime. Third, I examine *whose* costs and benefits should be considered in conducting benefit-cost analysis of a criminal justice policy. Fourth, I discuss the differences among average, marginal, and aggregate costs of crime. Finally, I consider some of the criticisms that have been articulated against the use of this methodology, both in general and in the context of criminal justice programs.

Criminal justice policy decisions always involve choices between two or more alternatives, each having its own costs and benefits. The enumeration of those costs and benefits puts the various alternatives on a level playing field and can help policymakers make more informed decisions that enhance society's well-being.

Why should we measure monetary costs and benefits?

The idea of measuring the monetary costs of crime and the monetary benefits of crime reductions has been around for many years. Gray (1979) reviews the history of the cost of crime and reports that one of the earliest estimates was published in a government-sponsored study in 1901. Subsequent Presidential commissions have been called on to report on the cost of crime. Many of these reports noted the difficulty and lack of progress over the years in adequately capturing the full costs of crime. However, they also acknowledged the importance of continuing this line of research.

To most economists, there is no question that crime costs should be estimated. Economics involves the allocation of scarce resources in society. Criminal

justice policy decisions always involve choices between two or more alternatives, each having its own costs and benefits. The enumeration of those costs and benefits puts the various alternatives on a level playing field and can help policymakers make more informed decisions that enhance society's well-being. Of course, if the enumerated costs and benefits are inaccurate, there is a risk that more information can lead to worse decisions. Further, many noneconomists would argue that there is neither a moral justification nor adequate empirical basis for placing dollar values on intangible factors such as pain, suffering, and lost quality of life. I will return to these issues later in this section. For now, I assume that such intangibles can be measured and consider three important policy-relevant purposes of measuring costs and benefits:

- Comparison of the relative harm caused by type of crime.
- Comparison of the aggregate harm from crime with that of other social ills.
- Benefit-cost analysis of alternative crime control policies.

Martin and Bradley (1964) provide a more detailed discussion of the importance of identifying and quantifying the costs of crime.

Relative harm by type of crime

Policymakers are often interested in comparing the harm caused by different types of crime. For example, most advocates of sentencing guidelines rely on victim harm as one component of their sentencing structure. Those who subscribe to a "just deserts" philosophy combine harm with culpability, whereas those who advocate a utilitarian approach combine harm with detectability and deterrability. Although one can tally up the various harms associated with each type of crime (e.g., value of property stolen, physical injuries, mental health-related injuries), without a common metric such as dollars, it is difficult to objectively compare these harms.

A few nonmonetary metrics have been proposed for comparing harms, such as the number of days for a victim to recoup from the financial loss or the number of life-years lost (see, e.g., Maltz 1975). These are primarily designed to overcome the perceived unfairness of valuing harms according to the wealth of the individual being harmed.⁴ However, these proposals also suffer from not having a common metric. One is still unable to compare 10 lost workdays with 10 lost life years.

Without a common metric to compare harms, the generally accepted approach to ranking the severity of crimes has been to survey the public (see Wolfgang et al. 1985; Cullen, Link, and Polanzi 1982; Rossi et al. 1974; and Rossi and Berk

1997). These surveys ask respondents to rank the seriousness of various crimes and result in relatively consistent rankings over time and across populations. However, they are based on subjective public perceptions concerning the severity of crimes—which may include misperceptions about the frequency of injuries in typical criminal events. For example, Cohen (1988b) argues that public perception surveys tend to underestimate the harm associated with violent crimes relative to property crimes. These studies are also generally unable to distinguish between the generic harm associated with an injury and the actual consequences of any particular victimization. This would be particularly important if one were interested in the extent to which the consequences of victimization vary across different segments of the population (e.g., age or sex), for example. Thus, although public perception surveys are useful for determining the public's attitudes toward crime, the surveys are limited in their ability to objectively measure and compare the seriousness of crimes.

Aggregate costs and benefits

One of the most common—yet probably least important—reasons for estimating the costs of crime is to tally up the aggregate cost to society. Multibillion dollar cost estimates can easily make their way into the popular press and political debate. There are two basic problems with tallying up the costs of intentional injury. First, having been told that crime costs the United States \$450 billion per year, what are we to do with this information? If we are successful in fully estimating the aggregate cost of crime, we can compare this total cost estimate with that of other social problems (e.g., cancer, auto crashes, homelessness). Whether one agrees that this is a useful exercise or not, various advocacy groups do compare “cost of crime” estimates with the cost of other social ills in an effort to affect policy decisions. Unfortunately, misuses of these data occur on both sides of the political debate.

Until recently, most estimates of the cost of crime (including estimates published by the Bureau of Justice Statistics) have significantly underestimated costs. For example, Irwin and Austin (1994) use the “official” estimate of \$19 billion to illustrate that crime is less of a problem than other social ills and to argue against increased prison sentences. A more comprehensive cost study sponsored by the National Institute of Justice (NIJ) reports that the annual cost of crime to victims is \$450 billion (Miller, Cohen, and Wiersema 1996). A *New York Times* article about that study quoted a Republican Congressman as saying the report “demonstrates that the cost of building prisons and adding police are justified” (Butterfield 1996). Despite the rhetoric, neither small nor large cost-of-crime numbers demonstrate that the cost of building more prisons is justified or that alternatives to incarceration are better than more prisons!

Even if properly measured,⁵ one cannot simply compare aggregate cost estimates of crime with estimates of the cost of other social ills and arrive at policy recommendations for future public spending priorities. Suppose, for example, that the cost of crime in the United States was estimated to exceed the cost of auto crashes. This does not necessarily mean that society should increase expenditures on crime prevention relative to the prevention of auto crashes. If the costs of preventing crimes and auto crashes are factored into the equation, it might be found that society is already spending too much on the former and not enough on the latter. The more relevant question is how much additional reduction in crimes (or auto crashes) would we observe if we spent more on prevention. This can be answered only if we know such things as the deterrent and incapacitative effect of various sanctions, increased police patrols, and so forth.

Subject to the previously mentioned caveat, comparing cost estimates of crime with other social ills can provide a basis of comparison on a common metric. For example, a study by Streff and colleagues (1992) estimated that the total cost of traffic crashes in Michigan was approximately three times the total cost of crime in that State. Although no immediate policy implications should be drawn from this comparison, it does help begin the process of identifying public policy priorities, and it puts crime in proper perspective. Over time, it might also be possible to quantify the magnitude of any change in crime rates by comparing costs per year.

A second problem with tallying up the costs of crime is that the true cost of crime is more than the sum total of its parts. If there were no more robbers or rapists, hitchhiking probably would be a way of life for a huge portion of the population. If violence was eliminated from society, organized crime might evaporate (as it depends on the threat of violence for its survival), and the standard of living for many inner-city residents would increase as businesses returned to previously abandoned storefronts. These massive changes in social structure could come about only with equally impressive changes in social behavior. Thus, any aggregate estimates of the cost of crime would need to account for these factors.

Cost-effectiveness and benefit-cost analysis of crime control policies

Perhaps the most important and controversial use of monetary estimates of the cost of crime is to compare the benefits and costs of alternative crime control policies. There is no shortage of crime prevention and crime reduction programs and proposals that would benefit from government funding. However, the government can only fund so many of these programs. One of the benefits

of using dollars as a common metric for analyzing criminal justice policy is that society spends dollars to try to prevent crime from occurring. Society's ability to control criminal behavior and reduce the incidence of victimization is limited by its ability to pay for police, courts, corrections, and prevention programs. In an effort to reduce crime and the severity of its consequences, society has undertaken many criminal justice experiments, including intensive probation, electronic monitoring of offenders, shock incarceration, targeted and community policing, and spouse arrest programs for domestic violence. As new policies are tested and options considered, one must be able to apply objective evaluation techniques.⁶ If two options have identical crime control effects but differing costs, the choice is simple. Unfortunately, few policy alternatives are so easily compared. In a more realistic case where a new policy reduces crime at some additional expense (or increases crime at a savings), one of the key questions is whether the reduced (increased) crime is worth its cost. Only by monetizing the cost of criminal victimization can one begin to answer that question.

One of the most compelling reasons to monetize the costs and benefits of crime control programs—and to attempt a benefit-cost analysis—is the consequences of *not* doing so. Whenever a criminal justice or prevention program is adopted or not adopted, society is implicitly conducting such an analysis and placing dollar values on crimes. For example, suppose one program costs \$1 million and ultimately will prevent 100 burglaries from occurring. Whether explicitly stated or not, the policymaker adopting that program has determined it to be worth spending at least \$10,000 to prevent each burglary (\$1,000,000 divided by 100 burglaries). If another \$1 million program that was *not* funded would have prevented 50 serious physical assaults, the policymaker is implicitly determining that each assault is worth *less than* \$20,000 (\$1,000,000 divided by 50). Thus, even the policymaker who has ethical concerns about placing dollar values on crime and conducting benefit-cost analyses *implicitly* makes a value judgment about the monetary value of crime.

Social versus external costs of crime

One of the most confusing and misunderstood concepts in the cost-of-crime literature is the difference between “social costs” and “external costs.” Many authors ignore this distinction or sweep it under the rug, making it difficult for the reader to know how to compare different estimates. This is not surprising, because there is no real agreement on which crime costs are social costs. Neither is there full agreement on whether social costs should be the relevant criteria for assessing the monetary cost or seriousness of crime. Indeed, I argue that the relevant concept for analysis of crime control programs is external cost, not social cost.

An external cost is a cost imposed by one person onto another, where the latter person does not voluntarily accept the negative consequence (through monetary payments or otherwise). For example, the external costs associated with a mugging include stolen property, medical costs, lost wages, and pain and suffering endured by the victim. The victim neither asked for, nor voluntarily accepted, compensation for enduring these losses. Moreover, society has deemed that imposing these external costs are morally wrong and against the law.

The concepts of social costs and external costs are closely related but not identical. Social costs are costs that reduce the aggregate well-being of society. In this case, medical costs and lost wages are clearly social costs because they are resources that could have been spent elsewhere in the economy, providing a socially productive activity. Although pain and suffering costs are not actual commodities or services exchanged in the marketplace, individuals are willing to pay real dollars and expend real resources to avoid the pain, suffering, and lost quality of life associated with becoming a crime victim. Thus, to the extent that society cares about the well-being of crime victims, these costs should also be considered social costs of victimization.

The value of the stolen property is more problematic. Some economists have argued that stolen property is an external, but not technically a social, cost because the offender can enjoy the use of the property. For example, Cook (1983) argues that the relevant concept should be the social cost—which would exclude transfers of money or property. However, Cook notes that he “presumes that the criminal is properly viewed as a member of society” (p. 374). In contrast, Trumbull (1990) argues that those who violate the criminal law are not entitled to have their utility counted in the social welfare function, i.e., their gain or loss is to be ignored. McChesney (1993) argues that criminal behavior is akin to “rent seeking” that has no social value. This example highlights the fact that social cost is a normative concept based on a subjective evaluation of whether an activity is socially harmful.

Regardless of whether one considers stolen property a transfer, there are other social costs associated with theft. Consider the case of an auto theft where the auto is never recovered, but the thieves use the car for their own private benefit. Although technically a transfer, the fact that cars are stolen forces potential victims to buy security systems, park in secure lots, and take other preventive measures. If the car or some of its contents are fenced, resources devoted to the fencing operations are considered a social cost, as these resources are diverted from socially productive uses. Thus, the value of stolen property might be used as a proxy for these lost resources, and are thus a measure of social cost (Becker 1968, 171, note 3).⁷

Regardless of whether stolen property is considered a social cost, society has an interest in enforcing property rights and has determined it is a crime to steal. There will be less productive investment—and therefore less social wealth—in a society where property rights are not enforced. The value of the stolen car must certainly be considered a cost of crime. For that reason, many economists who study the cost of crime rely on an external cost approach, including all costs imposed by a criminal on external parties, whether or not they are technically considered social costs.

Even the external cost notion of crime has pitfalls, however. Consider the victimless crime of drug abuse, which does not by itself create an external cost if the user voluntarily purchases drugs and reaps the full benefits and costs associated with its use. Nevertheless, drug abuse imposes many external costs: drug users might be less productive in the workforce and might commit crimes to support their drug habits, dealers might forego socially productive work activities, and society might be burdened with additional medical costs in treating drug addicts.⁸ Some of these costs (such as crime committed to support a drug habit and medical costs associated with drug overdoses) are clearly external and/or social costs, irrespective of whether drug use is illegal. However, some costs are social costs only because society has deemed drug use to be illegal. For example, economists generally consider the foregone legitimate earnings of a person in the illegal drug trade to be a social cost due to the socially valuable resources that are wasted. However, because illegal drug sales are voluntary transactions between two parties, these resources would not be considered social costs if drugs were made legal.

Another complicating factor in conceptualizing social and external costs is the fact that many crimes are allegedly committed as a form of self-help because the perpetrator feels wronged by the victim (Black 1998). Examples of this might be collecting on a bad debt, an original owner who steals back his property, and assaults committed in response to violent behavior. Although motives such as revenge or “self-help” do not justify criminal activity, they do raise the question of who is being harmed and whether those harms are external or social costs that society wishes to prevent.

From whose perspective are these costs and benefits to be measured?

The benefits and costs of criminal justice policy accrue to different parties—taxpayers, crime victims, offenders, government agencies, and so forth. Exhibit 1 contains a comprehensive listing of the costs of crime and who bears those costs. Exhibit 2 contains a similar listing of the cost of society’s response to

Exhibit 1. Costs of crime

Cost category	Party who directly bears cost*
Direct property losses	
Losses not reimbursed by insurance	Victim
Losses reimbursed by insurance	Society
Administrative cost: Insurance reimbursement	Society
Medical and mental health care	
Charges not reimbursed by insurance	Victim
Charges reimbursed by insurance	Society
Administrative overhead of insurance	Society
Victim services	
Expenses charged to victim	Victim
Expenses paid by agency	Society
Temporary labor and training of replacements	Society
Lost workdays	
Lost wages for unpaid workdays lost	Victim
Lost productivity for paid workdays	Society
Lost schooldays	
Foregone wages due to lack of education	Victim
Foregone nonpecuniary benefits of education	Victim
Foregone social benefits due to lack of education	Society
Lost housework	Victim
Pain and suffering/quality of life	Victim
Loss of affection/enjoyment	Victim's family
Death	
Value of life	Victim
Funeral and burial expenses	Victim's family
Loss of affection/enjoyment	Victim's family
Psychological injury/treatment	Victim's family
Legal costs associated with tort claims	Victim or victim's family
Long-term consequences of victimization	Future victims and society

* Ignores any recovery from offenders through legal action.

Source: Cohen, Miller, and Rossman 1994.

Exhibit 2. Costs of society's response to crime

Cost category	Party who directly bears cost*
Precautionary expenditures/effort	Potential victim
Fear of crime	Potential victim
Criminal justice system	
Police and investigative costs	Society
Prosecutors	Society
Courts	Society
Legal fees	
public defenders	Society
private	Offenders
Incarceration	Society
Nonincarcerative sanctions	Society
Victim's time	Victim
Jury's and witness' time	Jury/witness
Victim services	
Victim service organizations	Society/volunteers
Victim compensation programs	Society
Victim's time	Victim
Other noncriminal programs	
Hotlines and public service announcements	Society/volunteers
Community treatment programs	Society
Private therapy/counseling	Society/offender
Neighborhood watch and community prevention programs	Volunteers
Incarcerated offender	
Lost wages	Offender/family
Lost tax revenue and productivity	Society
Value of lost freedom	Offender
Psychological cost to family	Family of offender
"Overdeterrence"	
Innocent individuals accused of offense	Innocent individuals
Restriction of legitimate activity	Innocent individuals/society
Cost of detection avoidance by offenders	Offender/society/victim
"Justice"	
Constitutional protections to avoid false accusations	Society
Increasing detection to avoid differential punishment	Society

* Ignores any recovery from offenders through legal action.
Source: Adapted from Cohen, Miller, and Rossman 1994.

crime and who bears those costs. The terms “costs” and “benefits” are interchangeable depending on whether one is examining the effect of crime or the effectiveness of a crime reduction program. In other words, the cost of a crime is the same as the benefit of a crime that was prevented.

Economists often use the analogy of a market to describe crime (Cook 1986). Potential criminals commit (supply) crime based on the expected cost and benefits. These costs and benefits are determined by the actions (demand) of potential victims (e.g., whether they buy security alarms or take other costly preventative measures) and by the criminal justice community (e.g., the probability of being punished and the expected punishment). Many of the actions by potential victims and the criminal justice community that attempt to reduce the costs of crime are described in exhibit 2. Note that increasing some costs in exhibit 2 are expected to decrease corresponding costs in exhibit 1. For example, imposing longer prison sentences or increasing police protection are expected to decrease the chance of victimization and hence lower victim costs.

Some of the most significant costs of crime are the pain, suffering, and lost quality of life suffered by victims. Economists have long noted that “psychic” benefits and costs are part of individual utility and, hence, social welfare. Individuals are willing to trade tangible goods and services in exchange for some of these psychic benefits. Thus, they represent real social costs and benefits. Similarly, individuals who suffer the pain, suffering, and lost quality of life from becoming a crime victim would be willing to pay real dollars to reduce those psychic costs.

Although I include the lost productivity of the offender who is incarcerated, noticeably missing is the lost quality of life for the offender while he is denied freedom. When an offender is locked up and unable to be gainfully employed, not only does the offender lose wages, but society loses the value of those work hours. Hence, the offender’s lost productivity is included. However, the offender’s pain, suffering, and lost quality of life while in prison is not considered either an external or a social cost of crime because the offender is the only one who suffers. Although the offender is part of society, the conventional approach ignores the purely private losses. Similarly, although exhibit 1 includes the value of stolen property as a cost of crime that is borne by the victim, it ignores the value of stolen property that is a benefit to the offender who now has the use of the property. As these two examples clearly illustrate, benefit-cost analysis is not a value-free concept but instead involves definitions and explicit boundaries to determine whose costs and benefits matter.

Even though we might be able to identify all external costs associated with crime, some philosophical disagreement may still arise over which costs to

consider. For example, society might determine that we do not care about the negative consequences of imprisonment (psychological trauma, lack of freedom, etc.). Not all would agree with this approach, however, as antiprison activists might care very much about the treatment of imprisoned offenders. They might also care about the monetary and psychic costs to the family of the offender. However, those costs are indeed both social costs and external costs to the extent the family of the offender did not participate in the crime. Thus, to the extent possible, benefit-cost analysis should take into account the negative consequences on the family of the imprisoned offender as well.

On a more practical level, differences over which costs and benefits to consider may have important implications for criminal justice policy. Individual government agencies might ignore consequences that do not come out of their budget and that are not part of their mandate. For example, a public health agency might not budget enough money for drug treatment programs for indigents because it is primarily concerned with the direct budgetary implications of its decisions. However, it is well known that drug addicts have a high propensity to commit property crimes, often to support their habits. Although crime reduction benefits are important social benefits, they might not enter into the decision calculus of the public health agency. Similarly, the criminal justice system might ignore some of the social benefits of drug rehabilitation programs, despite growing evidence that drug courts can be effective mechanisms for reducing both drug abuse and subsequent criminal activity designed to support a drug habit (Sherman et al. 1997, ch. 9).

The issue of whose perspective to consider when calculating costs and benefits is far more complex than suggested by this example. For example, some of the funds for a criminal justice program might come from a Federal Government grant designed to expand the use of such programs. This will reduce the perceived costs of the program to the local agency. The result may be an overinvestment in that type of program—even if it is not socially beneficial but borne primarily out of political concerns. This might be one explanation for why police departments continue to staff Drug Abuse Resistance Education (D.A.R.E.®) officers despite the fact that there is little evidence the program reduces drug abuse. Although there is considerable evidence that increasing the number of police officers on the street can reduce crime, a police chief may not consider reducing D.A.R.E.® staff in exchange for more officers on the street, as the cost to the local community of hiring an additional D.A.R.E.® officer is considerably less than putting an officer on the street.

Average costs, marginal costs, and aggregate costs

Oftentimes, a distinction must be made between the average cost and marginal cost of a program. Conceptually, all costs of a crime prevention program should be included in a cost analysis *if* those costs would not otherwise be incurred. In practice, this requires an understanding of *incremental* (or *marginal*) costs versus *fixed* costs. It also requires an understanding of *opportunity* costs.

Fixed costs do not vary with the number of participants in the program. Thus, the annualized cost associated with maintaining a criminal court (compensation for the judge, debt retirement on the building, etc.), might not be affected by the number of cases actually tried in any year. Other costs, such as a drug rehabilitation program or feeding an incarcerated offender, vary with the number of participants. These are considered incremental (or marginal) costs. Unless fixed costs change with a policy decision under review, they should be ignored for purposes of assessing that policy.

A cost that is incremental for one decision might be fixed for another. For example, consider the problem of whether to increase the average sentence for violent offenders. If prison capacity is not a binding constraint, the incremental costs are primarily the cost of food, medical care, and so forth for these offenders. However, if this policy will require additional prison capacity, then the annualized cost of prison cells is part of the incremental costs.

There is a fundamental economic principle at work here: only the costs (and benefits) that vary with the decision should be considered. This is a general rule that should be applied to virtually any policy decision. However, implementing this rule is often less than straightforward, and it requires a careful analysis of which costs and benefits vary with a decision. For example, suppose a local court is considering whether to establish a new drug rehabilitation program that will serve as an alternative to incarceration for first-time nonviolent offenders. In that case, the analyst might want to compare the *average* cost of drug treatment with the alternatives because both fixed and incremental costs will vary with the decision about whether to start this program. Suppose the alternative is incarceration for those offenders, and there is adequate prison space available. In that case, we would compare the *average* cost of the drug treatment program with the *incremental* cost of incarceration. Because there are plenty of prison beds available, the fixed costs associated with prison construction and maintenance do not vary with our decision about whether to institute this drug treatment program. Hence, the fixed costs of incarceration are irrelevant in this case. Of course, this may not be true if the alternative to a drug treatment program is to build a new prison. On the other hand, average costs in some instances are irrelevant. For example, suppose a drug rehabilitation program is already

operational, and there is excess physical capacity. In this case, the only relevant costs to be considered in the decision to expand the program are the *incremental* costs of adding program participants.

Note that when estimating the cost of incarceration (or savings due to less incarceration), whether the annualized cost of constructing a prison cell should be included depends on the capacity constraint. If there are empty beds, the opportunity cost of the prison cell is zero, and the only costs are food, electricity, medical care, and so forth for the additional prisoner. In that case, the cost of the prison space itself is a “sunk” cost that is not relevant to the costs and benefits of the proposal under consideration. The financing cost of the prison will be incurred regardless of whether another drug abuser is incarcerated; thus, there will be no savings from diverting the drug abuser to a nonincarcerative treatment program. On the other hand, if the prison is operating at full capacity and lack of prison space is forcing authorities to incarcerate fewer individuals than they otherwise would, the annualized cost of the prison cell might be considered an opportunity cost that is saved by diverting the offender. In that case, money is indeed being saved by not having to build another cell and/or by not enduring a higher crime rate due to the inability to incarcerate other offenders.

Finally, for many types of policies under consideration, there are a host of cost categories that can be ignored. These aggregate costs of crime include fear of crime, deterioration of the quality of neighborhoods, and so forth. Unless a policy will have a significant impact on these communitywide costs, they can be ignored for purposes of considering any one policy.

Critiques of benefit-cost analysis and monetizing crime costs

Although few would disagree with the fact that *enumerating* costs and benefits is a worthwhile exercise, there is less agreement on whether costs and benefits should be *measured* and, if so, how much weight benefit-cost analysis should be given in policy analysis. At one extreme, many economists would argue that virtually any cost and any benefit can be measured—albeit with some uncertainty and often using indirect methods. Some economists might even argue that benefit-cost analysis should be the primary criteria used in making policy decisions. At the other extreme, some authors argue that not only is it difficult or impossible to measure some costs and benefits but, even if we could, benefit-cost analysis is inappropriate for use in many policy discussions. Kelman (1981) articulates several concerns over the use of this methodology on ethical and philosophical grounds. He argues that some things simply cannot be valued, such as free speech, pollution, or safety.⁹ He also argues that benefit-cost

analysis assumes that economic efficiency is the goal, at the expense of other socially desirable goals such as equity or fairness. This is not a criticism of the methodology—only of those who want to impose benefit-cost analysis as the sole criteria for public decisionmaking. Instead, when viewed as one policy tool available to policymakers, benefit-cost analysis has many benefits and only limited costs. Indeed, most texts on benefit-cost analysis include an analysis of the “incidence” of costs and benefits—i.e., who bears the costs and who reaps the benefits—as an integral part of benefit-cost analysis.

In the context of crime, Zimring and Hawkins (1995) are highly critical of recent attempts to monetize the cost of crime. They argue that the state of the art in economics has not developed to the point where we can adequately characterize the social costs and benefits—either in theory or in practice. Thus, economists have problems both in defining the social cost of crime and in measuring it in any meaningful way. Although there is some validity to both concerns, there is also much confusion about the proper role that benefit-cost analysis can play in policy debates. I will defer a discussion of the empirical concerns raised by Zimring and Hawkins until the section “Review of Literature on the Costs of Crime and the Criminal Justice System,” where I discuss monetary estimates of the cost of crime. Here, I will address their theoretical concerns.

Zimring and Hawkins note that recent attempts to estimate the monetary costs of crime fail to articulate a coherent theory underlying its cost estimates. Those who attempt to estimate the cost of crime have perpetuated much of the confusion; indeed, my writings in this area are partly to blame by not thoroughly explaining the underlying theory. Part of the problem is a misunderstanding of the difference between social costs and external costs, a subject that was previously discussed at length. As an example, Zimring and Hawkins (1995, 141) cite the theft of a \$50,000 Mercedes whose owner failed to take relatively inexpensive antitheft precautions. Noting that this might be a \$50,000 personal loss to the owner, they wonder what the social cost is. As discussed earlier, although there might be some disagreement about whether the \$50,000 theft is technically a social cost, there is no doubt that it is an external cost that society has an interest in preventing. Because society has laws making it a crime to involuntarily appropriate the property of others, and the harm to the victim is clearly related to the value of the item stolen, \$50,000 is a good estimate of the external cost of the crime.

Next, Zimring and Hawkins raise the concern that “any public expenditure to prevent it up to \$49,999 would be justified on a cost-benefit basis.” On the contrary, I would *not* argue that society should spend up to \$49,999 to prevent this theft. Although a simple benefit-cost analysis comparing the theft with a proposal

requiring an expenditure of \$49,999 to prevent the theft would conclude that benefits exceed costs; if alternative measures could prevent the theft at a lower cost, those alternatives would be preferred, and *they* would be economically efficient. To spend \$49,999 to prevent a theft that could be prevented for \$200 is economically *inefficient*. This example has important practical policy implications. It is *not* appropriate to examine only one policy option. Instead, policy analysts should examine many alternatives to find the one that has the highest benefit-cost ratio, or the most “bang for the buck.” Indeed, regulatory agencies are often required by law to consider all technically feasible alternatives to proposed regulations.

The distinction between social and external costs is most apparent with victimless crimes such as drug abuse, prostitution, and gambling. Although economists are often chided for their arguments that these crimes impose no social costs and ought to be legalized, that is a simplistic view of the economic arguments. It is true that there is no direct social cost associated with many of these crimes because they are voluntarily supplied and demanded, and the individuals who consume these illegal products incur both the direct cost and benefit of these products. However, society *has* made them illegal for some reason—often because of the collateral consequences that are socially undesirable, including medical/health concerns, external costs imposed on children or other family members, and so forth. To the extent these *external* costs can be identified and measured, they should be included as the cost of victimless crimes.¹⁰

Methodologies for Measuring Costs of Crime and Society’s Response to Crime

There are many different approaches to measuring society’s response to and the costs of crime. Broadly, costs can be classified as either tangible or intangible, and measurement methods can be classified as either direct or indirect. Tangible costs are those that involve monetary payments such as medical costs, stolen or damaged property, wage losses, prison cells, and police expenditures. These are costs that end up being tallied in the gross national product and are normally included in estimates of aggregate or individual wealth. Intangible, or nonmonetary, costs are those not normally exchanged in private or public markets, such as fear, pain, suffering, and lost quality of life. There are many methods for estimating the costs of crime. Broadly, these methods can be described as either direct or indirect. Direct methods use primary sources such as crime victim surveys or budgets of criminal justice agencies. Indirect methods use secondary sources such as property values or jury awards. This section reviews the state of the art in identifying and measuring both the tangible and intangible costs of crime.

Tangible costs of crime

At first, it might appear that the tangible costs of crime are relatively straightforward to estimate. In fact, aside from data on direct government expenditures on the criminal justice system, this is far from the truth. For example, there is no national accounting system tallying up the out-of-pocket losses to crime victims. The only direct source of crime victim costs is the ongoing National Crime Victimization Survey (NCVS), which interviews households and elicits information from those who have experienced a recent criminal victimization (Rand 1998). NCVS asks crime victims several questions about their out-of-pocket losses, including an estimate of the dollar cost of medical care, lost wages, and property loss. These estimates are periodically published by the Bureau of Justice Statistics (see, e.g., Klaus 1994). Despite their official-looking stature, NCVS crime cost estimates severely underestimate the tangible costs of crime to victims. First, the reference period for NCVS is crimes committed during the previous 6 months. Because the average crime will have occurred about 3 months prior to being reported, any medical costs are necessarily limited to those short-term costs. Even short-term costs are likely to be underestimated, however, because hospital bills often are sent directly to insurance companies and may arrive months after hospitalization. Second, some cost categories are simply excluded from NCVS. For example, respondents are not asked about mental health care, despite the fact that this is a significant cost of victimization (Cohen and Miller 1998). In addition, the consequences of victimization can be far-reaching and beyond the scope of any government survey. According to Burt and Katz (1985, 330), "During the weeks or months following the (rape), women frequently make costly changes in their lifestyles; this may involve moving to a 'better' neighborhood, buying expensive security systems, or avoiding work situations which they suddenly perceive as dangerous." I am not aware of any study that attempts to quantify these losses. Long-term implications of victimization may also be hidden and underestimated. For example, a recent study by Macmillan (2000) finds that educational attainment and lifetime earnings are lower for victims of childhood physical or sexual assault. These impacts have yet to be incorporated into cost-of-crime estimates.

Because the direct method of estimation is known to exclude significant costs, recent attempts to estimate medical costs and lost wages have relied on indirect methods. Miller, Cohen, and Wiersema (1996) obtained all available data on each crime victim in NCVS (e.g., type of injury, whether hospitalized, age and sex of victim) and combined that information with cost-of-injury data from other sources, such as worker compensation and hospitalization charges. This resulted in estimates of tangible costs that are considerably higher than NCVS—about four times higher in the case of robbery, more than 10 times higher for assault, and 20 times higher for rape (Miller, Cohen, and Wiersema 1996, table 9).

Other tangible crime costs that are reasonably easy to measure include police expenditures and the cost of the criminal justice system itself. Although aggregate costs may be available from government statistics, the cost *per crime* is not always available. For some purposes, we might be interested in these costs. For example, in studying the costs and benefits of an early release program, we would want to know the cost of recidivism created by those who are released from prison early. Thus, we might want to know the marginal cost of police resources associated with investigating a crime, as well as the marginal costs to the criminal justice system of having to reprocess a repeat offender. Such studies exist for specific jurisdictions and/or specific time frames. However, these studies are not routinely updated and the costs might vary considerably by location.

Economic/white-collar crimes such as fraud, theft of services, and antitrust violations, are notoriously difficult to quantify because victims often do not know they have been subject to a criminal offense. Even for those crimes in which victims are aware of their losses, there is no central government survey or reporting mechanism to tally these crimes or their costs. Government regulatory or enforcement agencies often collect these figures and may report them as they see fit. However, it is often difficult to verify their methodology and to know if any figures can be compared in a meaningful way. Most estimates of the cost of economic crimes are based on either surveys of potential victims to ascertain their experiences or collection of government data on prosecutions.

Finally, even *potential* victims suffer tangible costs of crime by taking costly preventive measures, such as buying security systems, deadbolt locks, cell phones, guard dogs, and guns purchased for defensive protection. Although direct measures of these expenditures should be relatively easy to obtain through survey methods, one difficulty in doing so is the fact that many of these expenditures serve dual purposes. The guard dog may also be a pet that provides companionship. The cell phone might provide a sense of security to a nighttime traveler, in addition to its use for other purposes. Sorting out the reason for purchase and the value obtained for each reason is not a trivial task.

Intangible costs of crime

Victims, potential victims, and communities all incur intangible costs of crime. Crime victims incur pain, suffering, and lost quality of life following the physical injury and/or psychological trauma associated with victimization. Potential

Victims, potential victims, and communities all incur intangible costs of crime. Several approaches have been used to estimate the monetary value of these intangible costs.

victims might have increased fear, which manifests as psychological anxiety and/or actual averting behavior (e.g., staying home at night, walking longer distances to avoid certain streets). Communities and businesses might suffer from reduced tourism and retail sales as outsiders perceive the community to be a high crime area. High crime rates might also inhibit economic development as employers and potential employees shun certain communities.

Several approaches have been used to estimate the monetary value of these intangible costs. Perhaps the earliest indirect method was to infer property owners' willingness to pay for a safer neighborhood through higher property values. To the extent that risk of victimization is capitalized in housing prices, we expect higher crime neighborhoods to have lower housing prices, controlling for all other factors that affect house prices (Thaler 1978). The methodology requires detailed location-specific housing characteristics (square feet, number of rooms, age, etc.), housing prices, crime rates, and other location-specific amenities (e.g., tax rates, school quality, distance to center city). Multiple regression analysis isolates the effect of crime on housing prices. The coefficient on the crime variable is then interpreted as the marginal willingness to pay for a reduction in the crime rate. Note that this is a marginal valuation, based on the current crime rate and small changes around that rate.

Data limitations have prevented these property value studies from isolating the cost of any individual crime type. Instead, studies to date have estimated the cost of an aggregate measure of crime such as the Federal Bureau of Investigation's (FBI's) Uniform Crime Reporting Index. In theory, a comprehensive dataset could isolate the effect of each crime type on housing prices. Property value studies necessarily rely on important assumptions about the competitiveness of the housing market and consumer information about neighborhood crime rates. They also ignore the effect that location-specific amenities—including crime—have on local wage rates. A few researchers have estimated both a housing and a wage equation to capture both effects (see, e.g., Hoehn, Berger, and Blomquist 1987). Although these models use two equations, they have yet to estimate simultaneous models that account for the interaction between housing prices and wages.

One of the positive features of the property value studies of crime is that they rely on actual market transactions. Although economists tend to favor market-based approaches in which actual market transactions (housing prices) are used, any market-based approach necessarily takes into account the wealth and income of the buyer. Thus, the fact that less wealthy individuals necessarily buy less expensive homes leads to an estimate of the value of crime that is based on "ability to pay." This issue will reappear several times in this chapter because many of the methodologies discussed depend on the income of the victim or potential victim.

The housing market is not the only indicator affected by crime rates. People buy handguns and security alarms, take cabs instead of walk, and use other precautions to avoid crime. Although all of these expenditures can be considered part of the cost of society's response to crime, they might also be used in estimating the cost of crime itself. For example, a study of the purchase of security alarms might allow us to infer the value that consumers place on a particular reduction in the probability of being victimized. For example, if the purchase of a double-bolt lock at the cost of \$25 reduces the risk of being burglarized from 1 in 500 to 1 in 1,000, we could infer that the individual who purchases the lock values the reduced risk by at least that amount.

Another method of estimating the nonmonetary costs of crime is to infer society's willingness to pay for reductions in crime from noncrime studies of society's willingness to pay for safety. Although there are several approaches, this growing literature primarily estimates wage rate differentials for risky jobs (Viscusi 1993). Thus, for example, if there is an additional \$50 wage rate premium for accepting an increased risk of death of 1 in 500,000, that is interpreted to mean that the collective "value of life" is \$25 million ($\$50 \times 500,000$). There is now an extensive literature on the value of life, which should *not* be interpreted as the value of any one particular life but, instead, is society's value of saving a "statistical" life. The first attempt to incorporate these value-of-life estimates with crime appears to be by Phillips and Votey (1981), who combined value-of-life estimates and out-of-pocket costs of crime with society's perception of the seriousness of crime to arrive at crime-specific monetary estimates. However, their methodology was unable to account for the risk of injury and death for many crimes.

Cohen (1988a) attempted to overcome these data limitations by combining estimates of the value of life with monetary estimates of the pain, suffering, and lost quality of life for nonfatal injuries. The approach used in Cohen is a hybrid of direct and indirect cost estimation. Direct costs are taken from NCVS data and several additional sources to augment some of the weaknesses of the government survey. Nonmonetary costs include the value of life for fatal crimes and pain, suffering, and lost quality of life for nonfatal injuries. These nonmonetary costs are estimated using indirect techniques. Risk of death is calculated directly from FBI data identifying the underlying crime in homicide cases. Risk of death probabilities are multiplied by the value of life to arrive at an estimate of the value of the risk of death component of each crime type.

The innovative—and most controversial—methodology introduced by Cohen (1988a) was the use of jury award data to estimate the monetary value of pain, suffering, and lost quality of life for *nonfatal* injuries. Cohen used jury awards in traditional tort cases and matched the type and severity of injury (e.g., broken

bones) with crime victim data in NCVS. This approach implicitly assumes that identical injuries are valued the same whether caused by an auto accident or an assault. However, crime victims might endure more pain and suffering due to the psychological trauma and fear of repeat victimization. More recently, Miller, Cohen, and Wiersema (1996) obtained data on jury awards to victims of physical and sexual assault and estimated crime costs using these cases.¹¹ These data were previously unavailable because civil lawsuits by crime victims are a relatively new phenomenon that has only recently grown to the point where adequate data exist. These lawsuits are generally against third parties for inadequate security, such as a parking lot owner not providing adequate lighting or an apartment complex owner not adequately securing a building.

The reason the jury award approach is controversial is primarily the popular notion that jury awards in the United States are unpredictable and/or unreasonably high. Theoretically, juries are asked to make the victim "whole" by compensating the victim for all out-of-pocket losses plus pain, suffering, and lost quality of life. Punitive damages are meant to punish the tortfeasor (defendant), not to compensate the victim; hence, they are excluded from the pain, suffering, and lost quality of life estimates. Despite popular beliefs to the contrary, considerable evidence exists that jury awards are predictable in a large sample. Popular press articles and calls for tort reform often focus on the outliers and punitive damage awards. The more common cases, however, are quite predictable, and jury awards are being used as a measure of pain and suffering in other contexts, including government regulatory agencies (e.g., Consumer Product Safety Commission). Perhaps most compelling, however, is the fact that our society has placed its tort system in the hands of juries and has decided that these awards are "just compensation."

Despite my defense of the use of jury awards to measure victim compensation for nonmonetary harms, this approach is theoretically *not* the most appropriate one for purposes of estimating the willingness to pay to reduce the risk of crime. Jury awards are *ex post* compensation designed to make a person whole. For policy purposes, the more relevant question is the "willingness to pay" (WTP) to reduce crime, which is an *ex ante* concept. The property value studies described previously are *ex ante* WTP approaches because they are based on actual market transactions, taking into account the prospective risk of criminal victimization. As noted, to date, researchers have been able to value an Index crime using only this method. For various reasons, the WTP for reduced crime is likely to be lower than the amount juries would likely award as compensation for an injury after the fact (see Cohen, Miller, and Rossman 1994, 73–74).¹²

An alternative approach to estimating the *ex ante* WTP for reduced crime is to directly survey the public (i.e., potential victims). This approach, often called

“contingent valuation,” is a methodology developed in the environmental economics literature and has been used extensively to place dollar values on non-market goods such as improvements in air quality or endangered species. There have been literally hundreds of contingent valuation studies, meta-analyses and textbooks written on the subject.¹³ Although there is some disagreement on the reliability of these surveys, they are continually used in benefit-cost analysis and natural resource damages litigation and for other purposes. A distinguished panel of social scientists, chaired by two Nobel laureates in economics (Arrow et al. 1993) was commissioned by the National Oceanic and Atmospheric Administration (NOAA) to assess the contingent valuation methodology. This panel was brought together because NOAA had drafted regulations calling for the use of this methodology when estimating natural resource damages in legal proceedings involving compensation for damaged public property. The panel concluded that this is a valid approach and provided a set of guidelines for conducting a reliable contingent valuation survey. Thus, if done properly, contingent valuation surveys can be useful policy tools. Although being used in many different policy contexts, contingent valuation is only beginning to be employed in criminal justice research.¹⁴

Finally, economists often rely on indirect measurement techniques by appealing to the notions of *opportunity cost* and *revealed preference*. In some instances, this is as straightforward as identifying foregone productive opportunities, such as the time an offender spends in prison or the time a victim spends out of work while dealing with the criminal justice process. In other instances, the costs are subtler. If consumers are rational and maximize their own utility, we can learn many useful things from their behavior—i.e., their revealed preference for one choice over another. Thus, the fact that individuals choose a leisure activity over working another hour provides us with a lower bound estimate of the value of that leisure activity—it must be at least as much as the opportunity cost of the time involved. This notion can be used to value the cost of many preventive or avoidance activities that people take to reduce their likelihood of victimization. Examples of these time costs include the time people take to lock and unlock cars and homes and the time needed to take a longer route home to avoid a bad neighborhood.

Some crimes with very large intangible costs, like treason or crimes that betray the public trust, may never be monetized. However, that does not invalidate the theory that would identify the social cost of treason to be the risk of harm to our national security, or the social cost of a public betrayal of trust to be a diminution of public trust and moral behavior. Cohen (1989) examines the social costs associated with numerous types of corporate crime, including fraud, environmental, food and drug, safety, export violations, and other regulatory offenses. Although it is difficult to estimate social costs for some of these crimes, the

vast majority of corporate crimes are frauds that can be easily measured, and techniques exist for estimating the harm in many other instances.

Review of Literature on the Costs of Crime and the Criminal Justice System

This section reviews the empirical literature estimating the costs of crime and society's response to crime. The purpose of this review is to provide the most recent estimates available. However, because some older studies used different methodologies, they are included for comparison purposes. Most crime cost studies to date have focused on traditional Index crimes, with some recent attempts to estimate drunk driving and child abuse. Other studies have examined the cost of drug abuse, white-collar crimes, and maintaining the criminal justice system. Few studies have systematically examined all crime types and all costs.¹⁵ The reader should be careful about comparing and/or adding up the costs in these various studies. Differences in methodologies, time periods, and potentially overlapping cost categories make such comparisons difficult. A more thorough study is required to undertake such an endeavor.

Traditional Index crimes

Exhibit 3 contains the most recent estimates of the cost of crime to victims, using the approach originally developed by Cohen (1988a)—combining out-of-pocket losses, the risk of death as measured by the value of life, and jury awards for nonfatal injuries. These estimates are taken from Miller, Cohen, and Wiersema (1996), an NIJ-sponsored research project. Exhibit 3 provides estimates of the cost *per criminal victimization*, which range from \$370 for larceny to more than \$2.9 million for murder (1993 dollars). These figures include attempted crimes that are unsuccessful and are averaged over all crimes—whether or not injury occurs. They include the cost of victim services provided by government and nonprofit agencies and the initial emergency police response (but not followup expenses to catch the offender). They exclude the risk of death because crimes resulting in death are included as a separate crime category. The Miller, Cohen, and Wiersema (1996) estimates are more comprehensive than previous efforts because they include domestic assault, child abuse, and drunk driving.

The estimates of tangible victim costs in Miller, Cohen, and Wiersema (1996) are considerably greater than in comparable government estimates derived from victim surveys. For example, Klaus (1994) estimates the average cost per rape in NCVS is \$234. In contrast, Miller, Cohen, and Wiersema (1996) estimate tangible rape victim costs to be \$5,100, including \$2,200 in lost productivity and \$2,200 in mental health care. As noted earlier, NCVS respondents are

Exhibit 3. Losses per criminal victimization (including attempts)

	Productivity	Medical care/ambulance	Mental health care	Police/fire services	Social/victim services	Property loss/damage	Subtotal: Tangible losses	Quality of life	Total
Fatal crime									
Rape, assault, etc.	\$1,000,000	\$16,300	\$4,800	\$1,300	\$0	\$120	\$1,030,000	\$1,910,000	\$2,940,000
Arson death	724,000	17,600	4,800	1,900	0	21,600	770,000	1,970,000	2,740,000
DWI	1,150,000	18,300	4,800	740	0	9,700	1,180,000	1,995,000	3,180,000
Child abuse	2,200	430	2,500	29	1,800	10	7,000	52,400	60,000
Sexual abuse (including rape)	2,100	490	5,800	56	1,100	0	9,500	89,800	99,000
Physical abuse	3,400	790	2,700	20	2,100	26	9,000	57,500	67,000
Emotional abuse	900	0	2,700	20	2,100	0	5,700	21,100	27,000
Rape and sexual assault (excluding child abuse)	2,200	500	2,200	37	27	100	5,100	81,400	87,000
Other assault or attempt	950	425	76	60	16	26	1,550	7,800	9,400
NCVS with injury	3,100	1,470	97	84	46	39	4,800	19,300	24,000
Age 0-11 with injury	2,800	1,470	100	84	46	39	4,600	28,100	33,000
Non-NCVS domestic	760	310	81	0	0	39	1,200	10,000	11,000
No injury	70	0	65	69	9	15	200	1,700	2,000
Robbery or attempt	950	370	66	130	25	750	2,300	5,700	8,000
With injury	2,500	1,000	65	160	44	1,400	5,200	13,800	19,000
No injury	75	0	66	110	15	400	700	1,300	2,000
Drunk driving	2,800	1,400	82	40	*	1,600	6,000	11,900	18,000
With injury	12,100	6,400	82	120	*	3,600	22,300	48,400	71,000
No injury	170	0	82	17	0	1,000	1,300	1,400	2,700
Arson	1,750	1,100	18	1,000	*	15,500	19,500	18,000	37,500
With injury	15,400	10,000	24	1,000	*	22,400	49,000	153,000	202,000
No injury	8	0	18	1,000	0	14,600	16,000	500	16,500
Larceny or attempt	8	0	6	80	1	270	370	0	370
Burglary or attempt	12	0	5	130	5	970	1,100	300	1,400
Motor vehicle theft or attempt	45	0	5	140	0	3,300	3,500	300	3,800

* Unknown.

Note: All amounts are estimates in 1993 dollars. Totals may not add due to rounding. Risk of death is excluded. Source: Miller, Cohen, and Wiersma 1996, table 2.

asked only for short-term costs, and some categories (e.g., mental health) are excluded altogether. Measurement issues also arise in estimating NCVS results. For example, in some instances, NCVS survey respondents indicated they did incur costs but were unable to provide estimates. Klaus (1994) treats these responses as zero costs, whereas Miller, Cohen, and Wiersema (1996) assumed these individuals incurred costs that were comparable to victims with similar injuries.

Generally, the largest component of crime costs is quality-of-life or intangible costs. Because intangible costs also are subject to the most uncertainty (and controversy), they are reported separately. Intangible costs are estimated from the pain and suffering component of jury awards for physical and sexual assault as well as the intangible value of the risk of death. Thus, they are calculated for all crimes except larceny. Even burglary is estimated to have an intangible cost of \$300 per offense, based on the fact that a small fraction of burglaries eventually result in a homicide.¹⁶ The ratio of intangible to tangible costs varies considerably by crime, with burglary being on the low end—intangibles being about one-third of tangibles—and rape being at the high end, with intangibles being 15 times greater than tangible losses.

Exhibit 4 compares the cost per *victimization* with the cost per *victim*, highlighting another important measurement issue. Some crimes, particularly physical and sexual assaults, are often repeated against the same victim. Thus, measuring the cost of *victimizations* might understate the impact on *victims*. Although there are a significant number of series victimizations in NCVS, there are few studies of series victimization. The methodology distinguishing between victims and victimizations needs further development. For example, we do not know if a victim who is assaulted 10 times incurs higher or lower costs than 10 individuals who were victimized once. We also do not know much about the validity of these series victimization responses. Thus, Miller, Cohen, and Wiersema (1996) truncated the number of victimizations against 1 person in a 6-month period to 10. However, they conducted a sensitivity analysis varying the maximum number of victimizations per individual and found about an 18-percentage point spread in total crime costs, depending on which estimate is used.

The crime costs estimated in exhibit 4 can be compared with the estimates derived using the property value studies discussed earlier. In Cohen (1990), I compared several property value studies with my earlier cost-of-crime estimates (Cohen 1988a) and found both methods yielded relatively close estimates. After adjusting for the number of crimes reported, I estimated the value of an average property crime in Rochester, New York, to be \$665 based on Thaler (1978), compared with \$869 based on Cohen (1988a). The average Index crime was

Exhibit 4. Crime severity measured by monetized losses per crime victimization/victim

Crime	Per victimization		Per victim	
	w/o risk of death	w/risk of death	w/o risk of death	w/risk of death
Child abuse: Sexual	\$99,000	*	\$125,000	*
Rape and sexual assault (excluding child abuse)	\$87,000	\$87,000	\$109,000	\$110,000
Child abuse: Physical	67,000	*	77,000	*
Child abuse: All types	60,000	63,000	70,000	74,000
Arson	38,000	54,000	38,000	54,000
Child abuse: Emotional	27,000	*	30,000	*
Drunk driving	18,000	26,000	18,000	26,000
Assault or attempt (NCVS)	9,000	19,000	12,000	31,000
Assault (any)	9,000	15,000	14,000	23,000
Robbery or attempt	8,000	13,000	10,000	16,000
Motor vehicle theft	4,000	4,000	4,000	4,000
Burglary	1,400	1,500	1,600	1,700
Larceny	370	370	400	400

* Deaths due to child abuse are not categorized by type of child abuse (e.g., sexual, physical, or emotional). Thus, the estimates do not include the risk of death. However, a combined child abuse category is included in this table, which includes the risk-of-death estimate.

Note: All amounts are estimates in 1993 dollars. Assault, robbery, motor vehicle theft, burglary, and larceny include attempted crimes that were not successfully carried out. If the other crime categories excluded attempts, the arson and drunk driving categories might drop in the rankings. See text.

Source: Miller, Cohen, and Wiersema 1996.

estimated to range from \$1,177 based on a study in Boston (Hellman and Naroff 1979) to \$2,285 in Chicago (Rizzo 1979), compared with the national estimate of \$2,210 in Cohen. Because the estimates in exhibit 4 do not differ dramatically from those in Cohen (1988a), an updated comparison would likely find similar results.

Exhibit 5 aggregates victim crime costs based on the number of victimizations in the United States between 1987 and 1990, resulting in aggregate costs of \$450 billion in 1993 dollars. This estimate includes only the cost of crime to victims and the cost of services provided to victims of crime. It excludes the

Exhibit 5. Aggregate annual costs of criminal victimization

Crime	Tangible	Quality of life	Total
Fatal crime (1990)	\$33,000	\$60,000	\$93,000
Rape/robbery/abuse/neglect/assault	25,000	46,000	71,000
Arson death	600	1,700	2,000
Drunk driving death (DWI)	7,200	12,300	20,000
Child abuse	7,300	48,000	56,000
Rape	900	8,000	9,000
Sexual abuse	1,400	12,800	14,000
Physical abuse	3,200	20,400	24,000
Emotional abuse	1,900	7,100	9,000
Rape and sexual abuse	7,500	119,000	127,000
Other assault or attempt	15,000	77,000	93,000
NCVS with injury	11,000	44,900	56,000
Age 0-11 with injury	600	3,900	5,000
Non-NCVS domestic	2,200	19,100	21,000
No injury	1,300	9,500	11,000
Robbery or attempt	3,100	8,000	11,000
With injury	2,500	6,600	9,000
No injury	600	1,100	2,000
Drunk driving	13,400	27,000	41,000
With nonfatal injury	11,300	24,600	36,000
No injury	2,400	2,500	5,000
Arson	2,700	2,400	5,000
With nonfatal injury	750	2,400	3,000
No injury	1,900	65	2,000
Larceny or attempt	9,000	0	9,000
Burglary or attempt	7,000	1,800	9,000
Motor vehicle theft or attempt	6,300	500	7,000
Total	\$105,000	\$345,000	\$450,000

Note: All amounts are estimates in millions of 1993 dollars. Totals were computed before rounding. No-injury cases involve no physical injury, but may involve psychological injury. NCVS fatal crimes are all crime deaths except drunk driving and arson. Personal fraud/attempt is excluded to prevent possible double counting with larceny.

Source: Miller, Cohen, and Wiersema 1996, table 4.

cost of prevention and the cost of the criminal justice system. Of this amount, tangible costs are estimated to be \$105 billion, or about 25 percent of the total. The crime-specific estimates in exhibit 5 *exclude* the risk of death because a category already exists for fatal crimes. To include the risk of death in aggregate crime cost data would be double counting. Although it would be tempting to update this figure to 1999 dollars, this is not a straightforward exercise. Because crime has been steadily declining in the United States since 1990, updating national crime costs requires recent data on victimization rates. It also requires recent data on the distribution and severity of injuries to determine if this has changed significantly since the 1987–90 timeframe.

Drunk driving is a special category of crime that has some unique measurement issues. It is a crime (and a risk to society) every time someone drives drunk. Yet, many drunk driving incidents occur without any collisions and, thus, there is no harm to victims. In other crime categories, attempted offenses are included, as they might involve some property loss, fear, anxiety, and trauma. No comparable data exist on drunk driving incidents that do not result in collisions. In addition, not all collisions where the driver was drunk are caused by drunk driving. Some of those accidents might have occurred in the absence of alcohol. Thus, some method of attributing collisions to their cause is necessary.

Drug abuse

A series of reports has been commissioned by U.S. Government agencies to determine the economic costs of alcohol and drug abuse in the United States. The most recent study, by Harwood, Fountain, and Livermore (1998), estimated the total cost of drug abuse to be \$98 billion in 1992. The bulk of these costs (\$69 billion) were productivity losses to drug abusers, including premature death, reduced productivity while at work, career criminals who did not enter the legitimate labor market, and crime-related costs such as victim losses and time spent by incarcerated offenders. An estimated \$10 billion was spent on drug abuse services and health care for drug-related illnesses. The remaining \$18 billion was the estimated cost of crime committed by drug abusers.

Harwood, Fountain, and Livermore (1998) only included tangible costs and ignored intangible costs to victims, families of drug abusers, and so forth. Because a significant portion of these costs was associated with victims of crime, there is some overlap between these estimates and those reported by Miller, Cohen, and Wiersema (1996).

The Harwood, Fountain, and Livermore (1998) report illustrates the difficulty of preparing credible estimates of the cost of drug abuse. First, the empirical evidence on the causal connection between drug abuse and crime is limited and largely unresolved (Miczek et al. 1994). Thus, the authors necessarily rely on

Unlike street crime, which is systematically measured through victim surveys and by the FBI, no comprehensive surveys of the incidence or cost of white-collar crimes exist.

If the estimates are to be believed, white-collar crime causes tangible losses far in excess of tangible losses associated with street crimes.

assumptions that are based on a few limited studies. In addition, they assume that average productivity losses for incarcerated drug offenders is the same as the population average, about \$39,000 per year (see Cohen 1999). Yet, it is known that the typical incarcerated offender is not as productive as the average person (Cohen, Miller, and Rossman 1994), and those engaged in street-level drug dealing have been found to have relatively low legitimate wage earnings potential (Reuter, MacCoun, and Murphy 1990).

The actual cost of purchasing illegal drugs is not included in the Harwood, Fountain, and Livermore (1998) study. According to a study by Abt Associates (1995), approximately \$53 billion was spent on ille-

gal drugs in 1992.¹⁷ Heavy cocaine users are estimated to spend approximately \$9,000 to \$10,000 per year on cocaine, and heroin addicts spend approximately \$17,000 per year (Executive Office of the President, Office of National Drug Control Policy 1991). However, adding these costs would largely result in double counting. Drug users who buy drugs transfer wealth from themselves to the seller, a voluntary transaction not resulting in direct external costs. However, external and social costs do result from the *activities* surrounding the purchase and consumption of drugs (e.g., theft to support a drug habit, medical costs associated with drug-induced illness). Cohen (1998, 19) argues that one could use the cost of drugs as a proxy for the opportunity cost of resources devoted to drug distribution. However, there is a significant risk premium associated with selling drugs, which presumably is reflected in the price of drugs. Noting that the Reuter, MacCoun, and Murphy (1990) study of street-level drug dealers finds legitimate hourly earnings to be about 25 percent of hourly earnings

from drug sales, I assumed as a first approximation that only 25 percent of the price of drugs represents a social cost—the lost productivity of a drug dealer not working in legitimate activities. The remainder represents a risk premium paid to dealers who must face a higher risk of being killed on the job.

Economic/white-collar crimes

Unlike street crime, which is systematically measured through victim surveys and by the FBI, no comprehensive surveys of the incidence or cost of white-collar crimes exist. Although various estimates exist, the sampling methodology and crime definitions are

seldom transparent, making comparability across crime particularly difficult. If the estimates are to be believed, white-collar crime causes tangible losses far in excess of tangible losses associated with street crimes. For example, a 1995 study by the Association of Certified Fraud Examiners (1995) reports that the average business loses about six percent of its total annual revenue to fraud and abuse committed by its own employees. This translated into about \$435 billion in 1995—about four times the tangible losses from street crime shown in exhibit 5.

Exhibit 6 lists various estimates of the cost of economic/white-collar crimes. For example, Titus, Heinzelmann, and Boyle (1995) conducted a national survey of the U.S. population to identify victims of personal fraud. They estimated the annual tangible costs to be \$45 billion. However, some of the fraud definitions include incidents that may not be considered criminal. Noticeably missing from exhibit 6 are many regulatory offenses such as antitrust, environmental,

Exhibit 6. The cost of criminal fraud

Industry	Fraud type	Cost (\$ billions)	Year
All firms	Employee theft and fraud	\$435	1996 ^a
Telecommunications	Theft of services	\$3.7–\$5.0	1995 ^b
Health care	Overcharge, services not rendered, kickbacks, etc.	\$70	1992 ^c
Insurance	False claims	\$120	1995 ^d
Entertainment	Bootlegging	\$2.3	1995 ^e
Telemarketing	Con artists, sweepstakes, phone scams	up to \$40	1995 ^f
All consumers	Fraud in general	\$45	1991 ^g

^a Association of Certified Fraud Examiners 1995.

^b Communications Fraud Association. Private communication.

^c U.S. General Accounting Office. 1992. *Health insurance: Vulnerable payers lose billions to fraud and abuse*. Washington, D.C.

^d Insurance Information Institute. 1996. *Insurance issues update*. New York.

^e Recording Industry Association of America. 1996. Cited in Music and performer groups act to curb piracy. *Reuter European Business Report*. London, 26 September.

^f Federal Trade Commission. *1995–1996 report: Staff summary of Federal Trade Commission activities affecting older Americans*. Retrieved 30 April 2000 from the World Wide Web: <http://www.ftc.gov/os/1998/9803/aging98.rpt.htm>.

^g Titus, Heinzelmann, and Boyle 1995.

and food and drug. I am unaware of any studies that attempt to measure aggregate costs of these regulatory crimes. Based on the relatively large fines, cleanup costs, and so forth that can accrue in these cases, these regulatory crimes have the potential for enormous costs. However, some regulatory crimes are primarily reporting requirements that involve little harm (Cohen 1989).

To date, I am unaware of any study that attempts to quantify the intangible costs of fraud. In addition, the studies to date have assumed that the tangible losses are limited to the dollar value of the fraud. Nevertheless, there is anecdotal evidence that losses can be significantly greater in certain cases. For example, some frauds prey on the elderly and uneducated poor. To the extent these victims lose their homes, are unable to afford health care, and so forth, the costs may far exceed the dollar value of the fraud. Whether these losses are common or significant in the aggregate is unknown.

One conceptual difficulty in estimating the cost of crime against business is how to value the items taken. If money is taken, the value is straightforward—the face value of the bills. However, if the loss is merchandise, whether the loss should be valued at retail or wholesale depends on the opportunity cost to the victim. If the victim can easily replenish the product as needed and does not lose retail sales, the loss is the cost to the owner—not the price at which it would sell. However, if the item is scarce and cannot be readily replaced, the loss is now the full value the owner could have expected to receive for the item. Some white-collar crimes involve theft of services that involve essentially zero marginal costs to the victim and might not have been purchased at all in the absence of the theft. For example, the telecommunications industry estimates it is defrauded of \$3.7 to \$5 billion per year in schemes that allow users to obtain free services. This is a loss to only the phone company, however, if the user would have otherwise purchased the service. If these services would not have been purchased, it is hard to label this a cost. This is particularly true with bootlegged music and counterfeit luxury goods. Of course, in all cases, there may be other more subtle costs associated with the loss, such as diminishing the value of the legitimate product to all law-abiding purchasers.

Criminal justice system

The Bureau of Justice Statistics periodically estimates annual justice expenditures in the United States. These data are compiled from various surveys of local, State, and Federal government agencies. Lindgren (1997) reports that direct expenditures in 1992 by all governments was \$94 billion. Nearly half (\$41 billion) was for police, with the remaining for corrections (\$31 billion) and judicial and legal services (\$21 billion). These estimates include costs that are not related to crime, such as traffic safety and the civil court system. Some

of these expenditures are likely to be included in the estimates of the cost of drug abuse previously mentioned. In addition, they exclude crime prevention activities undertaken outside the traditional police, sheriff, or law enforcement office.

Even if accurate, aggregate criminal justice expenditures are of little value in conducting benefit-cost analysis. As noted, police officers do more than enforce criminal laws; they also deal with traffic safety and other community issues. Although I am unaware of any estimate that attempts to sort these costs, some authors have taken a different approach in estimating criminal justice costs using a “bottom-up” approach. Cohen, Miller, and Rossman (1994, 126–134) attempted to piece together the cost of the criminal justice system on a per-crime basis from a few studies done in single localities. They estimated the criminal justice processing cost per offense in 1987 dollars to be \$5,925 for murder, \$2,050 for rape, \$1,125 for robbery, and \$1,225 for aggravated assault. This includes the estimated cost of both police investigations and court-related costs such as pretrial booking and jail, hearings, and trials. I am unaware of any similar attempts to estimate per crime the criminal justice costs for other crimes. Note that the appropriate unit of analysis is an issue that becomes important in these types of studies. Criminal investigation costs occur only for crimes reported to police. However, because not all crimes result in arrests or convictions, few other criminal justice costs occur unless an actual offender is apprehended. Thus, for example, although the estimated cost of processing an aggravated assault case was estimated to be \$1,225, the cost per *victimization* was only \$580. However, because only 28.1 percent of aggravated assaults that were reported to police were cleared by arrest, the cost per *arrested* offender was considerably higher—about \$4,400 in 1987.¹⁸

Note that police (and firefighter) emergency response to victimization is excluded from this section because it was included in the cost of victimization itself. As shown in exhibit 3, these costs are relatively small—generally less than \$200 per incident. The emergency response to arson and murder are higher, about \$1,000 to \$2,000. All other criminal justice costs are considered part of society’s response to victimization (and therefore excluded from the cost estimates in exhibit 3).

Private crime prevention

Not all crime prevention programs are paid for by government agencies.

According to an industry estimate, revenues in the security industry were about \$82.2 billion in 1996—nearly as much as governments spent on the criminal justice system itself.¹⁹

Zedlewski (1985) estimated the cost of firearms and guard dogs bought primarily for protective purposes. More recent estimates by Laband and Sophocleus (1992) and Anderson (1999) estimate the costs of protective firearms, guard dogs, locks, and other protections. Anderson's (1999) is one of the few studies to attempt to estimate the opportunity cost of time spent preventing crime. He estimated the average adult spends 4 minutes per day locking and unlocking doors and looking for keys. Based on the opportunity cost of their time, Anderson estimated this to be worth \$437 per year per adult, or \$89.6 billion annually in the United States. He also estimated the value of time spent by participants in neighborhood watch programs to be \$655 million.

Although one could identify types of crimes that are more likely to be prevented by certain types of private protection expenditures, it is impossible to apportion most of these costs to individual crimes. For example, even though a private home alarm might be purchased to protect against burglary, it also protects a homeowner who happens to be home during a burglary from robbery, assault, rape, and murder. However, if one takes the view that the cost of burglary includes the *risk* of being home and further victimized, then expenditures on home security are indeed attributable to burglary.

Review of Literature on Costs and Benefits of Criminal Justice Policy

The existing literature on benefits and costs of criminal justice policies generally takes one of two forms: cost-effectiveness or benefit-cost studies. A cost-effectiveness study seeks to answer questions such as "What is the cost per crime averted?" or "What is the cost per successfully treated offender who does not recidivate?" These questions require a thorough understanding of costs and the probability of a successful outcome. They do not, however, require the analyst to monetize the successful outcome. To do so would be to conduct a benefit-cost analysis. To date, there have been few published studies of criminal justice or prevention programs that attempt to either conduct cost-effectiveness or benefit-cost analysis. Instead, researchers generally stop at the question of whether a certain punishment deters potential offenders or whether a treatment program reduces recidivism. If so, the program "works." But at what cost? Are there alternative programs that would give us more bang for the buck? The fact that few studies have been conducted is not surprising, given that criminal justice researchers are seldom economists and not necessarily trained to analyze costs and benefits. The few existing studies are described as follows.

Cost-effectiveness studies

Greenwood and colleagues (1994) compared various incarceration alternatives that were considered during the three-strikes debate in California. In comparing five alternatives, the cost per serious crime prevented ranged from \$11,800 (third violent offense) to \$16,300 (third felony offense). Although focusing on only violent offenders would appear to give us the most bang for the buck, because we do not know the value of the crimes averted by each option, we do not know which is better.

Similarly, Greenwood and colleagues (1996) compared four child and youth intervention programs: home visits to new mothers and daycare for their children, parent training, high school graduation incentives, and delinquent supervision. They found that per \$1 million spent, graduation incentives prevented the largest number of serious crimes (258), followed by parent training (157), delinquent supervision (72), and home visit/daycare (11). Although one might begin to prioritize spending on the basis of such a study, it does not tell us if *all* or only some of these programs should be adopted.

Thus, although the studies by Greenwood and colleagues can help us determine which approach is most cost effective, they are not equipped to determine whether any one approach is socially desirable. A policy analyst must make a subjective determination that the option being considered is worthwhile. That does not mean, however, that cost-effectiveness studies are without merit. They may also provide important information about the relative benefits of two or more programs being compared. For example, Rydell and Everingham (1994) compared supply-control drug strategies (e.g., drug seizures) with demand-control strategies (e.g., drug treatment). Although comparing completely different programs, Rydell and Everingham were able to place these two approaches on equal footing by estimating the reduced cocaine consumption from each alternative. They found a 1-percent reduction in cocaine consumption could be obtained by spending either \$34 million on treatment or \$246 million on domestic drug enforcement (Rydell and Everingham 1994, 24, table 3.2). This study is often cited as providing evidence that treatment is seven times more cost-effective than drug control programs.

It is worth noting that some analysts ask questions such as, "What is the cost per offender?" Although important, this is purely a cost analysis and should not be construed as a cost-effectiveness study. To be a cost-effectiveness study, one must measure outcomes (e.g., crimes averted, recidivism rate), not just inputs (e.g., number of offenders admitted to the program). More importantly, to focus solely on costs can easily result in a conclusion to fund a program even if it has few benefits.

Outside of crime, a myriad of government programs reduce the risk of death from consumer and motor vehicle products, highway and workplace accidents, medical care, and so forth. Tengs and colleagues (1995) calculated the cost per life-year saved by more than 500 such programs. Although the median program cost \$42,000 per life-year saved, they reported a wide range of interventions, from the best, which save more resources than they cost, to the worst, which cost as much as \$10 billion per life-year. Thus, shifting resources between programs could save a greater number of life-years at a lower cost.

Given the growing interest in quantifying the effectiveness of criminal justice programs (e.g., Sherman et al. 1997), a similar project would be extremely valuable in the criminal justice arena. Such a project would compare criminal justice and prevention programs among themselves as well as across other interventions that save lives. This is a major endeavor, however, that will not be completed in the immediate future. Many issues have yet to be resolved, not the least of which is determining a common metric for comparison purposes. Although Tengs and colleagues (1995) used life-years saved, this assumes 1 year of life is valued equally, regardless of age, health status, or wealth. An alternative that has been developed in the health economics literature, quality adjusted life-year (QALY), weights life-years by the level of pain and/or impairment (see, e.g., Fabian 1994). Another alternative is to use the common metric of dollars, which puts us into the realm of benefit-cost analysis.

Benefit-cost studies

To date, only a few researchers have gone beyond cost-effectiveness analysis to explicitly compare the monetary costs with the monetary value of benefits. This is not surprising, given that, until recently, no credible monetary estimates existed for intangible costs of crime. Moreover, introducing a new metric such as intangible costs takes a considerable amount of time before it becomes mainstream in the literature. In addition, as noted earlier, this approach is not without controversy.

Among the authors who have used monetary estimates of the cost of crime (including intangible costs) in conducting cost-benefit analyses are: DiIulio and Piehl (1991), Gray (1994), Levitt (1996), and Donohue and Siegelman (1998). Among the programs studied by these authors are longer prison sentences, prison overcrowding, rehabilitation programs, and juvenile intervention programs. Welsh and Farrington (2000) summarized recent studies that measure costs and benefits and also discussed some of the methodological issues surrounding benefit-cost analysis.

In cases where a program passes a benefit-cost test using only tangible costs, the need for monetizing intangible losses is less obvious. For example, Prentky and Burgess (1990) show that the cost of incarcerated sex offender treatment is less than the tangible benefits from lower recidivism rates (e.g., lower reprocessing costs of recidivists and lower victim costs). No intangible benefits need to be estimated because the program already passes a benefit-cost test. However, a similar study by Austin (1986) of early release programs in Illinois concluded that the benefits (reduced prison costs) exceeded costs (tangible costs associated with increased crime due to recidivists). As I show in Cohen (1988a), if the cost of recidivism *includes* the intangible cost of crime to victims, the benefit-cost ratio goes the other way, and Illinois residents are better off building more prisons or finding another less costly but equally effective alternative.

Recently, Washington State has undertaken the most ambitious and policy relevant project to systematically compare the costs and benefits of crime prevention programs. The Washington State Institute for Public Policy was mandated by the State legislature to “evaluate the costs and benefits of certain criminal justice policies, violence prevention programs, and other efforts to decrease the criminal recidivism of juvenile and adult offenders, and certain at-risk behaviors of youth” (Aos et al. 1999). The project involves monetarily quantifying all tangible costs and benefits. It presents the benefit-cost analysis from two perspectives—taxpayers and crime victims. To the extent possible, it uses Washington State estimates (e.g., cost of criminal justice resources), although crime costs are taken from the national estimates in Miller, Cohen, and Wiersema (1996). Only tangible, crime-related benefits are considered.²⁰ The study focuses on prevention programs and does not include policing and sentencing policies.

Despite its limitations, the policy implications of the Aos and colleagues (1999) study are profound. First, a few programs (such as juvenile boot camps) simply do not reduce crime at any cost and instead actually increase crime. Second, although most programs result in modest crime reduction benefits, benefits generally exceeded costs. Third, the largest benefit-cost ratio was generally found in programs targeting juvenile offenders. For example, an aggression-replacement training program was estimated to pay back taxpayers \$19.57 for every \$1 spent. When victim benefits are included, these programs were estimated to pay back \$31.40 per \$1 spent. Inprison education and vocational training programs for adult offenders also generally had a positive benefit-cost ratio. Some early childhood education programs, such as nurse home visitation and early childhood preschool programs, were also found to be cost-beneficial. Others were found to have benefits that were less than the cost to taxpayers. Thus, shifting government resources between such programs could have major long-term social benefits.

An interesting study by Ayres and Levitt (1998) highlights the fact that the costs and benefits of crime prevention programs may accrue to different parties. They studied LoJack, a radio-transmitter device used for retrieving stolen vehicles that costs \$600 to install (or about \$97 per year on an annualized basis). Ayres and Levitt (1998, 45–46) estimated that “one auto theft is eliminated annually for every three LoJacks installed in high-crime central cities. There is little evidence that the reductions in central city auto thefts are simply being displaced either geographically or to other categories of crime.” Interestingly, most of this benefit is a positive externality, because the probability that any one LoJack owner will reap the benefits of an avoided theft is relatively low. They estimate the private benefits to be about \$150 per year in expected theft protection. Since much of this benefit goes directly to insurance companies, their study raises the issue of how to design proper incentives—either through insurance markets or government programs—to capture these positive externalities for the full benefit of society.

Measuring the effectiveness of crime prevention programs provides useful information that goes beyond the criminal justice community. For example, the often-cited Perry Preschool program has been shown to have long-term crime reduction benefits that exceed its costs, in addition to the intended consequences of higher graduation and employment rates (Barnett 1993). Similarly, a study of the costs and benefits of drug treatment programs by Rajkumar and French (1997) finds that a substantial benefit beyond reduced drug consumption is the monetary value of reduced crime committed by rehabilitated drug abusers. Based on the Treatment Outcome Prospective Study of 11,750 drug abusers, they estimated that the monetary value of reduced crime 1 year following treatment far outweighed the cost of the program.

Like any statistical tool, benefit-cost analysis is vulnerable to misapplication through carelessness, inexperience, or deception.

A somewhat different approach was taken by Cohen (1998), in which the generic question was asked, “What is the monetary value of saving a high-risk youth” from a life of crime, drug abuse, or dropping out of high school? I estimated the value of saving one high-risk youth from a life of crime to be \$1.3 to \$1.5 million in 1997 dollars (discounted to present value). Comparable estimates are \$370,000 to \$970,000 for a heavy drug abuser, and \$243,000 to \$388,000 for a high school dropout. These estimates provide a basis for others who want to conduct a benefit-cost analysis of their programs.

Issues in Implementation and Policy Analysis

This section briefly addresses a few important issues that need to be considered by any researcher or policy analyst seriously thinking about using a cost-effectiveness or benefit-cost methodology. Briefly, I look at the potential for misuse of benefit-cost analysis, uncertainty, treatment of future benefits or costs, issues of fairness and equity, and public perception of the risk of crime.

The use and misuse of benefit-cost analysis

Policymakers—the consumers of benefit-cost analyses—often have little understanding of the methodology and assumptions underlying the analysis. Like any statistical tool, benefit-cost analysis is vulnerable to misapplication through carelessness, inexperience, or deception. The technique is sometimes criticized because it presents an aura of precision and objectivity that might not be justified. The results can be no more precise than the assumptions and valuations that are employed. Thus, it is important that the analyst carefully spell out the assumptions, the basis for those assumptions, the projected benefits, how those benefits are valued, and how alternative assumptions might affect the results (see the following discussion of uncertainty).

The risk of using benefit-cost analysis is that regulatory agencies or proponents of a particular program might use the technique to justify a program they want funded and manipulate the numbers until a positive benefit-cost ratio is achieved. Opponents of a proposal can do the same. Yet, benefit-cost analysis forces analysts to explicitly characterize the assumptions so that the analysis is transparent. This lends itself to an open process where the issues can be debated on an informed basis.

Uncertainty

Despite a policy analyst's best attempt to base the analysis on sound assumptions, there will always be considerable uncertainty about both the costs and benefits of a proposed program. Oftentimes, the program being evaluated for potential implementation will be based on one already studied elsewhere. Yet, the two programs will rarely be identical to each other. Differing elements

Despite a policy analyst's best attempt to base the analysis on sound assumptions, there will always be considerable uncertainty about both the costs and benefits of a proposed program.

might include the demographics of the offending or treatment populations, the punishment or treatment protocols (e.g., level of security for a prison; type, length of, and number of counseling sessions), the program personnel (e.g., educational background, experience, commitment), and the time periods. Any one of these factors might alter both the costs and effectiveness of the proposed program.

One method for dealing with uncertainty is to conduct a sensitivity analysis of the results. For example, suppose previous studies found that a drug treatment policy is 20 percent effective at reducing crime committed by those who complete the program and that the benefits of the program exceed its costs. Instead of assuming a 20-percent effectiveness rate, one might vary this assumption to see how sensitive the benefit-cost analysis is to that rate. We would be much more confident in the program if it also passes a benefit-cost test with only a 10-percent success rate. However, we would be much less sanguine if only a slight reduction in effectiveness reversed the benefit-cost equation. Depending on the number of studies and type of data available, a more sophisticated statistical analysis might be performed to estimate a confidence interval around the costs and benefits.

Discounting to present value

Another problem of comparability often arises when the benefits of a program will not be realized for many years into the future. For example, investing in a program that treats young offenders involves expenditures today, but might yield benefits over an extended future timespan. Because a dollar spent today is not the same as a dollar received 15 years from now, future benefits must be discounted to present value when compared with the costs borne today. Programs that require a multiyear funding commitment might also be evaluated by discounting future costs to present value.

Although there is no general consensus of the appropriate discount rate for purposes of policy analysis, most cost-of-crime studies have used a yearly rate of between 2 and 3 percent, which is consistent with both the real (i.e., net of inflation) discount rate for worker wages over time and the real consumer interest rate over time.²¹ A similar consensus appears to have developed around a 3-percent net discount rate in health care economics (Gold et al. 1996). Some government agencies, however, have routinely used net discount rates of approximately 10 percent, and the Office of Management and Budget (1992) only recently reduced its required discount rate for regulatory policy analysis to 7 percent. The higher the discount rate, the lower the present value of future benefits. The choice of a proper discount rate is especially important in considering the benefits of youth prevention programs or other programs with benefits that

might not accrue for many years. Although this problem is nowhere near being settled, it is less of a problem in the criminal justice context than in the environmental arena, where it is common for latency periods to extend 20, 30, or 40 years or more (see Revesz 1999).

Fairness and equity

Benefit-cost analysis does not discriminate on the basis of socioeconomic status. A \$1,000 medical cost is valued at \$1,000, regardless of whether the injured person is rich or poor. Thus, the tool is politically neutral and can (and will) be overridden when other policy goals come into conflict. Thus, it is useful for the analyst to identify the demographic characteristics of the group that is likely to benefit most from a proposal.

A more subtle concern is the fact that the methodology itself may incorporate inequities in society. For example, if one is measuring lost wages to victims of crime—and those victims tend to be in the lower income quartiles—the benefits of a crime prevention program will be skewed downward based on the victim's income. If one were to compare a crime reduction program with another program that targets airline safety, for example, the typical wage rate might be higher for the airline accident victim than the crime victim. Further, if one were to conduct a contingent valuation survey of potential victims to determine their willingness to pay for crime reduction programs, the value is likely to be highly dependent on the wealth (i.e., ability to pay) of the respondent. Thus, from a public policy standpoint, benefit-cost analysis does indeed discriminate against society's less wealthy. If society deems this to be unfair, the analyst needs to make adjustments in the estimated costs and benefits to "neutralize" the effect of wealth on the estimated costs and benefits. This has been done to some extent in my earlier studies of the cost of crime. For example, the statistical value of life that is applied to the risk of death is based on the typical individual in the United States, not the typical crime victim. However, wage losses and reported short-term medical costs are necessarily taken from crime victim surveys.

Public perception versus objective measures of risk

Perhaps one of the most difficult issues that needs to be confronted as these methods are further developed and implemented is the fact that the public's perception of the risk of crime may not be the same as actual risk. Indeed, it has long been noted that as crime rates have declined over the past decade, the public's concern about crime has grown. There are many possible explanations for this disparity that are beyond the scope of this chapter (see Warr in this volume). Furthermore, any method that asks the public their willingness to pay for reduced crime inherently must confront the fact that the public might be

misinformed about the risk and severity of crime. Thus, public expenditures on crime prevention might be too high relative to what the public would demand if it was fully informed. The reverse is also true, of course, so that any objective measure of crime severity will ignore public perception and fear.

Concluding Remarks

Cost-effectiveness, benefit-cost analysis, and placing dollar values on the intangible costs of crime have all arrived at the criminal justice policy arena and will not go away. Increased scrutiny of government spending programs, coupled with new evidence that certain targeted prevention and rehabilitation programs work, provide the impetus for both new and innovative criminal justice policies *and* fierce public debate over their merits. This paper provides a framework for the future analysis of criminal justice policy from the perspective of solid empirical research.

This chapter has demonstrated the importance of considering both the costs and benefits of criminal justice and prevention programs. Although the practitioner might think in terms of cost per treated offender, the more relevant cost is cost per unit of benefit in which the benefit might include reduced crimes, successful drug treatment, and so forth. Even if a treatment program has been shown to reduce recidivism, it is important to know at what cost and at what benefit it does so. It does little good to know that a program costs \$10,000 per offender treated without knowing the corresponding benefit received for that \$10,000. Although I argue that dollars is the best metric to compare benefits and costs, it is not always possible to adequately quantify all costs or benefits. In such cases, it is still worthwhile to quantify as much as possible and to identify and list those that cannot be quantified, along with a qualitative description of their relative severity and importance.

This chapter should make clear, however, that we are far from the point where benefit-cost methods can be applied to criminal justice programs on a whole-sale basis. There is much more work to do on many of the components of estimating the cost of crime. In many cases, these same problems exist in other program areas that value lives and other intangibles. Among the issues that would benefit most from further work are: refinement and agreement on the statistical value of life, studies that directly elicit the public's willingness to pay for reduced crime (especially for property crimes, in which intangible losses are difficult to estimate), a better understanding of how to incorporate public perceptions into policy decisions, agreement on the proper discount rate for policy analysis involving long-term benefits, and measures of community wellness that go beyond individual crime victims. Perhaps most important,

however, is a continuing need for improved estimates of the crime control benefits of incarceration, alternative sanctions, and crime prevention programs.

My purpose in writing this chapter was twofold. First, because I am obviously in favor of encouraging the use of empirical tools in analyzing alternative criminal justice or crime prevention policies, I hope this chapter will encourage policy analysts to experiment with these tools and thereby improve their decisions. Although the techniques described in this chapter have been used for many years in other areas of public policy, they are just beginning to penetrate the criminal justice policy arena. The technique is not ideological but instead can be an important tool in the public policy debate. Both the hardline view of three strikes and you're out and the more compassionate view of focusing on prevention instead of punishment can be subjected to rigorous cost-benefit analyses in addition to political rhetoric.

My second goal is to encourage other researchers to devote serious time and energy to further improving the empirical evidence on the costs of crime and the benefits of crime prevention strategies. Criminal justice literature is far behind other areas of public policy, such as environmental protection and health care, that affect the health and well-being of our society. Literally hundreds of studies, peer-reviewed journal articles, conferences, and actual regulatory analyses have been conducted in these areas. It is time for the criminal justice research community to do the same.

I would like to thank Jonathan Caulkins and Brian Ostrom for helpful comments on an earlier draft. I would also like to acknowledge support from the Dean's Fund for Summer Research, Owen Graduate School of Management, Vanderbilt University. To contact me: phone 615-322-6814; fax 615-343-7177; e-mail mark.cohen@owen.vanderbilt.edu; or access <http://www.vanderbilt.edu/VCEMS/resume/cohen.html> on the Internet.

Notes

1. Gramlich (1981) contains a historical overview of benefit-cost analysis and a textbook treatment of the fundamentals of this technique. See also Mishan (1988) for a standard textbook on benefit-cost analysis.
2. President Ronald Reagan promulgated the first such requirement in 1981, Executive Order 12291 (46 *Federal Register* 13193). In 1993, President Bill Clinton issued Executive Order 12866 (58 *Federal Register* 51735). Although these Executive orders cannot supercede statutory provisions, they have had a dramatic effect on the manner in which regulatory agencies draft and analyze proposed rules.

3. For example, see Senate Bill S981, 105th Congress (1997), which would require all major rules to be accompanied by a benefit-cost analysis.
4. As discussed elsewhere in the chapter, there are methods that can be adopted to deal with the effect of wage inequality on estimating the cost of crime. In short, the analyst might adopt average wage rates in the United States in estimating the cost of lost wages. This puts all crime victims on an equal footing, regardless of their wealth.
5. As some critics have noted, all cost estimates are subject to considerable uncertainty, and categories of costs will be inevitably left unaccounted for (Zimring and Hawkins 1995). Thus, unless methods and assumptions are relatively consistent, or the unaccounted for costs are known to be relatively small, any such comparisons are likely to be of questionable value.
6. See Sherman et al. (1997) for a comprehensive examination of the effectiveness of alternative programs.
7. It would be double-counting to include both the value of stolen property and all collateral costs of the theft in an estimate of the social cost of theft.
8. French, Rachal, and Hubbard (1991) contain a useful discussion of the distinction between private, social, and external costs and provide a conceptual framework for estimating the costs of drug abuse.
9. See Zerbe and Dively (1994, 263–270) for a detailed discussion of the Kelman article and opposing views in support of the use of benefit-cost analysis.
10. Chapters 8 and 9 of Hellman (1980) provide a useful discussion of the economics of victimless crimes.
11. Details can be found in Cohen and Miller (1999a).
12. Regardless of the theoretical concerns, Cohen (1990) finds that the jury award method yields estimates of the cost of an Index crime that are consistent with the property value studies. Cohen and Miller (1999b) find that jury awards are consistent with the value of a life-year implied by the value-of-life studies based on worker wage rate differentials.
13. For an overview of the contingent valuation method, see Mitchell and Carson (1989). Smith (1996) compared the valuation from two different proposed environmental projects and found that citizens could make a clear distinction between the two projects.
14. The National Institute of Justice recently funded a more comprehensive public survey on attitudes toward sentencing and parole decisions that included a significant contingent valuation component to it, “Measuring Public Perception of Appropriate Prison Sentences,” NIJ grant no. 99–CE–VX–0001. For further details, contact Mark Cohen, the author of this chapter, who is the project manager for the survey. The only study the

author is aware of that employs a similar technique in the context of violence is by Ludwig and Cook (1999), who examine the public's willingness to pay for reduced gun violence.

15. A recent paper by Anderson (1999) attempts to estimate aggregate costs of all crimes.

16. One limitation of the approach used in Miller, Cohen, and Wiersema (1996) as well as in my earlier estimates is that the intangible costs of property crimes such as burglary are based on the probability of being a homicide victim. This is likely to underestimate the fear and feeling of being violated that accompanies being a burglary victim. Future research might improve on the intangible cost estimates of property crimes.

17. See Caulkins (in this volume) for a discussion of the difficulty of measuring drug costs.

18. These figures are calculated from Cohen, Miller, and Rossman (1994), tables 19–20.

19. This estimate includes alarms, guards, electronic and physical security, bomb and metal detection, access control, and so forth. Source: Bill Cunningham, Hallcrest Systems, Inc., as reported on the Security Industry Association Web site: http://www.siaonline.org/wp_giant.html.

20. As previously noted, excluding such ancillary benefits as increased productivity from high school graduation might skew policy decisions away from programs that have higher overall social benefits. Excluding intangible benefits might bias decisions away from programs that prevent violent crimes relative to property crimes and might also result in certain programs failing a benefit-cost test when the intangible benefits exceed costs.

21. Note that these are net discount rates, as they already account for inflation. Thus, for example, a 2-percent discount rate would be consistent with long-term cost-of-living increases of 4 percent and long-term interest rates of 6 percent.

References

Abt Associates Inc. 1995. *What America's users spend on illegal drugs, 1988–93*. NCJ 153852. Washington, D.C.: Executive Office of the President, Office of National Drug Control Policy.

Anderson, David A. 1999. The aggregate burden of crime. *Journal of Law and Economics* 42 (2): 611–642.

Aos, Steve, Polly Phipps, Robert Barnoski, and Roxanne Lieb. 1999. *The comparative costs and benefits of programs to reduce crime*. Olympia: Washington State Institute for Public Policy.

- Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. Report of the NOAA panel on contingent valuation. *Federal Register* 58. (15 January): 4601–4614.
- Association of Certified Fraud Examiners. 1995. *Report to the Nation on occupational fraud and abuse*. Austin, Texas: Association of Certified Fraud Examiners.
- Austin, James. 1986. Using early release to relieve prison crowding: A dilemma in public policy. *Crime & Delinquency* 32:404–502.
- Ayres, Ian, and Steven D. Levitt. 1998. Measuring positive externalities from unobservable victim precaution: An empirical analysis of LoJack. *Quarterly Journal of Economics* 113:43–77.
- Barnett, W. Steven. 1993. Benefit-cost analysis of preschool education: Findings from a 25-year follow-up. *American Journal of Orthopsychiatry* 63:500–508.
- Becker, Gary. 1968. Crime and punishment: An economic approach. *Journal of Political Economy* 78:169–217.
- Black, Donald J. 1998. *The social structure of right and wrong*. San Diego: Academic Press.
- Burt, Martha R., and Bonnie L. Katz. 1985. Rape, robbery, and burglary: Responses to actual and feared victimization, with special focus on women and the elderly. *Victimology: An International Journal* 10:325–358.
- Butterfield, Fox. 1996. Study reveals high cost of crime in U.S. *New York Times*, 22 April.
- Cohen, Mark A. 1999. Alcohol, drugs, and crime: Is ‘crime’ really one-third of the problem? *Addiction* 94:636–639.
- . 1998. The monetary value of saving a high risk youth. *Journal of Quantitative Criminology* 14:5–33.
- . 1990. A note on the cost of crime to victims. *Urban Studies* 27:125–132.
- . 1989. Corporate crime and punishment: A study of social harm and sentencing practice in the Federal courts, 1984–1987. *American Criminal Law Review* 26:605–660.
- . 1988a. Pain suffering and jury awards: A study of the cost of crime to victims. *Law & Society Review* 22:538–555.
- . 1988b. Some new evidence on the seriousness of crime. *Criminology* 26:343–353.

Cohen, Mark A., and Ted R. Miller. 1999a. The monetary value of pain and suffering due to criminal victimization: Evidence from jury awards. Working paper, Vanderbilt University, Nashville.

———. 1999b. Willingness to award nonmonetary damages and the implied value of life from jury awards. Working paper, Vanderbilt University, Nashville.

———. 1998. The cost of mental health care for victims of crime. *Journal of Interpersonal Violence* 13:93–100.

Cohen, Mark A., Ted R. Miller, and Shelli B. Rossman. 1994. The costs and consequences of violent behavior in the United States. In *Understanding and preventing violence, Volume 4: Consequences and control*, edited by Albert J. Reiss, Jr., and Jeffrey A. Roth. Washington, D.C.: National Academy Press.

Cook, Philip J. 1986. The demand and supply of criminal opportunities. In *Crime and justice: An annual review of research*, edited by Michael H. Tonry and Norval Morris. Vol.7. Chicago: University of Chicago Press.

———. 1983. Costs of crime. In *Encyclopedia of crime and justice*, edited by Sanford H. Kadish. New York: Free Press.

Cullen, Francis T., Bruce G. Link, and Craig W. Polanzi. 1982. The seriousness of crime revisited. *Criminology* 20:83–102.

DiIulio, John J., and Anne M. Piehl. 1991. Does prison pay? The stormy national debate over the cost-effectiveness of imprisonment. *Brookings Review* 9:28–35.

Donohue, John T., and Peter Siegelman. 1998. Allocating resources among prisons and social programs in the battle against crime. *Journal of Legal Studies* 27:1–44.

Executive Office of the President, Office of National Drug Control Policy. 1991. *What America's users spend on illegal drugs*. Technical paper. Washington, D.C.

Fabian, Robert. 1994. The quality approach. In *Valuing health for policy: An economic approach*, edited by George Tolley, Donald Kenkel, and Robert Fabian. Chicago: University of Chicago Press.

French, Michael T., J. Valley Rachal, and Robert L. Hubbard. 1991. Conceptual framework for estimating the social cost of drug abuse. *Journal of Health & Social Policy* 2:1–22.

Gold, Marthe R., Joanna E. Siegel, Louise B. Russell, and Milton C. Weinstein, eds. 1996. *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Gramlich, Edward M. 1981. *Benefit-cost analysis of government programs*. Englewood Cliffs, New Jersey: Prentice Hall.

- Gray, Charles M., ed. 1979. *The costs of crime*. Beverly Hills: Sage Publications.
- Gray, T. 1994. Research note: Using cost-benefit analysis to measure rehabilitation and special deterrence. *Journal of Criminal Justice* 22:569–575.
- Greenwood, Peter W., Karyn E. Model, C. Peter Rydell, and James Chiesa. 1996. *Diverting children from a life of crime: Measuring costs and benefits*. Santa Monica, California: RAND.
- Greenwood, Peter W., C. Peter Rydell, Allan F. Abrahamse, Jonathan P. Caulkins, James Chiesa, Karyn E. Model, and Stephen P. Klein. 1994. *Three strikes and you're out: Estimated benefits and costs of California's new mandatory-sentencing law*. Santa Monica, California: RAND.
- Hann, Robert G. 1972. Crime and the cost of crime: An economic approach. *Journal of Research in Crime and Delinquency* 9:12–30.
- Harwood, Henrick J., Douglas Fountain, and Gina Livermore. 1998. *The economic costs of alcohol and drug abuse in the United States, 1992*. Washington, D.C.: U.S. Department of Health and Human Services, National Institutes of Health.
- Hellman, Daryl A. 1980. *The economics of crime*. New York: St. Martin's Press.
- Hellman, Daryl A., and Joel L. Naroff. 1979. The impact of crime on urban residential property values. *Urban Studies* 16:105–112.
- Hoehn, John, P., Mark C. Berger, and Glenn C. Blomquist. 1987. A hedonic model of interregional wages, rents, and amenity values. *Journal of Regional Science* 27:605–620.
- Irwin, John, and James Austin. 1994. *It's about time: America's imprisonment binge*. Belmont, California: Wadsworth Publishing Company.
- Kelman, Steven. 1981. Cost benefit analysis—An ethical critique. *Regulation* 5 (January/February).
- Klaus, Patsy A. 1994. *The cost of crime to victims: Crime Data Brief*. NCJ 145865. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Laband, David N., and John P. Sophocleus. 1992. An estimate of resource expenditures on transfer activity in the United States. *Quarterly Journal of Economics* 107:959–983.
- Levitt, Steven D. 1996. The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. *Quarterly Journal of Economics* 111 (2): 319–351.
- Lindgren, Sue A. 1997. *Justice expenditure and employment extracts, 1992: Data from the annual general finance and employment surveys*. NCJ 148821. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Ludwig, Jens, and Philip J. Cook. 1999. The benefits of reducing gun violence: Evidence from contingent-valuation survey data. Working paper 7166, National Bureau of Economic Research, Cambridge, Massachusetts.

Maltz, Michael D. 1975. Measures of effectiveness for crime reduction programs. *Operations Research* 23:452–474.

Macmillan, Ross. 2000. Adolescent victimization and income deficits in adulthood: Rethinking the costs of criminal violence from a life course perspective. *Criminology* (May).

Martin, J.P., and J. Bradley. 1964. Design of a study of the cost of crime. *British Journal of Criminology* 4:591–603.

McChesney, Fred S. 1993. Boxed in: Economists and the benefits of crime. *International Review of Law and Economics* 13 (2): 225–231.

Miczek, Klaus A., Joseph F. DeBold, Margaret Haney, Jennifer Tidey, Jeffrey Vivian, and Elise M. Weerts. 1994. Alcohol, drugs of abuse, aggression, and violence. In *Understanding and preventing violence, Volume 3: Social influences*, edited by Albert J. Reiss, Jr., and Jeffrey A. Roth. Washington, D.C.: National Academy Press.

Miller, Ted R., Mark A. Cohen, and Brian Wiersema. 1996. *Victim costs and consequences: A new look*. Research Report. NCJ 155282. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Mishan, E.J. 1998. *Cost-benefit analysis: An informal introduction*. 4th ed. London: Unwin Hyman.

Mitchell, Robert C., and Richard T. Carson. 1989. *Using surveys to value public goods*. Washington, D.C.: Resources for the Future.

Office of Management and Budget. 1992. Guidelines and discount rates for benefit-cost analysis of Federal programs. *Federal Register* 57:53519.

Phillips, Llad, and Harold L. Votey, Jr. 1981. *The economics of crime control*. Beverly Hills: Sage Publications.

Prentky, Robert, and Ann Wolbert Burgess. 1990. Rehabilitation of child molesters: A cost-benefit analysis. *American Journal of Orthopsychiatry* 60:108–117.

Rajkumar, Andrew S., and Michael T. French. 1997. Drug abuse, crime costs, and the economic benefits of treatment. *Journal of Quantitative Criminology* 13:291–324.

Rand, Michael. 1998. *Criminal victimization 1997: Changes 1996–97 with trends 1993–97*. NCJ 173385. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Reuter, Peter, Robert MacCoun, and Patrick Murphy. 1990. *Money from crime: A study of the economics of drug dealing in Washington, D.C.* Santa Monica, California: RAND.

Revesz, Richard L. 1999. Environmental regulation, cost-benefit analysis, and the discounting of human lives. *Columbia Law Review* 99 (4): 941–1017.

Rizzo, Mario J. 1979. The cost of crime to victims: An empirical analysis. *Journal of Legal Studies* 8:177–205.

Rossi, Peter H., and Richard A. Berk. 1997. *Just punishments: Federal guidelines and public views compared.* New York: Aldine de Gruyter.

Rossi, Peter H., Emily Waite, Christine E. Boise, and Richard A. Berk. 1974. The seriousness of crimes: Normative structure and individual differences. *American Sociological Review* 39:224–237.

Rydell, Charles P., and Susan S. Everingham. 1994. *Controlling cocaine: Supply versus demand programs.* Santa Monica, California: RAND.

Sherman, Lawrence W., Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter, and Shawn Bushway. 1997. *Preventing crime: What works, what doesn't, what's promising.* Research Report. NCJ 165366. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Smith, V. Kerry. 1996. Can contingent valuation distinguish economic values for different public goods? *Land Economics* 72 (2): 139–151.

Streff, Fredrick M., Lisa J. Molnar, Mark A. Cohen, Ted R. Miller, and Shelli B. Rossman. 1992. Measuring costs of traffic crashes and crime: Tools for informed decision making. *Journal of Public Health Policy* 13:451–471.

Tengs, Tammy O., Miriam E. Adams, Joseph S. Pliskin, Dana Gelb Safran, Joanna E. Siegel, Milton C. Weinstein, and John D. Graham. 1995. Five-hundred life-saving interventions and their cost-effectiveness. *Risk Analysis* 15:369–390.

Thaler, Richard. 1978. A note on the value of crime control: Evidence from the property market. *Journal of Urban Economics* 5:137–145.

Titus, Richard M., Fred Heinzelmann, and John M. Boyle. 1995. Victimization of persons by fraud. *Crime & Delinquency* 41:54–72.

Trumbull, William N. 1990. Who has standing in cost-benefit analysis? *Journal of Policy Analysis and Management* 9:201–218.

Viscusi, W. Kip. 1993. The value of risks to life and health. *Journal of Economic Literature* 31:1912–1946.

Welsh, Brandon C., and David P. Farrington. 2000. Monetary costs and benefits of crime prevention programs. *Crime and justice: A review of research*, edited by Michael Tonry. Vol. 27. Chicago: University of Chicago Press.

Wolfgang, Marvin E., Robert M. Figlio, Paul E. Tracy, and Simon I. Singer. 1985. *The National Survey of Crime Severity*. Survey Report, NCJ 96017. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Zedlewski, Edwin W. 1985. When have we punished enough? *Public Administration Review* 45:771-779.

Zerbe, Richard O., Jr., and Dwight D. Dively. 1994. *Benefit-cost analysis in theory and practice*. New York: HarperCollins.

Zimring, Franklin E., and Gordon Hawkins. 1995. *Incapacitation: Penal confinement and the restraint of crime*. New York: Oxford University Press.

Measuring the Sexual Victimization of Women: Evolution, Current Controversies, and Future Research

by Bonnie S. Fisher and Francis T. Cullen

In the 1970s, the growing interest in the victimization of women prompted claims that rape and sexual assault in the United States, heretofore rendered invisible, were rampant. Existing data sources, including the Federal Bureau of Investigation's Uniform Crime Reports and the Bureau of Justice Statistics' National Crime Survey (later called the National Crime Victimization Survey), were roundly criticized for methodological flaws that led to the substantial underreporting of the sexual victimization women experienced. These concerns in turn led to the quest to construct measures that would more accurately assess the true extent of females' sexual victimization. This essay examines the development and key methodological issues characterizing this effort to measure the extent and types of sexual victimization perpetrated against women.

In the 1980s, Koss and others constructed detailed surveys devoted to measuring sexual victimization. They made three methodological advances: (1) the use of legal statutes as the basis for developing measures of rape and other victimizations; (2) the development of "behaviorally specific" questions that used graphic language describing the elements of a criminal victimization, so as to cue respondents to recall victimization incidents; and (3) the assessment of a wide, as

Bonnie S. Fisher is Associate Professor and Francis T. Cullen is Distinguished Research Professor with the Division of Criminal Justice, University of Cincinnati.

A
B
S
T
R
A
C
T

317

opposed to a narrow, range of sexually victimizing conduct. Researchers in the 1990s incorporated these advances into their surveys and merged them with the methodology utilized by the National Crime Victimization Survey. From Koss and others, they borrowed the strategy of asking behaviorally specific questions to cue or “screen” respondents who may have been sexually victimized. From the National Crime Victimization Survey, they employed an “incident report,” which relied on detailed questions to determine whether a criminal victimization took place and, if so, how it should be categorized.

The research in the 1990s has provided valuable findings but also has illuminated a host of methodological issues that warrant further attention. Many of these issues involve the potential inconsistencies, if not biases, in responses that are introduced by the screen questions—incident report methodology. To resolve these issues, the next generation of research on sexual victimization might benefit from more sophisticated surveys that obtain qualitative data from respondents within the confines of a structured survey instrument.

A
B
S
T
R
A
C
T

Until the 1970s, relatively little consideration was given to the ways in which females were victimized criminally (Belknap 1996). However, at this time, criminologists—many of them female scholars who embraced feminist theory and were politicized by the prevailing social movements of the 1960s—turned their attention to examining the sexual victimization of women in the United States. Rape was reconceptualized as a crime of power—of males using their physical and social dominance to sexually assault women “against their will” (see, e.g., Brownmiller 1975). There also was an increasing effort to illuminate that rape was not confined to “stranger” victimizations but could be perpetrated by men that females knew intimately. This observation resulted in the creation of a new language to speak of intimate rapes—terms such as “acquaintance rape,” “date rape,” and “marital rape” (see Parrot and Bechhofer 1991; Russell 1982, 1984). Advocates simultaneously campaigned to implement legal changes that afforded rape victims more protection in the criminal justice system, including, for example, rape shield laws, police units trained to work more sensitively and effectively with rape victims, and rape crisis centers (Geis 1977; Goldberg-Ambrose 1992; Jensen and Karpos 1993; Muehlenhard et al. 1994).

Feminist scholars often argued that the sexual victimization of women was integral to patriarchy and thus was widespread in the United States. The difficulty, however, was that the empirical literature on the extent and nature of sexual victimization was sparse. The main sources of the extent of criminal victimization in the Nation—the Uniform Crime Reports (UCR) and the National Crime Survey (NCS)—were deemed hopelessly biased. Scholars claimed that these data sources substantially underestimated the true incidence of rape—the UCR because it relied on reported crimes when many rapes were not reported to the police, and the NCS because it was ostensibly poorly designed to elicit reports of rape from interviewees who had, in fact, been raped (Gordon and Riger 1989; Koss 1992, 1993a, 1993b; Russell 1984). The critical issue was how to develop measurement strategies that would reveal the “true” extent in American society of not only rape but also other forms of sexual victimization.

This task was taken up by a number of scholars—with Mary Koss and her colleagues being the most influential—who over the past two decades have used specially designed surveys to ask women about the extent of their sexual victimization (e.g., DeKeseredy and Kelly 1993; Schwartz and Pitts 1995; Kilpatrick, Edmunds, and Seymour 1992; Koss and Oros 1982; Koss and Gidycz 1985; Koss, Gidycz, and Wisniewski 1987). This research generally revealed that the extent of rape was considerably higher than that reported by the official UCR statistics or by the Government-sponsored NCS data. Further, this work showed that a large percentage of women had experienced types of sexual victimization that neither the UCR nor the NCS measured, namely, sexual

Despite the growth in the number of research studies, investigators have continued to be plagued by the daunting issue of how to accurately measure the extent and nature of rape and other types of sexual victimization.

coercion and unwanted sexual contact. These data thus suggested that the sexual victimization of women was a pervasive and multifaceted social problem that warranted both the attention of policymakers and governmental intervention (Crowell and Burgess 1996; Fitzgerald 1996; Goodman et al. 1993; Koss 1992; Koss, Gidycz, and Wisniewski 1987).

Despite the growth in the number of research studies, investigators have continued to be plagued by the daunting issue of how to accurately measure the extent and nature of rape and other types of sexual victimization. Serious methodological controversies exist and affect who is, and is not, counted as a victim. In particular, researchers have come to realize that conceptually defining and then operationalizing sexual victimization are complicated and, to a degree, imperfect enterprises—especially when deciding when an unwanted sexual advance crosses the line from imprudence to criminal behavior.

These methodological challenges have opened the way for conservative commentators to charge that the supposed “epidemic of rape” is an invention of feminist scholars. What about the data showing that rape is widespread? The critics accuse feminists of merely finding what they set out to find—of using research methods that are so flawed and ideologically biased that they present estimates of rape that are inflated many times over (Gilbert 1991, 1995; Roiphe 1993). In particular, critics contend, the definitions of rape and the survey questions used to measure rape are so broadly or poorly phrased that they “pick up” and count as rape a wide diversity of conduct, most of which could hardly be considered criminal in a legal sense.

In this contentious context—and after nearly three decades of concern about and research into the rape and sexual victimization of women—we have reached the point where it is useful to “take stock” of what we have learned and what we need to learn. The purpose of this essay is to review how the major studies have defined and operationalized rape and other types of sexual victimization. We attempt to elucidate what is distinctive about each of these approaches and to convey the strengths and potential problems inherent in given measurement strategies. This analysis is intended to set the stage for discussing the definitional and operationalization issues that will have to be addressed and rigorously examined if we are to advance our understanding of sexual victimization beyond its current level.

Specifically, in the section on the National Crime Victimization Survey (NCVS), we pay special attention to the strengths and weaknesses of the NCVS' attempt to measure sexual victimization through a two-step process that involves "screen questions" and an "incident report." In the next section, we turn our attention to the evolution of surveys specifically designed to measure sexual victimization. We focus in particular on Koss' classic "Sexual Experiences Survey" instrument and the critical analyses it elicited. In the following section, we consider more recent studies that have attempted to employ more sophisticated methods for measuring sexual victimization. We next review our own research as a vehicle for exploring central methodological issues that continue to confront the measurement of sexual victimization. Finally, we conclude this essay with recommendations for future research that attempts to measure sexual victimization.

Before embarking on this excursion through the evolution of the methods used to measure sexual victimization, three additional considerations must be mentioned. First, this essay examines only research based on victimization surveys, the main method used to measure the sexual victimization of women. For two reasons, we do not examine official measures of sexual victimization—including revisions to the UCR through the National Incident-Based Reporting System (NIBRS)¹: the space available to us is restricted, and official statistics measure only crimes known to law enforcement, which are a small fraction of the sexual victimizations that occur. Lack of space also precludes us from examining sexual harassment. Its definition and measurement are plagued by some of the same issues that we discuss in this chapter (see, for review, Fitzgerald 1996; Gruber 1990).

Second, our essay focuses primarily on the nature and implications of the methodological choices researchers have made in developing questions and designing survey instruments to measure sexual victimization. We do not assess how other methodological choices—such as in drawing samples, in the context in which surveys are conducted, and in training interviewers—might affect individuals' responses on sexual victimization surveys and, in turn, the findings these surveys report. We urge interested readers to consult the emerging literature on these and related issues (see, e.g., DeKeseredy and Schwartz 1998; Lynch 1996a, 1996b; Koss 1992, 1993a, 1993b, 1996; Koss and Cook 1993; Schwartz 1998; see also Cantor and Lynch in this volume).

Third, it also is beyond the scope of this essay to examine systematically how legal definitions of rape and sexual assault might affect estimates of sexual victimization for specific jurisdictions. Thus, we do not examine whether variations in how States define rape and sexual assault potentially influence how many sexually victimizing acts are counted as "crimes" in given jurisdictions. We also do not consider whether specific legal definitions of these acts may

influence what acts victims in different States report to the police or report as crimes on victimization surveys. Even so, throughout this essay, we do consider how statutory legal definitions shape the content and interpretation of victimization surveys that are intended to measure *criminal* sexual victimization.

For example, the crime of rape traditionally has been defined as carnal knowledge of a person forcibly and against the person's will (see Searles and Berger 1987; Spohn and Horney 1992). But what does "carnal knowledge" encompass? At one time, this category included only penile-vaginal penetration, but legal reforms in the 1980s expanded this concept to cover, in many States, sexual penetration generally (i.e., vaginal, oral, and anal) and with different objects (e.g., tongue, bottle). Further, the construct of "against a person's will" now includes not only whether physical force was used or bodily harm threatened, but also whether the victim was unconscious or unable to give consent due to a young age, mental illness, mental retardation, or the administration of an intoxicating or anesthetic substance (Gilbert 1997; Searles and Berger 1987; Spohn and Horney 1992).

The challenge is how to develop measures that can count all acts that, according to these legal definitions, would fall under the rubric of rape. Researchers have generally followed one of two strategies: (1) select the statute of a specific State and use this to guide what legal elements are measured, or (2) select the elements of rape that are common to most State legal statutes and measure these. These strategies do not obviate fully the possibility of measurement error being introduced, but researchers learned early on that the failure to use a reasonable legal definition of rape to inform their analysis could result in far greater imprecision in what was "counted" as a rape. The trickier issue, as researchers discovered, was how to word questions in a survey instrument that were capable of measuring the legal elements of rape. How many questions were asked, how questions were sequenced, what behaviors were described, what specific words were used—all these became the focus of critical analyses and, in some cases, of ideological dispute. Much of the discussion to follow will dwell on these important details and on their implications for measuring rape and other forms of sexual victimization.

The National Crime Victimization Survey

The major victimization survey in the United States is the National Crime Victimization Survey, originally termed the National Crime Survey. The NCS and the NCVS have been major sources of information about sexual victimization, although they have also been widely criticized for potentially underestimating the amount of rape in America. The NCS was redesigned into the

NCVS during the late 1980s and early 1990s, in part to address this shortcoming in the measurement of sexual victimization (see Bachman and Taylor 1994; Canter and Lynch in this volume). Annual estimates of rape and sexual assault from the redesigned NCVS were made available starting in 1993.

The NCVS is under the auspices of the Bureau of Justice Statistics (BJS). Like its NCS predecessor, the redesigned NCVS is a national household-based survey of an individual's experiences with crime victimization, whether or not the incidents were reported to the police. Using a stratified multistage cluster sample, the NCVS collects victimization data from a sample of about 90,000 individuals living in about 45,000 housing units (U.S. Department of Justice [DOJ], BJS 1997, appendix II). The survey employs a rotating panel design of housing units, with each unit being in the sample for 3.5 years; new households are constantly being added to the sample as other households complete their time in the sample. All household members age 12 and older are interviewed by male and female interviewers every 6 months during this period (seven times in all). The first interview is not employed in the reported estimates but is used only to bound the second interview, thus establishing a 6-month reference period for respondents. Thereafter, the previous interview serves as the bound for the subsequent interview. The first and fifth interviews are done in person; all others are conducted by telephone. Further, the NCVS uses computer-assisted telephone interviewing (CATI).

How the NCVS measures victimization

The major feature of the NCVS is that victimization is measured in a two-step process. Respondents are first read a series of "screen questions" on whether a victimization may have occurred and, if so, they are then interviewed through a lengthy "incident report" about what may have occurred. In the NCVS, the interviews include seven individual-level screen questions. The intent of the screen questions is to "cue" respondents, or jog their memory, as to whether they had experienced a criminal victimization within the 6-month reference period. For this reason, these questions are intended not to be redundant and are designed to elicit a "yes" or "no" regarding victim incidents (Lynch 1996a, 1996b). When a respondent says "yes" to any screen question, the interviewer then asks him or her "what happened?" (i.e., "briefly describe the incident[s]"). The interviewer then asks "how many times" that type of incident occurred during the reference period. Examples of screen questions are furnished in exhibit 1, which is presented in the section on screen questions.

At the conclusion of the screen questions, the interviewer administers an "incident report" for each time a respondent mentioned an incident had occurred. This incident report contains detailed questions about the nature of the incident

(e.g., month, time, and place of incident; characteristics of the offender; police reporting behavior). It includes questions that ask whether the offender hit, tried to attack, or threatened the respondent, how the respondent was attacked or threatened, and whether injuries were suffered.

It is essential to understand that the NCVS uses the incident report *to determine, and thus count, whether a victimization has occurred*. Again, the purpose of the screen questions is to cue the respondent to remember victimization events. Answering “yes” to a screen question only allows the respondent to gain entry into the incident report. At this point, the respondent must answer “yes” to more detailed questions and/or give a verbal account describing a victimization to be counted as having experienced a particular type of victimization. An example of how incident report questions are used to categorize a victimization is presented later in exhibit 2.

Notably, most sexual victimization surveys use only a one-step process, essentially measuring victimization through a series of “cueing” questions that are meant to prompt respondents to recall victimization incidents. The risk of this one-step approach—that is, of not having an incident report—is that it may include reports of victimization that, on closer scrutiny, would not qualify legally as a rape or other type of sexual victimization. The advantage of the NCVS is that the incident-report questions in essence confirm or validate what occurred to the respondent and thus *potentially* allow for a more valid categorization of incidents as to (1) whether they took place and, if so, as to (2) what type of victimization transpired. Despite this advantage, the quality of the NCVS as a measure of sexual victimization is contingent on how well the screen questions cue respondents and whether the incident report validly classifies what respondents experienced in the course of a victimization incident. These issues will occupy much of our attention in the sections to follow.

The first step in measuring sexual victimization: NCVS screen questions

In the original NCS, the respondents were asked four screen questions that assessed whether they were “attacked” or otherwise physically threatened. The NCS assumed that this broad-based inquiry into being attacked would prompt respondents who had suffered an attack involving rape to answer “yes” to one of these four questions. As critics correctly pointed out, however, the respondents were never asked *directly* or explicitly if they had experienced a completed or attempted rape (Eigenberg 1990; Koss 1992, 1993a). The use of an *indirect* means on the NCS to “cue” respondents as to their rape victimization, critics argued, would almost certainly mean that some rape or attempted

rape victims would not respond “yes” to any of the four screen questions. Accordingly, the NCS was held to be biased in the direction of underestimating the true incidence of rape and attempted rape victimization (see, e.g., Koss 1992, 1993a, 1996). Nearly all studies of rape in the United States have reported estimates much higher than those found by the NCS (see Bachman and Taylor 1994). Moreover, the NCS only measured one type of sexual victimization, rape, and thus did not assess sexual assault.

Given these considerations, the major redesign efforts of the NCS were concentrated in the cueing strategy used in the screening interview. First, new screen questions were added to the NCVS that directly asked respondents about “rape, attempted rape, and any other type of sexual attack” and about “forced or unwanted sexual acts.” Second, all the NCS screen questions were reworded and new “cues” or information about a potential victimization incident were added. The purpose of adopting more specific cues in the screen questions was to expand the frame of reference for the respondents so as to better stimulate their recall of an incident, thus helping to reduce underreporting because of forgotten incidents.

Thus, the revised NCVS screen questions begin with a reference to a type of criminal victimization that may have been experienced (e.g., “were you attacked or threatened”), which is followed by a list of short cue responses about the potential victimization. This list includes cues regarding specific places or situations in which the victimization could have occurred (e.g., “at work or at school”), objects that could have been used (e.g., “with any weapon, for instance, a gun or knife”), actions that could have been associated with the victimization (e.g., “face-to-face threats”), and people who potentially might have perpetrated the criminal act (e.g., “a relative or family member”) (U.S. DOJ, BJS 1994). Each of these cues is intended to diminish the effects of subjective interpretations of the questions and to help the respondent structure the recall task before answering the question, “Did any incidents of this type happen to you?” (See exhibit 1.)

Exhibit 1 lists the four screen questions used on the NCVS to screen for potential incidents of rape and sexual assault. The interviewer asks each respondent these screen questions. The interviewer also is instructed to “briefly describe incident(s)” that the respondent noted. This verbatim account is recorded by the interviewer but is not part of the data files that BJS archives at the Inter-University Consortium for Political and Social Research at the University of Michigan. The numbers used in exhibit 1—40a, 41a, 42a, and 43a—correspond to the question numbers used on the NCVS.

Exhibit 1. Specific NCVS individual-level screen questions designed to elicit reports of rape, sexual assault, and unwanted sexual contact

NCVS screen questions
<p>40a. (Other than any incidents already mentioned,) since _____, 19 _____, were you attacked or threatened OR did you have something stolen from you—(a) At home including the porch or yard—(b) At or near a friend’s, relative’s, or neighbor’s home—(c) At work or school—(d) In places such as a storage shed or laundry room, a shopping mall, restaurant, bank, or airport—(e) While riding in any vehicle—(f) On the street or in a parking lot—(g) At such places as a party, theater, gym, picnic area, bowling lanes, or while fishing or hunting—OR (h) Did anyone ATTEMPT to attack or ATTEMPT to steal anything belonging to you from any of these places?</p>
<p>41a. (Other than any incidents already mentioned,) has anyone attacked or threatened you in any of these ways (<i>Exclude telephone threats</i>)—(a) With any weapon, for instance, a gun or knife—(b) With anything like a baseball bat, frying pan, scissors, or stick—(c) By something thrown, such as a rock or bottle—(d) Include any grabbing, punching, or choking—(e) Any rape, attempted rape, or other type of sexual attack—(f) Any face-to-face threats—OR (g) Any attack or threat or use of force by anyone at all? Please mention it even if you are not certain it was a crime.</p>
<p>42a. People often don’t think of incidents committed by someone they know. (Other than any incidents already mentioned,) did you have something stolen from you OR were you attacked or threatened by (<i>Exclude telephone threats</i>)—(a) Someone at work or school—(b) A neighbor or friend—(c) A relative or family member—(d) Any other person you’ve met or known?</p>
<p>43a. Incidents involving forced or unwanted sexual acts are often difficult to talk about. (Other than any incidents already mentioned,) have you been forced or coerced to engage in unwanted sexual activity by—(a) Someone you didn't know before—(b) A casual acquaintance—OR (c) Someone you know well?</p>
<p>_____ Note: Question numbers are from the individual screen questions in the basic screen questionnaire (Perkins et al. 1996, 124–125).</p>

In the research literature, most discussion of the NCVS has centered on questions 41a, 42a, and 43a in exhibit 1 (see Bachman and Taylor 1994; Koss 1992, 1996). BJS uses these questions to screen for all types of personal crimes that the NCVS measures (Perkins et al. 1996). For our purposes, we also included the first question in exhibit 1 (question 40a) because it specifically asks the respondent about being threatened in different places. Note that verbal threats are part of the NCVS' definition of rape and sexual assault (see next section). This means that the NCVS uses incidents involving only threats—as opposed to limiting counts to attempted or completed acts—in its estimates of rape and sexual assaults.

Again, the screen questions are, in essence, “gatekeepers” to the respondent proceeding on to complete an incident report. As a result, the wording of each question is important to the reporting of an incident. In the second question in exhibit 1 (question 41a), respondents are explicitly asked about rape, attempted rape, and other types of sexual attack (instead of the omnibus “attack you” or “try to attack you” question that was used in the NCS screen questions). The term “types of sexual attack” is a broad cue to respondents and may work to capture a range of sexual victimization incidents that can then screen into an incident report for further clarification.

Critics of the second question point out that this type of question rests on several assumptions: the respondent knows how rape is defined, perceives what happened to her or him as rape, and remembers the experience with this conceptual label (Gilbert 1997; Koss 1992, 1993a). What do the empirical data show with respect to these assumptions? First, in one study (Russell 1982), some incidents reported by the respondents as rape or attempted rape victimizations did not satisfy the legal criteria for rape—a source of overreporting. Second, other studies point to the possibility of underreporting, because rape victims frequently fail to realize that their victimization qualifies as a crime and avoid using the term “rape” to describe their experience. Koss (1988) reported, for example, that only 27 percent of the college women labeled their experiences with forced, unwanted intercourse as rape.

Although BJS employs question 42a on the NCVS instrument to screen for all types of personal crimes, this question could be used to screen for rape and sexual assault because it makes reference to potential offenders who could sexually victimize a respondent. Given that a majority of rapes and other forms of sexual victimization are committed by someone the victim knows, this question helps to dispel notions that the survey is only measuring incidents committed by strangers (see Crowell and Burgess 1996; Koss 1992). Although the term “attacked or threatened” used in this question is an indirect way to ask about rape or sexual assault, the term is broad enough to capture a range of incidents

that may, or may not, be sexual victimizations (or even victimizations of any kind) once the incident-level responses are taken into account to classify the incident. In this respect, the question could encourage reporting. Use of the term, however, assumes that respondents will make the connection between being attacked or threatened and unwanted sexual experiences (Koss 1992). If respondents do not make this connection, then the gatekeeping function of this question may hinder respondents on the NCVS from reporting a rape or sexual assault incident.

Question 43a on the NCVS explicitly provides the respondents, especially hesitant ones, with a second chance to report a sexual victimization; this is also the case for respondents who do not use the term “rape” or “sexual assault” to label their respective incident. Within this question, a general definition of what is being measured is provided to the respondent so that she or he knows what type of experience the interviewer is asking about. The question is behavior specific; it twice specifically asks respondents about incidents involving “forced or unwanted sexual acts” and about the respondent being “forced or coerced to engage in unwanted sexual activity.” The terms used are explicit, yet they are broad enough to include a range of incidents that can be clarified in the incident report. As a gatekeeping question, it could cue or otherwise encourage respondents to report their victimization incidents to the interviewer. Koss (1992) points out several other advantages to this line of questioning that include minimizing the street-violence context within the item. She also suggests, however, that this question might be strengthened if respondents were told to mention the incident even if they were not certain it was a crime (e.g., see question 41a).

How well do these four questions screen for incidents whose final classification is a rape or a sexual assault? Using data from the 1992 to 1996 incident file, we examined the degree of agreement between the screen question on which the incident screened into an incident report and the final crime classification given to each incident. We found that 93.4 percent of the incidents counted as rapes and 94.8 percent of the incidents counted as sexual assaults were screened into the incident report on one of these four questions. By comparison, in the NCS, the proportion of incidents counted as rape that had screened into the incident report (on the four questions used to screen for rape victimization) was about 10 percentage points lower (Dodge 1984). Taken together, these results suggest that the revised NCVS screen questions were an improvement over those used in the NCS and were the main source for those incidents counted as sexual victimizations by the survey in the 1992 to 1996 data. Note, however, that the figures cited do not tell us the extent to which the four NCVS screen questions were able to cue all respondents who had been sexually

victimized to report this fact to the interviewer. As such, we cannot estimate how many incidents were not cued by these incident questions and thus never reached the incident report.

The second step in measuring sexual victimization: The NCVS incident report

Operationalizing completed, attempted, or threatened rape

In the NCVS, rape is defined as follows:

Forced sexual intercourse including both psychological coercion as well as physical force. Forced sexual intercourse means vaginal, anal, or oral penetration by the offender(s). This category also includes incidents where the penetration is from a foreign object such as a bottle. Includes attempted rapes, male as well as female victims, and both heterosexual and homosexual rape. Attempted rape includes verbal threats of rape. (U.S. DOJ, BJS 1997, 149)

This definition incorporates the legal definition of rape found in the rape reform legislation passed in the 1970s and 1980s. Koss (1993a) criticizes the NCVS' definition for being ambiguous with respect to the term "psychological coercion." She suggests that this term is probably meant to refer to verbal threats of bodily harm or rape, which she notes are crimes. She warns, however, that the term may suggest to respondents "such situations as those involving false promises, threats to end the relationship, continual nagging and pressuring, and other verbal strategies to coerce sexual intercourse" (p. 60), which, as she points out, are undesirable but not crimes.

As presented in detail in exhibit 2, there are four ways in which the respondent in an incident report can be counted as having experienced a completed rape. In essence, these are the ways the NCVS uses to operationalize its definition of rape. Even if respondents indicate on the screen question that they have been sexually victimized, they do not count as victims unless they answer one of these four sequences of questions in the incident report in a manner consistent with the NCVS' measurement criteria for rape (e.g., a "don't know" response on a key question may cause an incident not to be counted as a rape victimization).²

First, respondents who answer "yes" to question 24 in exhibit 2 about being hit or attacked may, in response to the subsequent question regarding how they were attacked, state that they were raped. Note that the interviewer asks this as an open-ended question (see footnote a in exhibit 2). Second, respondents may answer "yes" to question 24 but then state that something other than rape

Exhibit 2. Operationalizing completed rape in the NCVS incident report

Initial question	Conditional response	Skip question	Raped ^b	Final type of crime classification
24. Did the offender hit you, knock you down, or actually attack you in any way?	Yes →	29. How were you attacked? Any other way? <i>Mark (X) all that apply.</i> ^a	Raped ^b →	Completed rape
<p>^a The interviewer asked this as an open-ended question and then listened to the respondent's reply while marking items on the list. Included were the following: raped, tried to rape, sexual assault other than rape or attempted rape, shot, shot at (but missed), hit with gun held in hand, stabbed/cut with knife/sharp weapon, attempted attack with knife/sharp weapon, hit by object (other than gun) held in hand, hit by thrown object, attempted attack with weapon other than gun/knife/sharp weapon, hit, slapped, knocked down, grabbed, held, tripped, jumped, pushed, and other.</p> <p>^b If raped, the respondent was asked: "Do you mean forced or coerced sexual intercourse?" If no, the interviewer asked the respondent, "What do you mean?"</p>				
24. Did the offender hit you, knock you down, or actually attack you in any way?	Yes →	29. How were you attacked? Any other way? <i>Mark (X) all that apply.</i>	31. What were the injuries you suffered, if any? Anything else? <i>Mark (X) all that apply.</i> ^c	Completed rape
<p>^c The interviewer asked this as an open-ended question and then listened to the respondent's reply while marking items on the list. Included were the following: none, raped, attempted rape, sexual assault other than rape or attempted rape, knife or stab wounds, gun shot, bullet wounds, broken bones or teeth knocked out, internal injuries, knocked unconscious, bruises, black eyes, cuts, scratches, swelling, chipped teeth, or other.</p> <p>^d If raped, the respondent was asked: "Do you mean forced or coerced sexual intercourse?" If no, the interviewer asked the respondent, "What do you mean?"</p>				

Exhibit 2 (continued)

Initial question	Conditional response	Skip question	Final type of crime classification
24. Did the offender hit you, knock you down, or actually attack you in any way?	No →	27. What actually happened? Anything else? <i>Mark (X) all that apply.</i>	Completed rape
25. Did the offender TRY to attack you?	→	<i>Unwanted sexual contact with force</i> →	
26. Did the offender THREATEN you with harm in any way?	→		
<p>^e The interviewer asked as an open-ended question and then listened to the respondent's reply while marking items on the list. Included were the following: something taken without permission, attempted or threatened to take something, harassed, argument, abusive language, unwanted sexual contact with force (grabbing, fondling, etc.), unwanted sexual contact without force (grabbing, fondling, etc.), forcible entry or attempted forcible entry of house/apartment, forcible entry or attempted forcible entry of car, damaged or destroyed property, attempted or threatened to damage or destroy property, or other.</p> <p>^f If the respondent said "yes," the interviewer was instructed to go back to question 24 and change it to a "yes" response, delete entries for questions 25, 26, and 27, and then proceed accordingly (see row 1).</p>			
24. Did the offender hit you, knock you down, or actually attack you in any way?	No →	28b. How were you threatened? Any other way? <i>Mark (X) all that apply.</i>	Completed rape
25. Did the offender TRY to attack you?	No →		
26. Did the offender THREATEN you with harm in any way?	Yes →		
<p>^g If the respondent said "yes," the interviewer was instructed to go back to question 24 and change it to a "yes" response, delete entries for questions 25, 26, and 28, and then proceed accordingly (see row 1).</p>			

occurred when they were attacked. However, when asked whether they suffered any injuries, they may note that they were raped.

Third, if a respondent says “no” to the initial three bodily harm questions (questions 24, 25, and 26 in exhibit 2), she or he is then asked what actually happened. If the respondent mentions “unwanted sexual contact with force,” then the interviewer asks, “Do you mean forced or coerced sexual intercourse including attempts?” If the respondent answers “yes,” then the interviewer changes the “no” response to a “yes” for the question, “Did the offender hit you, knock you down, or actually attack you in any way?” The interviewer then proceeds as in the first example of how rape is operationalized. Fourth, the interviewer employs a similar process in the last means of operationalizing completed rape if the respondent says that she or he was threatened with “unwanted sexual contact with force.” Note that in these question sequences, it is assumed that there is a shared definition among the respondents about the word “intercourse,” and that the respondents know the kinds of penetration this word covers. Koss (1992, 73) points out, however, that it is “unknown whether women who have had forms of unwanted penetration other than vaginal, and whether men who have been sodomized, will respond to this wording.” Her point is well taken; well-designed experiments will be needed to further examine the possibility Koss raises.

In addition to completed rape, the NCVS includes measures of attempted rape and verbal threats of rape. There are six ways to operationalize attempted rape and two ways to operationalize verbal threat of rape. Recall that the BJS definition of attempted rape includes verbal threats of rape. Note that few other studies include verbal threats when counting attempted rapes (see, e.g., Fisher and Cullen 1998; Koss, Gidycz, and Wisniewski 1987).

Four of the six ways to operationalize attempted rape use the identical series of questions employed to operationalize completed rape, with the exception that the respondent either (1) answers “yes” that an attack was tried or physical harm threatened, and/or (2) volunteers that a rape was attempted. There are also two questions that explicitly ask if the offender tried to attack the respondent (yes or no response) (see NCVS question 25) and how the offender tried to attack the respondent (open-ended question) (see NCVS question 28a). In the latter question, if the respondent says she or he was verbally threatened with rape and some other type of bodily harm, then the incident is classified as an attempted rape. Similarly, the series of questions that begins with asking the “threaten you with harm in any way” question (see NCVS question 26) follows a similar line of questioning and responses for an attempted rape classification.

Verbal threats of rape follow the same question series as those for attempted rape. If the respondent says “verbal threat of rape” when asked how the offender tried to attack or threatened the respondent, the incident is classified as a verbal threat of rape. In the NCVS, these are counted under the “attempted rape” estimates.

Operationalizing sexual assault

In the redesign of the NCS, sexual assault was incorporated into the NCVS to measure types of sexual victimization other than rape—an important methodological change. Indeed, many researchers have argued and empirically documented that sexual victimization is not unidimensional and limited to rape but is multidimensional, covering a variety of types of sexual transgressions (see Fisher and Cullen 1998). In any event, consistent with the statutory reforms of this period, sexual assault is defined as:

a wide range of victimizations, separate from rape or attempted rape. These crimes include attacks or attempted attacks generally involving unwanted sexual contact between the victim and offender. Sexual assaults may or may not involve force and include such things as grabbing or fondling. Sexual assault also includes verbal threats. (U.S. DOJ, BJS 1997, 149)

Five types of sexual assault are operationalized in the NCVS: (1) sexual attack with serious assault, (2) sexual attack with minor assault, (3) sexual assault without injury, (4) unwanted sexual contact without force, and (5) verbal threats of sexual assault other than rape.

The respondent is asked the same series of questions as in the rape sequence of questions. The differences are in the respondent’s descriptions of how the offender attacked, tried to attack, or threatened to attack the respondent. How sexual assault is operationalized is a bit more complex than how completed rape, attempted rape, or verbal threats of rape are operationalized. Therefore, more clarifications in the line of questioning are needed to ensure that a sexual assault is being measured and not a rape. For example, if the respondent indicates that the offender tried to attack her or him with unwanted sexual contact with force (e.g., grabbing, fondling), the interviewer is instructed to ask the respondent, “Do you mean forced or coerced sexual intercourse including attempts?” If the respondent says “yes,” then the interviewer changes the “did the offender hit you” question to “yes” and proceeds accordingly to the “how were you attacked” question. Here the respondent can describe a rape, attempted rape, sexual assault other than a rape or attempted rape, or anything else that happened.

Comparison of the NCS and the NCVS victimization estimates

The introduction of the changes noted above produced substantial differences in BJS victimization estimates from the NCS and the NCVS data. Split sample tests of the two designs produced a 40-percent increase in the overall level of victimizations reported in the NCVS as compared with those reported in the NCS (Lynch 1996a). More relevant to our concerns are whether changes in the definition of rape and in how it is operationalized affected the estimates of rape victimization in the NCVS. Recall that a criticism of the NCS was that it substantially underestimated rape victimization.

First, Taylor and Rand (1995) reported that the revisions resulted in the NCVS-method rates that were 323 percent higher for completed rape and 96 percent higher for attempted rape than the NCS-method rates. Second, the split test results also revealed a 250-percent increase in rape estimates using the NCVS methods compared with the NCS methods (Lynch 1996a). Caution should be exercised, however, when attempting to compare the old estimates of rape with the new ones, even when adjustments have been made to the data that take into account these methodological improvements. Anomalies in the distribution of male and female victims in the 1992 NCS rape estimates raise questions about the adjustments of rape estimates (see Rand, Lynch, and Cantor

1997). Third, the redesign estimates show that the number of completed rapes exceeds the number of attempted rapes. For example, in 1997, the NCVS estimates that there were 115,000 completed rapes and 79,000 attempted rapes. This pattern was not evident in the NCS; with the exception of one year (1988), the number of attempted rapes was always greater than the number of completed rapes.

The incident report has the major advantage of asking detailed questions that might count victimization incidents that other methods would omit or omit victimization incidents that other methods would erroneously include.

The NCVS and beyond

Several major strengths characterize the NCVS: a large national sample; panel design; the possibility of comparing rates of sexual victimization over time and to other forms of criminal victimization; and a clearly bounded reference period. With regard to measuring the sexual victimization of women, however, critical methodological issues will continue to surround the potential biases in the use of the screen question-incident report format. The NCVS screen questions are a marked improvement over those used

to measure rape in the NCS. Even so, as we will see later, other researchers have argued that more detailed and more graphically worded questions would be more likely to elicit reports of sexual victimization from women who had, in fact, been victimized.

The use of the incident report—much understudied by scholars in this area—is a source of potential methodological rigor and bias. The incident report has the major advantage of asking detailed questions that might allow the NCVS (1) to count victimization incidents that other methods would omit or (2) not to count victimization incidents that other methods would erroneously include. Alternatively, the incident report might also result in women who had been victimized being deleted from the victimization count because they failed to answer questions in a way that conformed with the NCVS' coding scheme. We will return to these issues in the pages ahead.

The Specially Designed Sexual Victimization Survey: Koss' Contribution

The NCVS (and its predecessor, the NCS) was designed to measure criminal victimization generally in the United States, of which sexual victimization was but one form. Given its broader mandate, the NCVS may have limited ability to measure rape and other forms of sexual victimization. An alternative strategy would be to develop victimization surveys that focus exclusively on women's sexual victimization. By exploring such victimization in both more depth and breadth, it might be possible to arrive at a more accurate empirical assessment of the extent to which women are sexually victimized in the United States.

Early attempts at such surveys date back to at least the 1950s. Research by Kirkpatrick and Kanin (1957; Kanin 1957), for example, attempted to define and empirically measure "erotic aggressiveness" or "erotic offensiveness" by males against females in dating-courtship relationships on a university campus. Their methods and sampling design are still used today in sexual victimization research. They developed and distributed a self-report "schedule" to female students enrolled in one of 22 "varied" university classes during the academic year (September 1954 to May 1955). Their questionnaire distinguished five degrees of erotic aggressiveness: attempts at (1) "necking," (2) "petting above the waist," (3) "petting below the waist," (4) "sex intercourse," and (5) "sex intercourse

The uses of specially designed surveys generally reveal that sexual victimization is not rare and is more widespread than that found by official statistics and by the NCS/NCVS.

with violence or threats of violence.” The questions focused on the extent to which the respondents were “offended” by intimacy level, frequency, and number of men during the academic year (Kirkpatrick and Kanin 1957, 53). In essence, this was a victimization survey. Among the 291 female students, they found that a large proportion had experienced a sexual victimization. During the academic year, 55.7 percent of women reported being offended at least once at some level of erotic intimacy, with 6.2 percent stating that they had been subjected to “aggressively forceful attempts at sex intercourse in the course of which menacing threats or coercive infliction of physical pain were employed” (p. 53).

Kirkpatrick and Kanin’s research did not trigger a movement to study women’s sexual victimization. In fact, their work was largely neglected until rediscovered two decades later when, sensitized to females’ victimization by a changed social context, scholars returned to this topic. Still, their research is important in showing a finding that would tend to be repeated in later studies: the uses of specially designed surveys generally reveal that sexual victimization is not rare and is more widespread than that found by official statistics and by the NCS/NCVS.

In this regard, the most influential instrument constructed to measure sexual victimization is Koss’ “Sexual Experiences Survey,” which was developed in the 1980s and used extensively by subsequent researchers. This section will focus on the nature and criticisms of this survey. First, however, we will examine the work of Russell (1982), whose early research also introduced important methodological considerations.

Russell’s sexual victimization survey

Much of the sexual victimization research—including Koss’ pathbreaking study—has been conducted using college student samples, in part because of their convenience and in part because this is a social domain in which such victimization is elevated. In contrast, Russell (1982) randomly selected 930 adult female residents in San Francisco from a probability sample of households. Sixty-four percent of the original sample of 2,000 completed the interview. Sensitive to the possible effects of the gender of the interviewer, Russell employed professionally trained female interviewers; their race and ethnicity were matched to that of each respondent. Whenever possible, she had them interview selected respondents in person and in a private setting. The interviews were conducted during the summer of 1978.

Several features of this study are noteworthy. First, previous research had provided respondents, if at all, with only a brief or ambiguous definition of rape. In contrast, Russell’s definition of rape was patterned after the legal definition

of extramarital rape in California as “forced intercourse (e.g., penile-vaginal penetration) or intercourse obtained by threat of force, or intercourse completed when a woman was drugged, unconscious, asleep, or otherwise totally helpless and hence unable to consent” (1982, 84). Second, as presented in exhibit 3, she then operationalized rape using several “behaviorally specific” questions with respect to rape (e.g., “38 questions on sexual assault and abuse,” p. 85). A “behaviorally specific” question is one that does not ask simply if a respondent “had been raped” but rather describes a victimization incident in graphic language that covers the elements of a criminal offense (e.g., someone “physically forces you . . . to have sexual intercourse”). Notably, researchers have found that multiple, behaviorally specific questions are associated with greater disclosure by respondents about being sexually victimized (see Crowell and Burgess 1996, 35). Third, for every episode of rape and attempted rape elicited, the interviewer administered a separate questionnaire. Included was a “description of the assault sufficiently detailed to ensure that one of the criteria

Exhibit 3. Examples of Russell’s questions used to elicit experiences of rape or attempted rape

Rape or attempted rape questions^a
<p>1. Did a _____^b ever physically force you, or try to force you, to have any kind of sexual intercourse (besides anyone you’ve already mentioned)?</p> <p>2. Have you ever had any unwanted sexual experience, including kissing, petting, or intercourse with a _____^b because you felt physically threatened (besides anyone you’ve already mentioned)? IF YES: Did [^c] (any of them) either try or succeed in having any kind of sexual intercourse with you?</p> <p>3. Have you ever had any kind of unwanted sexual experience with a _____^b because you were asleep, unconscious, drugged, or in some other way helpless (besides anyone you’ve already mentioned)? IF YES: Did [^c] (any of them) either try or succeed in having any kind of sexual intercourse with you?</p> <p>4. At any time in your life, have you ever been the victim of a rape or attempted rape? _____</p>
<p>^a Russell provided the wording only for these 4 of her 38 questions.</p> <p>^b The interviewers asked the respondents these questions three times: first about strangers, second about acquaintances or friends, and third about dates, lovers, or ex-lovers.</p> <p>^c Russell used the pronoun “he” here because she had already asked the respondents about any unwanted sexual experiences with females.</p>

for defining the assault as a rape or attempted rape had been met" (Russell 1982, 86). Fourth, for the first three questions in exhibit 3, she asked if they had been perpetrated by (1) strangers; (2) acquaintances or friends; and (3) dates, lovers, or ex-lovers.

Russell's development and use of behaviorally specific questions based on the legal criteria for rape set a new standard for the operationalization of rape—one that the best of subsequent research would build on. Her approach potentially reduced measurement error inherent in previous studies. Thus, the use of a legally based definition of rape meant that she was likely to have assessed victimizations that would qualify legally as a rape. The use of behaviorally defined questions both increased the likelihood that respondents would be cued to victimization incidents that had transpired and diminished the likelihood that respondents would "read into" and thus differentially interpret the victimization questions they were asked to answer. Further, similar to the NCS and the NCVS, Russell suggested the importance of using followup questions to further explore or to "confirm" responses to initial questions about sexual victimization experiences, thus minimizing the possibility of counting as rape incidents those that did not qualify legally for this categorization. Finally, by asking about victimizations perpetrated not only by strangers but also by intimates, she potentially cued respondents to include "acquaintance" and "date" rapes that might otherwise have gone unreported to the interviewer.

The choice of definition of rape, which has played a large role in the exchange between the feminist scholars and their critics, has a significant impact on estimates that are reached regarding the magnitude of rape estimates.

All these factors—the number of questions asked, the manner in which they were presented, and her followup questions—likely contributed to Russell's reported rape estimates (1982, 85). She found that 41 percent of the women reported at least one completed or attempted extramarital rape during their lifetime. Over the past year (12 months prior to interview), 3 percent of the women reported that they had experienced a completed rape or attempted rape. Further, Russell was among the first researchers to compare her survey results with those reported in the UCR and the NCS and to question why statistical discrepancies existed. She tried, for example, to make her incidence rape rates as comparable as she could to the UCR and the NCS rates. She reported that her rates were higher than both the UCR and the NCS. Although Russell's rape estimation and extrapolation procedures and her response rate have been criti-

cized, this critical line of thinking about government-produced rape estimates helped to give direction to future researchers (see Gilbert 1997, 121–123).

Context for Koss' research

Much of the sexual victimization research through the 1980s and even into the 1990s focused on measuring one type of sexual victimization—rape. Even within this narrow definition of sexual victimization, the most persistent and troubling problems had been the lack of a widely agreed-upon definition of rape. With the proliferation of rape research, this issue became even more salient to estimating the extent to which rape occurs. Some studies provided no definition of rape or any type of sexual victimization that was being measured; others provided a broad definition; and still others rigorously defined rape using criteria derived from legal rape reform statutes (see Kilpatrick et al. 1987; Leidig 1992; Muehlenhard and Linton 1987; Rivera and Regoli 1987). Even within those rape studies that used legally derived definitions, there were discrepancies. Some studies, for example, included rape of a person who was given alcohol or drugs (Koss, Gidycz, and Wisniewski 1987) or intoxicated (Schwartz and Pitts 1995), while others excluded this type of victimization (Kilpatrick, Edmunds, and Seymour 1992). The choice of definition, which has played a large role in the exchange between the feminist scholars and their critics, has a significant impact on estimates that are reached regarding the magnitude of rape estimates. Even when similar definitions are used, variations in estimates across studies can result because of other methodological choices, such as sample composition, reference frame, and question wording (see Lynch 1996a, 1996b; Koss 1992, 1993a, 1996).

Separate from but related to this problem is the lack of a generally agreed-upon operational definition of sexual victimization—one that can be used in research and theory building in this area (see Bachman and Taylor 1994; DeKeseredy 1995; Koss 1996, 1993a, 1993b; Gilbert 1997, 1995). For example, although rape (and other types of sexual victimization) includes the notion of nonconsensual sexual behavior, the operational definitions for rape used by researchers have varied along several dimensions. For example, studies using a single question—no matter how broad it is—obtain lower rates of rape (e.g., Gordon and Riger 1989) than studies that use several questions that ask explicitly about a specific type of behavior (e.g., Russell 1982; Koss, Gidycz, and Wisniewski 1987; see also Crowell and Burgess 1996). Thus, whether a study counts a sexual encounter as a rape is affected by which criteria are used to determine the following: whether the sexual penetration is a rape (e.g., by bodily part, object), whether the sexual encounter has been nonconsensual, and whether an offender actually used force or the threat of force. Measurement issues also arise when the age at which the victimization occurred is used in calculating rape estimates. All these methodological issues underlie who eventually determines when a rape has occurred—the investigator or the survey respondent. In essence, the issues of defining and operationalizing rape and other forms of

sexual victimization are epistemological because they raise critical issues about what we know and how we know it.

To address the lack of a standardized instrument for measuring rape (and thereby address the aforementioned methodological issues) and to measure a broad range of different types of sexual victimization, Koss and associates developed the Sexual Experiences Survey (SES). As we will discuss, her instrument was not without its critics (Gilbert 1995, 1997; Roiphe 1993). Nonetheless, the SES and subsequent modified versions have been used in many sexual victimization studies both inside and outside the United States (DeKeseredy and Schwartz 1998; Lane and Gwartney-Gibbs 1985; Schwartz and Pitts 1995).

Koss' Sexual Experiences Survey instrument

Koss and associates (1982, 1985, 1987) had two goals in their study of sexual victimization. First, they sought to broaden the scope of what was assumed to fall under the category of sexual victimization. Working with Oros and Gidycz, Koss developed a measure of sexual victimization that represented a continuum of sexual victimization, what she refers to as a “dimensional view” where “rape represents an extreme behavior on a continuum with normal male behavior within the culture” (Koss and Oros 1982, 455). Their continuum ranged from intercourse achieved through verbal coercion (e.g., continual arguments and pressure) and threatened force, to intercourse achieved against consent through use of physical force (i.e., rape). Second, they wanted to develop an instrument capable of measuring what Koss refers to as “hidden” rapes—incidents that were not reported to the police but nonetheless met the legal definition of rape (Koss and Oros 1982, 455).

Origins of the SES

Koss and Oros' first attempt at developing a behavioral self-report instrument was called “The Sexual Experiences Survey” (see first column of exhibit 4). It contained 13 yes-no formatted questions about a range of coercive sexual experiences, including completed and attempted rape. Their instrument included, like Russell's survey, a direct question to the women respondents about whether they had ever been raped (see question 13). Out of a sample of 2,016 college women, 23.6 percent reported ever having been raped (questions 10, 11, 12, and 13: n=62, 165, 129, and 120, respectively) and 26.9 percent reported ever having experienced an attempted rape (questions 8 and 9: n=368 and 175, respectively).

Exhibit 4. Versions of Koss' Sexual Experiences Survey

Sexual Experiences Survey: Original version (Koss and Oros 1982) ^a	Sexual Experiences Survey: Revised version (Koss, Gidycz, and Wisniewski 1987) ^b
<p>Have you ever:</p> <ol style="list-style-type: none"> 1. Had sexual intercourse with a man when you both wanted to? 2. Had a man misinterpret the level of sexual intimacy you desired? 3. Been in a situation where a man became so sexually aroused that you felt it was useless to stop him even though you did not want to have sexual intercourse? 4. Had sexual intercourse with a man even though you didn't really want to because he threatened to end your relationship otherwise? 5. Had sexual intercourse with a man when you didn't really want to because you felt pressured by his continual arguments? 6. Found out that a man had obtained sexual intercourse with you by saying things he didn't really mean? 7. Been in a situation where a man used some degree of physical force (twisting your arm, holding you down, etc.) to try to make you engage in kissing or petting when you didn't want to? 8. Been in a situation where a man tried to get sexual intercourse with you when you didn't want to by threatening to use physical force (twisting your arm, holding you down, etc.) if you didn't cooperate, but for various reasons sexual intercourse did not occur? 9. Been in a situation where a man used some degree of physical force (twisting your arm, holding you down, etc.) to try to get you to have sexual intercourse with him when you didn't want to, but for various reasons sexual intercourse did not occur? 10. Had sexual intercourse with a man when you didn't want to because he threatened to use physical force (twisting your arm, holding you down, etc.) if you didn't cooperate? 11. Had sexual intercourse with a man when you didn't want to because he used some degree of physical force (twisting your arm, holding you down, etc.)? 12. Been in a situation where a man obtained sexual acts with you such as anal or oral intercourse when you didn't want to by using threats or physical force (twisting your arm, holding you down, etc.)? 13. Have you ever been raped? 	<ol style="list-style-type: none"> 1. Have you given in to sex play (fondling, kissing, or petting, but not intercourse) when you didn't want to because you were overwhelmed by a man's continual arguments and pressure? 2. Have you had sex play (fondling, kissing, or petting, but not intercourse) when you didn't want to because a man used his position of authority (boss, teacher, camp counselor, supervisor) to make you? 3. Have you had sex play (fondling, kissing, or petting, but not intercourse) when you didn't want to because a man threatened or used some degree of physical force (twisting your arm, holding you down, etc.) to make you? 4. Have you had a man attempt sexual intercourse (get on top of you, attempt to insert his penis) when you didn't want to by threatening or using some degree of force (twisting your arm, holding you down, etc.), but intercourse <i>did not</i> occur? 5. Have you had a man attempt sexual intercourse (get on top of you, attempt to insert his penis) when you didn't want to by giving you alcohol or drugs, but intercourse <i>did not</i> occur? 6. Have you given in to sexual intercourse when you didn't want to because you were overwhelmed by a man's continual arguments and pressure? 7. Have you had sexual intercourse when you didn't want to because a man used his position of authority (boss, teacher, camp counselor, supervisor) to make you? 8. Have you had sexual intercourse when you didn't want to because a man gave you alcohol or drugs? 9. Have you had sexual intercourse when you didn't want to because a man threatened or used some degree of physical force (twisting your arm, holding you down, etc.) to make you? 10. Have you had sex acts (anal or oral intercourse or penetration by objects other than the penis) when you didn't want to because a man threatened or used some degree of physical force (twisting your arm, holding you down, etc.) to make you?
<p>^a Version administered to females.</p> <p>^b The following experiences are operationalized with the noted items: (1) sexual contact [1, 2, and 3], (2) attempted rape [4 and 5], (3) sexual coercion [6 and 7], and (4) rape [8, 9, and 10].</p>	

The revised SES

A few years later, Koss and Gidycz (1985, 422) reworded the SES questions to increase the clarity of key terms (e.g., sex play), improve communication to the respondents about the legal definition of rape, and measure more “degrees” of sexual victimization. This new version is presented in the second column of exhibit 4. As can be seen, the 10 yes-no formatted questions (compared with 13 questions in the original version) were developed to reflect Koss and Oros’ original continuum of sexual victimization from the most severe form (rape) to no victimization. There are other noteworthy changes in the revised SES. First, the original SES contained a question that directly asked the respondent if she had been raped (number 13), but the revised version does not. Second, the original instrument did not contain either rape or attempted rape questions that include the criterion “because a man gave you alcohol or drugs” (see revised version questions 5 and 8).

The SES made several methodological improvements and contributions to the measurement of sexual victimization. First, similar to Russell’s rape questions but unlike the NCS, Koss, Gidycz, and Wisniewski (1987) embedded a definition of what they were measuring into their questions. As question 1 reads, Koss and associates defined sex play for the respondents as “fondling, kissing, or petting, but not intercourse.” They use colloquial terms like Russell did to stimulate recall in some respondents, but they also raise the measurement standards; they provide definitions of what terms mean so as to minimize measurement error between what they actually want to measure and how the respondent interprets the terms. With respect to the term “sexual intercourse,” Koss and Gidycz move beyond the limited clarification or lack of clarification given to respondents in past studies. What they mean by sexual intercourse is clearly defined within the context of their rape questions.

Second, similar to Russell, who grounded her definition of rape in the California statutes, Koss and Gidycz (1985) define rape according to the Ohio Revised Code (1980). Thus, their rape questions (questions 4, 5, 8, 9, and 10) are explicit as to the legal criteria for rape—type of penetration, force or threat of force, and no consent (see exhibit 4). To operationalize penetration, they use the term “sexual intercourse” and, as can be seen in question 4 that measures attempted rape, they tell the respondent exactly what they mean (“get on top of you, attempt to insert his penis”). In a footnote to their table 3, they wrote, “sexual intercourse was defined as penetration of a woman’s vagina, no matter how slight, by a man’s penis. Ejaculation was not required” (p. 167). It is not clear where in the questionnaire this definition appeared (e.g., the directions or introduction) because it is not explicitly written into any of the 10 questions.

To measure other forms of penetration that the Ohio rape law also encompasses, they ask on the SES about experiencing “sex acts—anal or oral intercourse or penetration by objects other than the penis” (see question 10). Force or threat of force is operationalized as physical force, and examples are provided for the respondents (e.g., “twisting your arm”; see question 9). Lack of consent is defined for the respondent as “when you didn’t want to” (for example, see question 8).

Third, similar to Russell, Koss and Gidycz (1985) use behaviorally specific language to measure a specific type of sexual victimization. This is unlike the NCS, which did not provide respondents with graphic descriptions of sexual victimizations. Instead, the NCS depended on the respondent to interpret and label an incident as a completed or attempted rape using a question that asked about a wide range of types of incidents, of which rape might be one. Wyatt and Peters (1986) have used the expressions “wide funnel” and “inverted funnel” to make the distinction between the two types of questions (as cited in Koss 1993a).

The use of behaviorally specific questions, according to Koss (1993a, 209), attempts “to put before the respondent detailed scenarios for the type of experiences the interviewer seeks to identify.” Within each of their 10 questions, Koss and Gidycz (1985) specify the behaviors involved. For example, in question 4, which measures attempted rape, the question contains descriptive words about the behavior of the offender (e.g., “get on top of you, attempt to insert his penis”) rather than words about the respondent’s state of mind (e.g., thinking that the offender intended to have sexual intercourse). This distinction is important. In addition to providing explicit definitions of critical terms, this form of questioning potentially minimizes the measurement error that may occur because of a discrepancy between the investigator’s and the respondent’s classification of a victimization incident. What the investigator labels as rape and what the respondent labels as rape may differ considerably (see Fisher and Cullen 1999; Koss 1988). For example, Koss reports that nearly three-fourths of college women who met the legal definition for rape failed to use this term as the label for their experiences.

Unlike most surveys at this time, which used a single item to measure rape, Koss and Gidycz (1985) employed three separate questions to measure rape. Each question specifies a different experience that, according to the law at this time, constituted a rape. They also utilized multiple questions to measure the three other types of sexual victimization. Once again, each question presents a different description of the experience to the respondent.

Findings from the SES

Moving beyond the numerous studies of college students undertaken at a single university or a limited number of institutions, Koss, Gidycz, and Wisniewski (1987) conducted a national-level study of college women. They used a two-stage sampling design to choose schools and then students. First, to select schools, they used a cluster sampling design to sample every *x*th cluster, according to the proportion of total enrollment accounted for by the region. Ninety-three colleges and universities were selected; 32 agreed to participate. Second, from these schools, classes were randomly selected into the sample (for the details of the sampling design, see Koss, Gidycz, and Wisniewski 1987, 163–165). The revised SES was part of a 330-question self-report questionnaire (National Survey of Inter-Gender Relationships) administered by post-master's degree psychologists (men and women) to those students who attended the selected classes that day. The response rate was 98.5 percent. The exact field period was not stated, but it was during the 1984–85 academic year.

To measure four types of sexual victimization, Koss and Gidycz created four categories using the following procedures: (1) rape (a yes response to questions 8, 9, or 10 and any lower numbered items); (2) attempted rape (a yes response to questions 4 or 5 but not to any higher items); (3) sexual coercion (a yes response to questions 6 or 7 but not to any higher numbered items); and (4) sexual contact (a yes response to questions 1, 2, or 3 but not to any higher numbered questions) (see exhibit 4). They also created a “not victimized” category (no to all the questions). Using these scoring procedures, the respondents were classified according to the highest degree of sexual victimization that they reported. Two of the types of sexual victimization on the SES are criminal—completed rape and attempted rape—and two are not—sexual coercion and sexual contact. By including all of these types of sexual victimization, they broadened the definition of sexual victimization to include experiences that may not be criminal but nonetheless victimize women.

To estimate the prevalence of sexual victimization, Koss and associates (1987) asked the respondents about their experiences since age 14. To obtain 1-year estimates, they asked respondents about their experiences since the previous academic year, from September to September (for freshmen, this would have been their senior year in high school).

More than half of the college women (53.7 percent) reported some form of sexual victimization since age 14. Just over 14 percent (14.4 percent) of the women had experienced sexual contact, and 11.9 percent had experienced sexual coercion. Most notably, 12.1 percent had experienced attempted rape, and 15.4 percent had been raped. Taken together, these latter two figures meant that

since age 14, more than a quarter (27.5 percent) of the sample had suffered a victimization that met the State of Ohio's legal definition of rape.

In terms of 1-year estimates, Koss and colleagues (1987) reported that 323 women, 10.1 percent of their sample, had experienced an attempted rape, and that 207 women, 6.5 percent of their sample, had been raped. They also calculated the 1-year rate for attempted/completed rape at 166 per 1,000 female students. When examining the number of sexual victimization incidents (as opposed to the number of women victims), the 1-year incident rate for the Koss, Gidycz, and Wisniewski data is 278 attempted/completed rapes per 1,000 female students (167.2 for attempted rape and 110.8 for completed rape) (p. 168).

Like Russell, Koss, Gidycz, and Wisniewski (1987) then took the important step of comparing their results to those published in the NCS. To avoid a biased comparison, they recalculated their data to include only those incidents that met the definition of rape employed by the NCS, which limited rape to penile-vaginal penetration (and excluded acts such as oral and anal intercourse and sexual intercourse made possible by intentionally incapacitating a victim). Even under this more restrictive definition, Koss and associates (1987, 168) concluded that the rape victimization rate computed from their survey was "10–15 times greater than rates that are based on the NCS." This finding was truly startling, for it suggested that rape victimization was extensive and thus a serious social problem.

Methodologically, Koss, Gidycz, and Wisniewski's study raised two important issues. First, it called into question the accuracy of the questions used by the NCS to measure rape. Their data seemed to confirm that the NCS methodology was incapable of prompting numerous rape victims to report their victimizations to interviewers. Second and relatedly, Koss and associates showed that when multiple behaviorally defined questions are used—questions that assess the full domain of acts that qualify as rape under legal statutes—many sexual victimizations that would not be measured by less detailed and rigorous methods are reported by respondents. Underlying their work is the assumption that many women who have been raped do not readily report this fact on victimization surveys unless they have been asked to answer questions that specifically describe a particular kind of behavior. Koss, Gidycz, and Wisniewski thus were most concerned with overcoming the problem, found in previous works, of underestimating the true extent of sexual victimization. As we will see, this concern results in Koss and associates being criticized for using a methodology that may have created the opposite measurement error, of overestimating the extent to which women are sexually victimized.

Gilbert's criticisms of Koss' definition and operationalization of rape

Gilbert (1997), among others, has accused Koss and associates of engaging in "advocacy research," which he defines as "playing fast and loose with the facts in the service of a noble cause" (1992, 7) (see also Roiphe 1993). Gilbert bases his accusations on methodological limitations of her work that he has detailed in several publications (1991, 1992, 1997). His central thesis is that Koss' research has exaggerated the extent of rape, and that her findings have been uncritically accepted because they reinforce feminist notions that entrenched patriarchal relationships in America generate widespread sexual exploitation of women. Gilbert rests his case on two main charges.

First, of the five questions used to measure rape, two involved a man attempting or completing forced intercourse "by giving you alcohol or drugs." Koss and colleagues used this phrasing to operationalize those acts that qualify as rapes under the Ohio Revised Code, which read "for the purpose of preventing resistance the offender substantially impairs the other person's judgment or control by administering any drug or intoxicant to the other person" (as cited in Koss, Gidycz, and Wisniewski 1987, 166). Notably, 44 percent of the rape victims in Koss and colleagues' study were counted as victims because they answered "yes" to these two questions that involved rape accomplished through purposeful intoxication.

Gilbert characterizes these two rape questions as "awkward and vaguely worded" because they lack any notion of the man's intention, how much alcohol the respondent ingested, and whether the alcohol or drugs led the respondent not to offer her consent. For example, what does having sexual intercourse with a man because he "gave you drugs or alcohol" mean? Did he order a beer or wine for the respondent? Was the respondent too intoxicated to consent (Gilbert 1991, 59)? Gilbert goes so far as to suggest that perhaps "the woman was trading sex for drugs, or perhaps a few drinks lowered her inhibitions so that she consented to an act that she later regretted" (1997, 116). He contends that the question could have been worded more clearly to denote "intentional incapacitation of the victim" (1997, 117). The larger point, of course, is that an unknown number of respondents might have answered "yes" to these two items even though their sexual experiences did not qualify legally as a rape (see also Muehlenhard et al. 1994).

Consistent with Gilbert's criticisms, subsequent research using Koss' SES instrument to measure rape and attempted rape in both the United States and Canada has altered the phrase regarding someone giving a woman drugs or alcohol. For example, in their study of rape among college women at Ohio

University, Schwartz and Pitts (1995, 17) changed question 8 to read: "Have you had sexual intercourse when you didn't want to because you were drunk or high?" DeKeseredy and Kelly (1993) used the same wording in their national study of women abuse on college and university campuses.

Second, Gilbert questions more fundamentally whether the questions developed by Koss and associates are, in any methodologically rigorous way, capable of validly measuring sexual victimization. Two troubling anomalies are found in Koss and associates' data. First, nearly three-quarters (73 percent) of the women categorized as rape victims in the study did not, when asked, believe they had been raped (see also Roiphe 1993). Second, about four in ten stated that they subsequently had sexual relations with the person who had purportedly raped them (Gilbert 1997, 116). Gilbert argues that it is highly unlikely that such a large proportion of college-educated women would be so uninformed or sexually inexperienced as (1) to misinterpret when they had, in fact, been raped and/or (2) to become involved again with a "rapist." The more plausible interpretation, contends Gilbert, is that Koss and associates' measure is hopelessly flawed, cueing respondents to answer "yes" to questions measuring rape even though the nature of their sexual experience would not, if examined in detail, qualify legally as a rape.

Koss has offered reasonable rebuttals to Gilbert's criticisms. For example, Koss and Cook (1993) note that even when the two items involving rape due to alcohol and drugs are removed from statistical calculations, the extent of rape in Koss and associates' sample remains disquietingly high (9.3 percent of the sample experiencing, in one year, attempted or completed rape). Further, Koss does not find it so implausible that many women, raised with a limited conception of rape as involving only attacks by strangers, might fail to define forced intercourse by an acquaintance as a rape. It also is possible that women might subsequently have sexual relations with their attacker because they blamed themselves for the previous encounter or because this person again attacked them (see Gilbert 1997).

The heated nature of the debate typically leads scholars to take sides for or against Koss or Gilbert. We suspect that becoming an advocate for a given position obfuscates more than it illuminates. In any case, we would share one broad comment on the Koss-Gilbert exchange. Koss' research must be seen as rigorous and pathbreaking for its time, but it should not be treated as sacrosanct. The challenge is to move beyond the initial state-of-the-art measure—in this instance, Koss' SES—to develop more valid measures of sexual victimization. Gilbert's concerns cannot be dismissed simply because they are conveyed in an inhospitable manner. Much more quantitative and qualitative research is needed to settle the methodological issues raised in the Koss-Gilbert debate,

including how question wording affects the accuracy of respondents' answers and to what extent women who have experienced a legally defined rape comprehend this fact.

Measuring Sexual Victimization in the 1990s

During the 1990s, investigators built on the methodological insights from researchers such as Koss and her associates in the previous decade, from the Koss-Gilbert methodological debates, and from criticisms of the NCS and the redesigned NCVS. Several features characterized the studies undertaken by this "next generation" of investigators. Thus, they (1) broadened the scope of sexual victimization to include sexual coercion and unwanted physical contact, (2) included newly criminalized offenses such as stalking, and (3) used nationally representative samples of women. Perhaps most important, however, these investigators sought to operationalize sexual victimization by using behaviorally specific questions that were consciously grounded in the existing legal definition of rape (or other types of sexual victimization).

Again, this was the approach taken, most notably, by Koss in her SES. The next generation of investigators attempted to improve on Koss' work in two ways. First, some researchers in the 1990s employed questions that were more graphically explicit than those in the SES. The new series of questions not only included a description of the behavior in question but also provided a definition of what behaviors that act in question entailed. Second, some scholars in this "next generation" incorporated into their measures a strength that, as we noted, has marked the NCVS—an "incident report." The incident report not only allows researchers to gather more details about a given victimization incident but also helps to classify the incident as to the *type* of sexual victimization, if any, the respondent had experienced.

Some of the methodological advancements we have identified were incorporated in two major national studies completed during the 1990s that examined sexual victimization among women, which we discuss in the next two sections.

The National Women's Study

According to Lynch (1996a, 1996b), the most frequently cited estimates of the incidence of rape in the United States are drawn from the NCS/NCVS and from the National Women's Study undertaken by Kilpatrick, Edmunds, and Seymour (1992). The National Women's Study (NWS) is a 3-year (1990–92) longitudinal study of a sample of women 18 years and older. The NWS used a

probability sample of 4,008 adult Americans, including a group of 2,000 women ages 18 to 34 who were oversampled. Kilpatrick and colleagues designed three waves of telephone interviews to collect information about the respondents' major mental health problems and alcohol and drug-related problems and consumption. In the first and second waves, they measured forcible rapes that had occurred (1) any time during the respondent's lifetime and (2) within the past 12 months, respectively. Professionally trained female interviewers were employed by the survey firm Schulman, Ronca, and Bucvalas (SRBI) to administer the survey.

With respect to forcible rape, the first wave of interviews was unbounded, but the second wave was bounded by the previous interview. In the first interview, respondents were asked about their lifetime forcible rape experiences. In the second interview, respondents were asked to report their forcible rape experiences for the year since their last interview—that is, the 1-year period between wave one and wave two. In both wave one and wave two, those respondents who in the screen questions reported an incident, skipped into a sequence of questions about the characteristics of the rape incident(s) (e.g., whether they reported the incident to the police; their relationship to the attacker). Responses to these questions were not used to verify what happened but to classify the event; the responses to the screen questions were used to estimate the extent of lifetime and annual rape, respectively (see exhibit 5). The third interview did not contain any questions about forcible rape or any other forms of sexual victimization.

Eighty-five percent of the women contacted participated in wave one. At wave two, 81 percent of the wave one participants (n=3,220) were located and participated in the study. The participation rate at time two was therefore 68.9 percent of the original sample.

Definition and operationalization of forcible rape

Kilpatrick and colleagues (1992) admit to using a “very conservative definition of rape—one which would be legally defined as forcible rape or criminal sexual assault in most States.” They define rape as:

an event that occurred without the woman's consent, involved the use of force or threat of force, and involved sexual penetration of the victim's vagina, mouth, or rectum.

Attempted rape was not covered by this definition; accordingly, the NWS measured and reports data only on completed rape.

Kilpatrick, Edmunds, and Seymour (1992, 15) used four questions to operationalize their definition of rape, which according to them “provide clear answers for the first time to the critical elements of forcible rape: use of force or threat of force, lack of consent; and sexual penetration.” Similar to Koss’ SES instrument, each question is directly worded to describe a narrow behavior. For example, in the NWS questionnaire (question 48), the respondent is asked, “Has a man or a boy ever made you have sex by using force or threatening to harm you or someone close to you?” It is possible that some respondents might find this question to be ambiguous; if so, then measurement error might be introduced because these respondents might potentially overreport or underreport their sexual victimization experiences. To minimize the possibility that respondents would be confused about what kinds of experiences the question covers, Kilpatrick and colleagues use a followup statement to clarify the specific type of behavior they are asking the respondent about. Thus, in question 48, they follow up with a statement that defines what “sex” means: “Just so there is no mistake, by sex we mean putting a penis in your vagina.” A similar followup statement containing an explicit definition of the behavior in question is also incorporated in the oral sex question (see question 49 in exhibit 5).³

Unlike the SES instrument, Kilpatrick and colleagues broaden the criterion of “threat of force” to include threats of harm not only to the respondent but also to “someone close to” the respondent (see question 48 in exhibit 5). This wording is used explicitly only in the question measuring penile-vagina penetration and not in the questions measuring other types of penetration (i.e., oral, anal). Further, there is no explanation why this wording on “threats to someone close to you” is used to operationalize rape but is not mentioned in the definition of rape employed by the NWS researchers (see definition quoted previously).

According to the NWS, 13 percent of the women in the sample reported having experienced a completed rape at least once during their lifetime (Kilpatrick, Edmunds, and Seymour 1992, 2). Less than 1 percent (0.7 percent) of the women surveyed had experienced a completed rape within the past 12 months. Kilpatrick and associates (1992, 2) compare their estimate of the number of women age 18 or older who were raped during a 12-month period—683,000—with the 1990 NCS annual estimate of completed and attempted rapes for females age 12 and older—130,000—and claim that “the *National Women’s Study* estimate was still 5.3 times larger than the NCS estimate.”

This comparison must be interpreted with caution, or at least placed within an appropriate context. First, the NCS and the NWS have different definitions of completed rape. The NCS includes only penile-vaginal penetration, whereas the NWS includes more types of penetration. Second, the NCS’ estimate includes completed and attempted rapes, whereas the NWS is limited to completed

Exhibit 5. NWS questions used to elicit experiences of completed rape

Completed rape questions^a

Another type of stressful event that many women experience is unwanted sexual advances.^b Women do not always report such experiences to police or discuss them with friends or family. The person making the advances isn't always a stranger, but can be a friend, boyfriend, or even a family member. Such experiences can occur any time in a woman's life—even as a child. Regardless of how long ago it happened or who made the advances:

48.^c Has a man or boy ever made you have sex by using force or threatening to harm you or someone close to you? Just so there is no mistake, by sex we mean putting a penis in your vagina.

49. Has anyone ever made you have oral sex by using force or threat of harm? Just so there is no mistake, by oral sex, we mean that a man or boy put his penis in your mouth or somebody^d penetrated your vagina or anus with his mouth or tongue.

50. Has anyone ever made you have anal sex by force or threat of harm?^e

51. Has anyone ever put fingers or objects in your vagina or anus against your will by using force or threat?

52.^f During your lifetime, how many times (different occasions) have you been forced to have (sex/oral sex/anal sex) or been forcibly penetrated with fingers or objects? Please include any incidents that may have happened when you were a child.

53. Did this (any of these incidents) occur before you were 18 years old?

54. Did this incident (any of these incidents) occur within the past 12 months or since the last time you were interviewed?

^a Unless noted, the wording of the questions came from the appendix of *Rape in America: A Report to the Nation* (Kilpatrick, Edmunds, and Seymour 1992).

^b This sentence is taken from Lynch (1996b, 139).

^c Question numbers were taken from Lynch (1996b, appendix).

^d Lynch's (1996b) version reads ". . . or someone, male or female. . . ."

^e Lynch's (1996b) version reads "[J]ust so there is no mistake, by anal sex we mean that a man or boy put his penis in your anus."

^f Questions 52 through 54 come from Lynch (1996b).

rapes. (Note, though, that the NCS does divide its rape counts into completed (n=60,710) and attempted (n=63,760); 1990 figures cited here.) Third, the NWS includes females age 18 and older, whereas the NCS includes females age 12 and older. The NCS' results cannot be easily aggregated to match the sample of those 18 years and older because published reports on the NCS use the noncomparable age category of 16 to 19 years old. Note, however, that the NCS data tapes are publicly available, and thus comparable estimates by age could potentially be calculated from these data. Fourth, the NCS estimates cited in the *Rape in America* report include males and females; the NWS includes only females. Fifth, as Lynch (1996a, 1996b) points out, comparing estimates may be confounded because the NWS and the NCS use different procedures for bounding the victimization reference period.

Methodological lessons learned

What methodological advances in measuring rape did the NWS introduce or emphasize? First, moving beyond the many case studies of college women (i.e., surveys of women attending one or two universities), the NWS used a nationally representative sample of adult women. As Koss (1993b, 1063) points out, the sample did exclude "several potentially high-risk groups for rape," for example, women living in college residences and women serving in the military. When the NWS was undertaken, however, it was the only national-level study of women in the general U.S. population other than the NCS. Relatedly, wave two of the NWS, which was used to compute the NWS' annual estimates of rape victimization, was bounded by the wave-one interview. Bounding has been shown to reduce measurement error associated with "telescoping"—that is, of respondents counting as victimizations events that occurred outside the reference period that the survey covered (see Lehen and Skogan 1981, 1984). Because panel studies are rare in this area of research (again, the NCVS being the notable exception), this type of bounding is not found in the extant published literature.

Second, unlike many previous surveys, the investigators furnished the respondents with a clearly worded introduction to their rape questions, which contained several cues (1) as to the possibility that a sexual experience may have occurred even though it had not been disclosed to the police, family, and/or friends and (2) as to whom potential offenders might include (i.e., not "always a stranger, but can be a friend, boyfriend, or even a family member"). This introduction is meant to guard against respondents not reporting rape incidents to interviewers because they did not believe that victimizations unreported to others, or committed by people they knew, "counted" as rape.

Third, a special contribution of the NWS is the wording the researchers developed in the four questions used to measure completed rape. They attempted to employ wording that described the specific behavior in question in very graphic detail (e.g., “by oral sex, we mean that a man or a boy put his penis in your mouth or somebody penetrated your vagina or anus with his mouth or tongue”). These behaviorally specific questions were used to cue the respondent to the particular domain of behavior that was being measured. In this way, the researchers were attempting to minimize measurement error by making sure that respondents would understand what kinds of sexual experiences were covered by the questions being asked.

Despite these advantages, the NWS had three main limitations—issues that later research would address: (1) the definition of rape does not include incidents when the victim was incapacitated, (2) it measured only forcible rape and not attempted rape or other forms of sexual victimization, and (3) it did not employ an incident report to “check” or validate whether respondents who answered “yes” to the rape questions should be counted as rape victims.

The National Violence Against Women Survey

Building on the NWS, Tjaden and Thoennes (1998a) designed and executed the National Violence Against Women (NVAW) Survey, introducing it to respondents as a survey on personal safety. The NVAW Survey includes questions about general fear of violence and about incidents of actual or threatened violence experienced during the respondent’s lifetime and annually by different types of perpetrators. Sixty types of perpetrators were assessed (e.g., specific parent, spouse, specific ex-spouse, specific partner, cousin). Specific types of violence included sexual assault (i.e., rape and attempted rape), physical assault (i.e., slapping, getting beat up, using a gun on the victim), and stalking.

The NVAW Survey is a nationally representative sample of 8,000 English-speaking and Spanish-speaking women 18 years of age and older who reside in households throughout the United States (Tjaden 1996, 1). Using random-digit dial within U.S. Census regions to draw the sample, eligible women in each household were identified (see Tjaden and Thoennes 1998a, 14). If, for example, more than one woman was eligible, a designated respondent was randomly selected using the “most recent birthday method” (see Tjaden 1996, 2). Using CATI, professionally trained female interviewers employed by SRBI pretested the survey and administered the NVAW Survey from November 1995 to May 1996. The NVAW Survey’s participation rate was 72 percent (see Tjaden 1996, 3–4).

Definitions and operationalization of rape

Tjaden and Thoennes (1998a, 13) defined rape as:

an event that occurred without the victim's consent, that involved the use or threat of force to penetrate the victim's vagina or anus by penis, tongue, fingers, or objects, or the victim's mouth by penis. The definition included both attempted and completed rape.

Exhibit 6 shows how the NVAW Survey operationalized rape. Note that unlike the NWS, the NVAW Survey measures both completed rape and attempted rape.

The influence of the NWS is evident in the questions used by Tjaden and Thoennes. First, both surveys begin the questioning about rape with an introduction for the respondents as to the nature of the questions that follow. Second, both studies use a lifetime reference frame of "regardless of how long ago it happened" to cue respondents. Third, the questions used in the NWS and the questions used in the NVAW Survey provide the respondent with a behaviorally specific definition.⁴ For example, in question F1 in exhibit 6, the question asks the respondent if "a man or boy ever made you have sex by using force or threatening to harm you or someone close to you?" Fourth, to minimize any respondent confusion regarding what behavior the question is asking about, the respondent is provided with a definition as to type of behavior in question. For example, question F3 includes the phrase "by anal sex we mean that a man or boy put his penis in your anus" (see Tjaden 1996, F1–F6). According to Tjaden and Thoennes (1998a, 3), "These questions were designed to leave little doubt in the respondents' minds as to the type of information being sought."

By contrast, differences in the wording of the screening questions used in the NVAW Survey and the NCVS are striking. As we have discussed, the NVAW Survey uses five behaviorally defined screen questions to prompt respondents to report sexual experiences that meet the legal definition of rape used in many States. This approach thus relies on multiple questions, all narrowly worded, to ask about only one type of crime—rape. The NCVS, however, uses a more general approach in its screen questions. Thus, the screen questions employ general terms—such as asking respondents whether they have experienced "forced or unwanted sexual acts"—to capture a wide range of sexual victimizations, of which rape is but one type. The word rape is also used in the NCVS screen questions but not in the NVAW Survey. Finally, the NCVS classifies incidents based on responses to the questions in the *incident* report and not on responses to the screen questions (see Bachman 1998). As we will discuss, the NVAW Survey uses screen questions, not its perpetrator report, to count whether a sexual experience has been a completed rape, attempted rape, or other type of victimization.

Perpetrator report for rape

After the NVAW Survey interviewer asks the five questions presented in exhibit 6, she then asks the respondent questions concerning the type of perpetrator(s). Unlike the NCVS, which has the interviewers administer a separate incident-level report for the number of times the respondent indicates the incident happened, the NVAW Survey has the interviewers administer a detailed sexual assault report for each type of perpetrator. Within this report, the respondent is asked on how many different occasions the specific perpetrator forced or tried to force the respondent to have sex or forcibly penetrate the respondent with

Exhibit 6. NVAW Survey questions used to elicit experiences of rape and attempted rape

Rape and attempted rape questions*

F1. We are particularly interested in learning about violence women experience, either by strangers, friends, relatives, or even by husbands and partners. I'm going to ask you some questions about unwanted sexual experiences you may have had either as an adult or a child. You may find the questions disturbing, but it is important we ask them this way so that everyone is clear about what we mean. Remember the information you provide is confidential.

Regardless of how long ago it happened, has a man or boy ever made you have sex by using force or threatening to harm you or someone close to you? Just so there is no mistake, by sex we mean putting a penis in your vagina.

F2. Has anyone, male or female, ever made you have oral sex by using force or threat of force? Just so there is no mistake, by oral sex we mean that a man or boy put his penis in your mouth or someone, male or female, penetrated your vagina or anus with their mouth.

F3. Has anyone ever made you have anal sex by using force or threat of harm? Just so there is no mistake, by anal sex we mean that a man or boy put his penis in your anus.

F4. Has anyone, male or female, ever put fingers or objects in your vagina or anus against your will by using force or threats?

F5. Has anyone, male or female, ever **attempted** to make you have vaginal, oral, or anal sex against your will, but intercourse or penetration did not occur?

* These questions were asked only of the female respondents.

his/her fingers or other object; the respondent is then asked when this incident happened. In the case of a single incident, the respondent is asked when this incident happened with respect to the number of years ago or in the past 12 months. If there is more than one incident, then the respondent is asked when was the first time this incident happened and when was the most recent time it happened. Once again, the response is with respect to the number of years ago or in the past 12 months (see Tjaden 1996, J1–J2). Unlike the NWS and the NCVS, the NVAW Survey is not bound by a previous interview because it is a cross-sectional design.

Classifying victimizations

In the NVAW Survey, the interviewers ask about four types of offenses—each with their own set of screen questions. For rape, the questions in exhibit 6 are used to determine whether the respondent experienced a completed rape or an attempted rape. The perpetrator report is used to determine when the rape occurred, not if a rape occurred. If a respondent answers “yes” to any of these questions, she is then “counted” as a rape victim. However, Tjaden and Thoennes do provide separate estimates for completed rape and attempted rape (see following discussion).

Thus, the NVAW Survey assumes that coherent sets of questions that cover a given domain of conduct (e.g., rape) and that these questions worded in a behaviorally specific way will yield accurate responses from women as to whether they have been sexually victimized. In contrast, the NCVS assumes that the main purpose of such questions—which it calls “screen questions”—is to cue the respondent to recall that she had experienced some type of sexual victimization during the reference period. The NCVS then assumes that the set of questions in the incident report is needed to probe more carefully the detailed nature of the victimization experience (“what actually happened”). Accordingly, the NCVS uses the second set of questions to classify victimizations.

Comparison of rape results

Tjaden and Thoennes results revealed that 17.6 percent of the surveyed women reported having experienced a completed (14.8 percent) or attempted rape (2.8 percent) during their lifetime. The NVAW Survey’s estimate of completed rape is only slightly higher than the NWS’ lifetime estimate of 13 percent. The NVAW Survey also found that 0.3 percent of the women reported experiencing a completed or attempted rape in the previous 12 months.

Tjaden and Thoennes (1998a, 4) note that it is difficult to make direct comparisons between their estimates and the NCVS rape estimates. Bachman (1998,

15–16), however, attempted to address this issue by disaggregating the NCVS data to create rape estimates that would be more comparable to the NVAW Survey. Thus, she compares annual rape estimates for women 18 years and older who were raped by a lone offender. She reports that the NVAW Survey's rape estimate of 0.35 victims per 100 women is higher than the NCVS estimate of 0.16 victims per 100 women. She argues that "this difference underscores the very sensitive nature of estimation procedures and how slightly different methodological procedures can result in quite diverse estimates" (p. 16).

Making comparisons between the NWS and the NVAW Survey is also complicated. As noted, these studies share many features: (1) moving beyond the common reliance on college student samples, both used a nationally representative sample of women age 18 and older; (2) they employed the same survey firm (SRBI) whose professionally trained interviewers used CATI; (3) both used a lead-in introduction to the screen questions to alert respondents that victimizations could involve incidents unreported to the police and committed by intimates; (4) their screen questions contained similar wording and both are behaviorally specific about the types of experiences in question; and (5) both used responses to their respective screen questions and timeframe questions to determine lifetime and annual estimates. Despite these similarities, the NWS asked questions only about completed rape, not about attempted rape, whereas the NVAW Survey measured both completed rape and attempted rape. Of course, it might be possible to compare the NWS and the NVAW Survey figures for completed rapes. However, Tjaden and Thoennes (1998a) do not report annual estimates for completed and attempted rape separately. As a result, at this time, no comparison can be made to Kilpatrick, Edmunds, and Seymour's (1992) NWS annual estimate for completed rape.

Stalking

In the NVAW Survey, Tjaden and Thoennes also collected data on a form of victimization that has earned increasing public and legal attention: stalking. To date, only two national-level studies do so: the NVAW Survey and Fisher and Cullen's (1998) college women study (see discussion following). Building their definition of stalking from the model antistalking codes for States developed by the National Institute of Justice, Tjaden and Thoennes (1998b, 2–3) defined stalking as:

a course of conduct directed at a specific person that involves a repeated visual or physical proximity, nonconsensual communication, or verbal, written or implied threats, or a combination thereof, that would cause a reasonable person fear with repeated meeting on two or more occasions.

They clarify this definition by stating that “the NVAW Survey does not require stalkers to make a credible threat against the victim, but it does require victims to feel a high level of fear.”

Space precludes a detailed review of how stalking is operationalized by Tjaden and Thoennes (1998b, 17). Still, we can note that in the NVAW Survey, 8.1 percent of the women reported being stalked at some time in their life; only 1 percent of the women reported being stalked in the previous 12 months. Tjaden and Thoennes’ (1998b) estimates of stalking, however, are dependent on what level of fear they use to “count” a respondent as a victim. Unlike other criminal offenses, legal statutes often assert that stalking is a crime only if it induces fear of being harmed in a reasonable person.

The question that emerges is what it means legally to say that someone is “fearful” and, in turn, how this concept should be measured in victimization surveys. Tjaden and Thoennes made the decision to use a stringent criterion, counting as stalking victims only those women who said that stalking behavior had made them feel “very” frightened. However, if the standard is lowered to include women who said they felt “somewhat” or “a little” frightened, then the prevalence of stalking victimization in the NVAW sample rises markedly. Thus, the lifetime estimate increases from 8.1 to 12 percent, and the estimate for the past 12 months jumps from 1 to 6 percent. These results again reveal the challenge of measuring the victimization of women and how methodological decisions can affect the estimates that researchers produce.

The National College Women Sexual Victimization Study: Assessing Screen Question-Incident Report Methodology

As we have reviewed, two very different methods have evolved over the past two decades for measuring different types of sexual victimization. First, Koss’ SES, Kilpatrick’s NWS, and Tjaden and Thoennes’ NVAW Survey all used behaviorally specific questions both to cue respondents to disclose their sexual victimization and, if so, to classify victims as to the type of victimization experienced. Second, the NCVS used an incident report to classify incidents as to the type of crime, if any, that the victim experienced. The strength of the behaviorally defined questions is that the respondent is provided with descriptive cues within a scenario framework. Using *only* behaviorally specific questions, however, assumes both that the respondent understands the experience she is being asked about and that these questions are able to cue accurate recall by the respondents (i.e., a rape question cues all rape victims to answer yes; an attempted rape question cues all attempted rape victims to

answer yes; and so on). These assumptions are problematic, especially in the absence of followup questions to probe in detail what actually transpired in any given incident.

Herein lies the strength of the NCVS and its use of an incident report. As discussed, the NCVS' measurement process includes screen questions about the criminal act followed by a series of short cues. Each victimization the respondent reports (answers "yes" to) is then followed up with a detailed incident report that contains multiple questions about what occurred during the incident. The responses to these questions are then used to classify the type of victimization that occurred.

In our national study of sexual victimization among college women, we combined these two approaches to operationalize rape and other types of sexual victimization (see Fisher and Cullen 1998). From Koss and similar researchers, we borrowed the idea to include in the design a series of behaviorally specific, graphically worded cueing or screen questions across a range of sexual victimizations. From the NCVS, we incorporated into the design the screen question-incident report method. The result is a method that measures sexual victimization by cueing and screening potential victims with behaviorally defined questions and then classifies the type of victimization, if any, through a detailed incident report.

Overview of the National College Women Sexual Victimization Study

We designed the National College Women Sexual Victimization (NCWSV) Study to estimate the extent of different forms of sexual victimization among college women and to examine risk factors associated with such victimizations. We collected sexual victimization data from a random sample of female undergraduate and graduate college students during the 1996–97 academic year. A total of 4,446 college women enrolled at 233 2-year and 4-year schools were selected using a two-stage probability sampling design (see Fisher and Cullen 1998, ch. 2).

Approximately 2 weeks before a respondent was called, she was sent a cover letter that explained the nature of the study and its procedures (e.g., telephone call from a female interviewer, an 800 number and an e-mail address to contact for more information, voluntary participation, confidentiality). The cover letter was clear that the intent of the study was to examine the extent and nature of unwanted sexual victimizations. Like the NWS (Kilpatrick, Edmunds, and Seymour 1992) and the NVAW Survey (Tjaden and Thoennes 1998a, 1998b), we contracted with SRBI to administer our surveys using CATI; they employed

professionally trained female interviewers. The field period began in late February and ended in early May 1997. The response rate was 84.6 percent. Before undertaking the survey, the content and flow of the cover letter, the introduction to the survey, and the screen questions were discussed during two focus groups conducted at the University of Cincinnati. We also pretested the survey instruments in spring of 1996 with a random sample of 100 female students enrolled at the University of Cincinnati.

Defining rape

Broadening the types of sexual victimization that previous researchers had assessed (cf. Koss, Gidycz, and Wisniewski 1987; Muehlenhard and Linton 1987), we measured 12 different forms of sexual victimization that included rape, sexual coercion, unwanted sexual contact, various threats, and stalking. Following the tradition of grounding our definition of rape in legal statutes, we defined rape to include unwanted penetration (completed and attempted) by force or threat of force. Penetration included penile-vaginal, mouth on genitals, mouth on someone else's genitals, penile-anal, digital-vaginal, digital-anal, object-vaginal, and object-anal. As shown in exhibit 7, we used 12 behaviorally specific questions to screen for different types of sexual victimization. All contain the same reference period—"since school began in the fall of 1996." The reference period was approximately a half of a year long, which is similar to that used in the NCVS.

Similar to the NWS (Kilpatrick, Edmunds, and Seymour 1992) and the NVAW Survey (Tjaden and Thoennes 1998a, 1998b), we included an introduction that explained the context of the study, cued respondents as to different situations and various potential perpetrators, and alerted them that graphic language would be used on the survey. The rape questions (questions 7, 8, 9, 10, and 12) were similar, if not identical, to the ones used in the NWS (Kilpatrick, Edmunds, and Seymour 1992) and the NVAW Survey (Tjaden and Thoennes 1998a). Within the first four questions (i.e., the completed rape questions), we provided a graphic description of the behavior in question and a definition of what was meant by each term we used.⁵ We designed the other screen questions to cue respondents to sexual coercion (questions 18, 19, and 20), unwanted sexual contact (questions 14 and 16), and stalking (question 24).

When a respondent said "yes," that she had experienced the type of behavior asked about in the screen question, she was asked which different incidents of this type had happened to her (see footnote 1 in exhibit 8). Similar to the NCVS, for every different incident, the interviewer completed an incident report.

Exhibit 7. NCWSV Study screen questions

Sexual victimization screen questions

Women may experience a wide range of unwanted sexual experiences in college. Women do not always report unwanted sexual experiences to the police or discuss them with family and friends. The person making the advances is not always a stranger, but can be a friend, boyfriend, fellow student, professor, teaching assistant, supervisor, co-worker, somebody you meet off campus, or even a family member. The experience could occur anywhere: on- or off-campus, in your residence, in your place of employment, or in a public place. You could be awake, or you could be asleep, unconscious, drunk, or otherwise incapacitated. Please keep this in mind as you answer the questions.

Now, I'm going to ask you about different types of unwanted sexual experiences you may have experienced since school began in the Fall 1996. Because of the nature of unwanted sexual experiences, the language may seem graphic to you. However, this is the only way to assess accurately whether or not the women in this study have had such experiences. You only have to answer "yes" or "no."

7. Since school began in the Fall 1996, has anyone **made** you have **sexual intercourse** by using **force or threatening to harm** you or someone close to you? Just so there is no mistake, by intercourse I mean putting a penis in your vagina.
8. Since school began in the Fall 1996, has anyone **made** you have **oral sex by force or threat of harm**? By oral sex, I mean did someone's mouth or tongue make contact with your vagina or anus or did your mouth or tongue make contact with someone else's genitals or anus.
9. Since school began in the Fall 1996, has anyone **made** you have **anal sex by force or threat of harm**? By anal sex, I mean putting a penis in your anus or rectum.
10. Since school began in the Fall 1996, has anyone ever used **force or threat of harm to sexually penetrate you with a foreign object**? By this, I mean for example, placing a bottle or finger in your vagina or anus.
12. Since school began in Fall 1996, has anyone **attempted but not succeeded** in making you take part in any of the unwanted sexual experiences that I have just asked you about? *This would include threats that were not followed through.* For example, did anyone threaten or try but not succeed to have vaginal, oral, or anal sex with you or try unsuccessfully to penetrate your vagina or anus with a foreign object or finger?
14. Not counting the types of sexual contact already mentioned, have you experienced any **unwanted or uninvited touching of a sexual nature** since school began in the Fall 1996? This includes forced kissing, touching of private parts, grabbing, and fondling, even if it is over your clothes. Remember this could include anyone from strangers to people you know well. Have any incidents or **unwanted or uninvited touching of a sexual nature** happened to you since school began in the Fall 1996?

continued

Exhibit 7 (continued)

Sexual victimization screen questions

16. Since school began in Fall 1996, has anyone **attempted or threatened but not succeeded in unwanted or uninvited touching of a sexual nature?**

18. I have been asking you about unwanted sexual contact that involved force or threats of force against you or someone else. Sometimes unwanted sexual contact may be attempted using threats of nonphysical punishment, promises of *rewards* if you complied sexually, or simply continual *verbal pressure*. Since school began in Fall 1996, has anyone made or tried to make you have sexual intercourse or sexual contact when you did not want to by making **threats of non-physical punishment** such as lowering a grade, being demoted or fired from a job, damaging your reputation, or being excluded from a group for failure to comply with requests for any type of sexual activity?

19. Since school began in the Fall 1996, has anyone made or tried to make you have sexual intercourse or sexual contact when you did not want to by **making promises of rewards** such as raising a grade, being hired or promoted, being given a ride or class notes, or getting help with course work from a fellow student if you complied sexually?

20. Since school began in the Fall 1996, has anyone made or tried to make you have sexual intercourse or sexual contact when you did not want to by simply being **overwhelmed by someone's continual pestering and verbal pressure?**

22. Not counting any incidents we have already discussed, have you experienced any other type of unwanted or uninvited sexual contact since school began in the Fall? Remember, this could include sexual experiences that may or may not have been reported to the police or other officials, which were with strangers or people you know, in variety of locations both on- and off-campus, and while you were awake, or when you were asleep, drunk, or otherwise incapacitated.^a

24. Since school begin in the Fall 1996 has anyone—from a stranger to an ex-boyfriend—someone *repeatedly* followed you, watched you, phoned, written, e-mailed, or communicated with you in other *ways in a way that seemed obsessive and made you afraid or concerned for your safety?* This includes waiting outside your class, residence, workplace, other buildings, or car.^b

^a Each question was asked using a “yes-no” response set. After each series of questions or question (7, 8, 9, and 10; 12; 14; 16; 18, 19, 20; 22), the following question was asked: “How many different incidents of [type of sexual victimization] happened to you since school began in the Fall 1996?”

^b After the stalking screen question, the following question was asked: “How many people exhibited this type of behavior toward you since school began in the Fall?”

Exhibit 8. NCWSV Study: Operationalizing rape and attempted rape

Initial question	Response	Skip question	Response	Skip question	Response	Final classification
R12. Was the sexual contact in this incident threatened, attempted, or completed (at least some sexual contact actually happened)?	<i>Completed</i> →	R13. Tell me which of the following actually occurred to you during this incident. Just say yes or no. Did you experience . . . READ LIST AND MULTIPLE RECORD	Any response from 1 through 8 in footnote 1 →	R17. Was physical force actually used against you in this incident? R18. Were you threatened with physical force in this incident?	Yes →	Completed rape
R12. Was the sexual contact in this incident threatened, attempted, or completed (at least some sexual contact actually happened)?	<i>Attempted</i> →	R15. What (other) type of unwanted sexual contact was attempted? READ LIST AND MULTIPLE RECORD	Any response from 1 through 8 in footnote 1 →	R17. Was physical force actually used against you in this incident? R18. Were you threatened with physical force in this incident?	Yes →	Attempted rape

^a The list included: (1) penis in your vagina, (2) a mouth on your genitals, (3) your mouth on someone else's genitals, (4) penis in your anus or rectum, (5) finger in your vagina, (6) finger in your anus or rectum, (7) another object in your vagina, (8) another object in your anus or rectum, (9) touching, grabbing, or fondling of your breasts or genitals under your clothes, (10) touching, grabbing, or fondling of your breasts or genitals over your clothes, (11) kissing, licking, or sucking, (12) some other form of unwanted sexual contact, and (13) none of these.

^b If the respondent said "no" to R17, she was then asked R18.

^c The list included: (1) penis in your vagina, (2) a mouth on your genitals, (3) your mouth on someone else's genitals, (4) penis in your anus or rectum, (5) finger in your vagina, (6) finger in your anus or rectum, (7) another object in your vagina, (8) another object in your anus or rectum, (9) touching, grabbing, or fondling of your breasts or genitals under your clothes, (10) touching, grabbing, or fondling of your breasts or genitals over your clothes, (11) kissing, licking, or sucking, (12) some other form of unwanted sexual contact, and (13) none of these.

^d If the respondent said "no" to R17, she was then asked R18.

Operationalizing rape in the incident report

Modeled after the NCVS incident report, we also designed two incident reports: (1) one for rape, sexual coercion, unwanted sexual contact, and threats, and (2) one for stalking. In the first incident report, we obtained information (1) to determine exactly what type(s) of sexual victimization occurred and to what degree (completed, attempted, or threatened), (2) to document information about the characteristics of the incident, and (3) to understand the reporting behaviors of the victim.

To determine if the respondent had experienced a rape (either completed or attempted), we developed a series of questions as shown in exhibit 8. First, the interviewer asked the respondent if the incident was threatened, attempted, or completed (see question R12 in exhibit 8). Depending on her response, the interviewer then asked the respondent which sexual act(s) was completed, which act(s) was attempted, and/or which act(s) was threatened. We collected information on these three degrees of victimization because we believed that an incident could be a single victimization, but it could also involve a series of victimizations that led to the most serious type of victimization experienced.

A respondent could answer one of the three responses or all three responses because it was possible that a single incident resulted in more than one victimization, either of the same type or of a different type. For example, if a respondent reported that there was attempted vaginal-penile penetration with force and completed unwanted sexual contact (e.g., touching of her breasts or buttocks) with the threat of force, then there were two victimizations during this one incident: an attempted rape and completed sexual coercion. Another incident could have included the same type of victimization: a completed penile-vaginal penetration with force and an oral-genital penetration with force (both are completed rapes). Information was collected on all victimizations for that respondent arising from a single incident.

Because some incidents involved more than one type of victimization, counting each would have inflated our counts of the different types of victimization. To address this methodological concern, we classified each incident as to the most severe type of sexual victimization that the respondent experienced with that specific incident. Koss and associates (1987, 165) also used this "most severe" scoring procedure for respondents in her study. The NCVS uses this type of procedure, too.

As shown in row 1 of exhibit 8, if the respondent indicated that the sexual contact was completed, she was then asked which type(s) of penetration were completed. The interviewer read her a list of different types of penetration (see footnote 1 in exhibit 8). The respondent answered with a "yes" or "no" answer to one, some, or all the types of penetration.

The respondent was then asked two questions about the use of physical force or threatened use of physical force (see questions R17 and R18 in exhibit 8). First, the interviewer asked her if physical force was used against her (see R17 in exhibit 8). If she said “yes,” the incident was classified as a completed rape. If she said “no” to question R17, the interviewer asked the respondent if she was threatened with physical force (see R18 in exhibit 8). If she said “yes,” the incident was classified as a completed rape.

As shown in row 2 of exhibit 8, an incident was classified as an attempted rape using the same series of questions that we discussed for a completed rape incident. The one difference is that the respondent indicated that the sexual contact included an attempted type(s) of penetration.

Study results

Close to 2 percent (1.7 percent) of the college women in our sample experienced a completed rape since school had begun in fall of 1996. Slightly more than 1 percent (1.1 percent) of the sample experienced an attempted rape. The percent of the respondents who experienced either a rape or an attempted rape was 2.5 percent. Note that in the 1997 NCVS, this pattern of the number of completed rapes being higher than attempted rapes is also found.

How do these results compare with other studies? To our knowledge, no published studies used a similar reference period and employed a national sample of randomly selected women who were currently enrolled in either 2-year or 4-year schools. Koss’ study, however, comes the closest to ours in terms of methods (behaviorally defined questions, sample). If we take her 1-year estimates for completed rape and attempted rape where we have a comparable definition (i.e., alcohol and drugs question estimates are excluded from the calculations)⁶ and project them for our reference period (6.91 months),⁷ we can roughly compare our results with her results. We see that our estimates of completed rape are similar to Koss’ estimates (1.7 compared with 2.1 per 1,000 female students). Our estimate of attempted rape is lower than her estimate (1.1 compared with 3.3 per 1,000 female students).

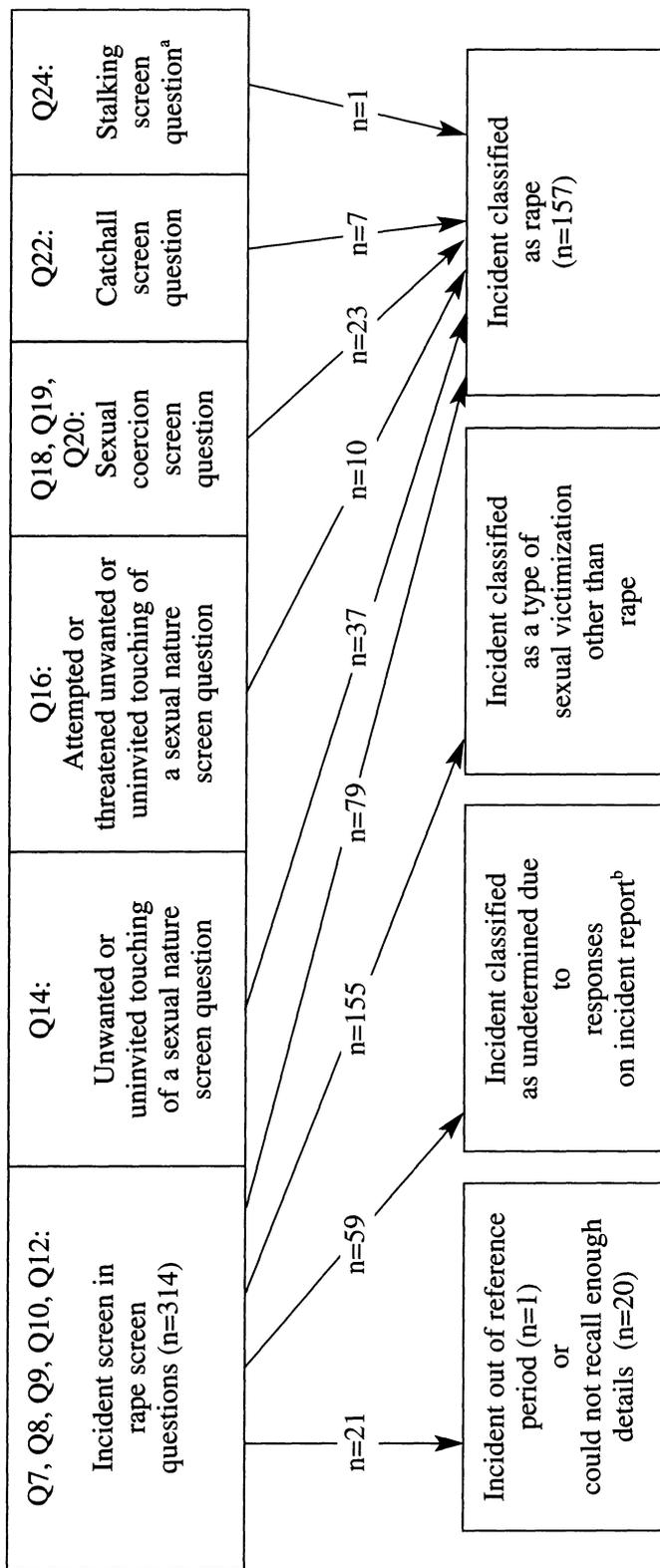
As with the measurement of any phenomena, measurement error can and does occur.

The issue then becomes to identify the potential sources of bias and determine their effects.

Potential sources of measurement error

As with the measurement of any phenomena, measurement error can and does occur. The issue then becomes to identify the potential sources of bias and

Exhibit 9. NCWSV Study: Incident classification



^a In the stalking incident report, if the respondent indicated that the perpetrator made or attempted to make the respondent have unwanted sexual contact, she was then asked if she had mentioned this incident(s) when asked about unwanted sexual intercourse or contact earlier in the survey. If she said "no," the interviewer then administered a sexual victimization incident report for that incident(s).

^b We could not determine what happened because the respondent refused to answer or said "don't know" to questions that would have allowed us to classify the incident.

determine their effects. In this next section, we discuss the ways in which we believe our screen question-incident report method addresses some of the sources of measurement bias that we argue are inherent in studies that estimate rape from behaviorally specific questions and those that use the incident report format.

Answering “yes” to a behaviorally specific rape question(s)

Studies such as Koss’ college women, Kilpatrick’s NWS, and Tjaden and Thoennes’ NVAW Survey use behaviorally specific questions to cue respondents as to the type of behavior in question. They then use a “yes” response to the rape questions to count the number of rape victims to generate their rape estimate. The assumption here is that since the question is explicit in describing what is being measured, measurement error will be minimized. That is, they make the assumption that a “yes” response to any of their respective rape screen questions is a necessary and sufficient condition to classify whether or not the respondent is a rape victim.

In exhibit 9, we present results from our NCWSV Study suggesting that this assumption is not supported and that using only a single step of “cueing” questions to measure victimization could result in the number of rape victims being *overestimated*. Thus, in the NCWSV Study, 314 victimization incidents screened into the incident report on the behaviorally specific screen questions that were designed to cue respondents to report this fact to the interviewer. Of these 314 incidents, however, only 25.2 percent were ultimately classified as a rape once respondents were probed further as to what occurred with questions in the incident report (see exhibit 9). In contrast, nearly half (49.4 percent) of the incidents that entered the incident report via the rape screen questions were classified as a type of sexual victimization other than a rape. Further, the fact that a respondent answered “yes” to a rape screen question does not necessarily mean that it can be determined that a victimization had, in fact, occurred. Slightly more than a fourth (25.5 percent) of the incidents cued by the screen questions could not subsequently be classified in the incident report. Here, 18.8 percent of these incidents could not be classified because the respondent refused to answer or said “don’t know” to questions that would have allowed us to classify the incident (e.g., refused to answer the force questions but indicated that penetration took place). In 6.7 percent of the incidents, the respondent could not recall enough details to complete an incident report, or the incident was out of the reference period. Research studies that count incidents such as these as rapes thus risk including incidents in their victimization totals that may not qualify legally as rapes.

Use of the incident report dramatically changed our rape estimate. If we had counted all 314 incidents that screened in on our rape screen question, our

victimization rate would have been 1.6 times higher than our rate calculated from the incident report—70.6 per 1,000 college women, compared with 44.8 per 1,000 college women.

Answering “yes” to a non-rape screen question

Conversely, previous research—such as the studies that use Koss’ SES—assume that women who answered “yes” to a cueing question meant to measure sexual victimizations other than rape did *not* experience a rape victimization. Again, this method assumes that behaviorally specific cueing questions are capable of directly measuring a specific type of victimization. It does not consider that non-rape cueing or screen questions might prompt a “yes” answer that, upon further questioning, is discovered actually to have been a rape. If this occurs, then relying exclusively on cueing or screen questions can produce a second kind of measurement error in which rape incidents are *underestimated*.

As we discussed, our method allowed respondents to screen into an incident report from any of our 12 screen (or cueing) questions. Screening in on a non-rape screen question did *not* mean that the respondent in the incident report could not answer questions in such a way that we would eventually classify her incident as a rape. Instead, regardless of the screen questions that the respondent answered “yes” to, the detailed incident report questions as to what happened allowed for the possibility that the incident could be classified as a rape if the respondent’s answers meet the required criteria.

As the results in exhibit 9 show, close to half (49.7 percent) of the incidents that we classified as a completed rape or an attempted rape screened in on our *non-rape* screen questions. A closer look at our findings reveal that all five non-rape screen questions yielded at least one incident that we classified as a rape. For example, the two non-rape screen questions that subsequently resulted in the highest number of classified rape incidents were the unwanted or uninvited touching question and the sexual coercion question (23.6 percent and 14.6 percent of the rape incidents, respectively). Again, a method that relied only on cueing questions and that did not include an incident report would have omitted these incidents from the count of rape incidents experienced by women in the sample.

These results raise two fundamental questions which are at the core of measuring rape and other forms of sexual victimization: (1) Why do women who answer “yes” to a rape screen question subsequently not answer questions in the incident report that would allow the incident to be classified as a rape? And (2) why do women who answer “no” to a rape screen question and “yes” to some other screen question eventually answer “yes” to questions in the incident

report that allow the incident to be classified as a rape? Although we are unable to provide data on these questions, we raise them because they are issues that are central to understanding the measurement of rape and thus need to be examined further by future researchers.

Is it rape?

As noted, Gilbert (1997) has questioned whether measurement devices such as Koss' SES are validly measuring rape victimization when a large proportion of those categorized as rape victims (nearly three-fourths in Koss' study) do not define what happened to them as a "rape." Our study attempted to examine this issue. In each incident of rape, we asked victims, "Do you consider this incident to be rape?" For the 86 incidents categorized by our definition of completed rape, in 46.5 percent (n=40) of the incidents women answered "yes" to this question, in 48.8 percent (n=42) of the incidents they answered "no," and in 4.7 percent (n=4) of the incidents they answered "don't know." For attempted rape incidents, in only 2.8 percent of the incidents (n=2) did the respondents define their victimization as rape. In 95.8 percent (n=68) of the incidents women answered "no," and in 1.4 percent (n=1) of the incidents women answered "don't know."

How should we interpret these data? The attempted rape data are problematic because we did not ask specifically whether the respondents believed a rape had been "attempted." Putting this issue aside, what should be made of the data showing that only about half of the women who qualified as completed rape victims in our study defined their victimization in this way?

At best, we can present two competing perspectives. Skeptical of victimization survey data, conservative commentators are reluctant to count any event that the victims themselves do not label as "rape" (Gilbert 1997). After all, it can be argued, adult women know when a rape has occurred. In contrast, feminist commentators wish to count as a rape any event that conforms to the legal standard for rape: unwanted sexual penetration by force or threat of force (Koss 1992, 1996). In their view, female victims may manifest a lack of knowledge of the law or false consciousness when they define forced sexual penetration as something other than a rape. In reality, to our knowledge no research has yet been undertaken that systematically explores why women on surveys who are victims of coerced sexual assault do, or do not, define that incident as a rape. Accordingly, which of these two interpretations is more or less correct cannot be definitively substantiated.

Other data from our study, however, complicate this issue still further. The question, "Do you consider this incident to be rape?" was asked not only of

those categorized as rape victims but also *for every respondent who indicated some form of sexual victimization*. Beyond the completed rape and attempted rape victims, 40 women in the sample defined their sexual victimization as a rape.

It is possible, of course, that these female students who defined themselves as rape victims did not know the legal requirements of rape and thus mischaracterized what happened to them. An alternative possibility, however, is that surveys on sexual victimization using close-ended questions—even when using behaviorally specific screen questions and carefully worded incident-report questions—fail to capture rapes that actually take place. For example, the line between a “threat” and an “attempt” may, in real circumstances, be a thin one that only can be ascertained by more probing closed-end questions or by asking open-ended questions that then lead to further questions that ask the respondent to examine what she meant. It also may be that forms of “pestering/verbal pressures” may escalate to the point where they become “force or threat of force.”

Again, issues such as these can only be clarified by further research that uses followup questions to ask respondents why they did or did not define an act as a rape. At this stage, we must admit that victimization surveys leave significant methodological questions unanswered and thus can provide only “ballpark” estimates of how much rape and other types of sexual victimization occur. It is important to realize, however, that response biases on these surveys may not, as conservative commentators contend, only be in the direction of overestimating the extent of rape. As our results suggest, there also is the possibility of underestimating how much rape victimization actually takes place.

Stalking

To our knowledge, our NCWSV Study is the first national-level survey to estimate the extent of stalking committed against college women. Using State-level legal statutes as our guidelines, we defined stalking much the way Tjaden and Thoennes did: the same person exhibiting repeated behavior that seemed obsessive and made the respondent afraid or concerned for her safety.

The screen question used to measure stalking is shown in exhibit 7 (see question 24). Given that stalking by definition involves repeated behaviors, we developed a separate incident report for this form of victimization. Similar to our sexual victimization incident report, the stalking incident report contained questions about when the stalking took place, characteristics of the offender(s), the nature of any injuries sustained, and police reporting behavior.

To operationalize stalking, we asked the respondent about what the person did that seemed obsessive and frightening to the respondent. These included: followed you, waited outside or inside places for you, watched you from afar, telephoned when you did not want the person to, sent unwanted letters or cards, sent unwanted e-mail messages, or made other unwanted contact. To document that the incident was repeated, we asked the respondents how often these events occurred.

Our results revealed that 13.1 percent of the women in our sample had been stalked at least once since the academic year had begun. Recall that the NVAW Survey's annual stalking estimate ranged from 1 to 6 percent, depending on how fearful a woman had to be to count as a stalking victim. There are three main differences between our study and the NVAW Survey that could affect our respective estimates of stalking. First, we did not include some of the behaviors that the NVAW Survey did (e.g., vandalized property, left unwanted items); alternatively, the NVAW Survey did not include behaviors that we queried about (e.g., stalking through e-mail). Second, as we have noted, Tjaden and Thoennes (1998b) used a stringent criterion in their definition of stalking; respondents had to feel very frightened or fear bodily harm to be counted as a stalking victim. We used the terms "made you afraid or concerned for your safety" and "frightening" to operationalize the fear criterion used in many State-level statutes of stalking. Third, we employed a national-level college women sample. The NVAW Survey employed a national-level sample of women from the community at large. The limited research to date suggests that stalking victims tend to be young, and thus we may have measured stalking among a group of high-risk women (Fisher, Cullen, and Turner 1999; Tjaden and Thoennes 1998b).

Although important insights about stalking can be drawn from Tjaden and Thoennes' NVAW Survey and from our NCWSV Study, the methodology for measuring stalking remains in its preliminary stages. Far more research is needed to develop more valid screen questions and an incident report capable of capturing more precisely what occurred during the panoply of victim-offender encounters that combine to comprise a stalking "incident."

Verbatim responses in the National Violence Against College Women Study

During the same time as the NCWSV Study, we conducted another national-level study to examine the extent and nature of violence among college women—the National Violence Against College Women (NVACW) Study. Acts of violence that we collected data on included rape, sexual assault, robbery, aggravated assault, simple assault, and unwanted sexual contact.

We revised and tailored BJS' NCVS to the needs of this study. Accordingly, in terms of measuring violence—specifically rape and sexual assault—we used all of the violent screen questions contained in the NCVS. In the incident report, we included those questions that were needed to determine the type of violent incident that the respondent had experienced (e.g., what happened, what was attempted, what was threatened, use of a weapon). Since we were dealing with a college student sample, we tailored some response sets so that they would be more relevant to this sample. For example, we asked whether the incident happened on or off campus, and we followed up with a closed-ended question that asked where specifically on campus or off campus the incident had transpired (for details, see Fisher and Cullen 1999). The response rate for this study was 91.6 percent.

The research design used in the NVACW Study is identical to the one used in the NCWSV Study in every way except three. First, the wording of the screen questions and of questions used in the incident report differed from the NCWSV Study. Except for the changes we noted, the screen questions and incident report questions were identical to those used in the NCVS. Second, rape and sexual assault were defined and operationalized according to how BJS defines, and the NCVS operationalizes, these two types of sexual victimizations.

Third, like the NCVS, the NVACW Study collected two sets of *verbatim responses* from open-ended questions in which respondents *used their own words* to describe their victimization experience to the interviewer. The first of these questions was asked for each incident reported in the screen question. The interviewer asked the respondent to “briefly describe the incident.” The second of these questions was asked at the end of the incident report. The interviewer instructed the respondent again to describe what happened. After each question, the interviewer recorded the respondent’s description using the respondent’s own words.

These verbatim responses are not structured or designed to measure in a systematic way what kind of victimization occurred. We included them originally in the survey only because they are part of the NCVS interview schedule. After completing the study, however, we realized that the respondents’ personal descriptions of their victimization incidents might be used as a potential source of data. In particular, these descriptions offered the opportunity to compare how a woman’s responses were coded using the closed-ended questions with what she stated had occurred to her in her verbatim responses.

Because the verbatim responses were not devised to collect detailed information on the victimization incident, it is at times difficult to discern whether the incident described by a given respondent would meet all the legal criteria for the

criminal offense of rape. To address this problem, we coded each verbatim response in two ways: (1) the most serious offense the verbatim seemed to suggest occurred, and (2) the least serious offense the verbatim seemed to describe. Across all respondents, we were then able to aggregate this coding to reach, based on the verbatim responses, a count that would range from a low number (i.e., only cases “definitely” described as a completed rape are counted as a completed rape) to a high number (i.e., all cases that “seem” to be a completed rape are counted as a completed rape). In essence, we used qualitative data to secure a quantitative estimate of the lower and upper bounds of rape in the sample, based on the content of the verbatim responses. In exhibit 10, we provide a sample of verbatim responses that we coded as a completed rape or an attempted rape. The first two columns are the verbatim responses, and the third column is the range of our coding classification for that incident (see Fisher and Cullen 1999).

Again, we analyzed the verbatim responses for all victimization incidents (rape and other forms of sexual and violent victimization). Across all the verbatim responses analyzed, how many could be interpreted as describing a completed, attempted, or threatened rape? (see exhibit 11). This is the upper bound estimate. Alternatively, using a “conservative” interpretation of the verbatim responses, what is the least number of rapes one might conclude occurred among the female students surveyed? As can be seen from exhibit 11, the lower bound and upper bound range is fairly wide: Based on the verbatim responses, there could be as few as 12 rape incidents in the sample or as many as 42 rape incidents in the sample. It is noteworthy that the upper bound figure is more than twice as high as the number of rape incidents computed from the responses to the closed-ended questions (n=20).

The findings on the verbatim responses suggest two further conclusions. First, the ambiguity of the descriptions of the incidents given by many respondents makes it risky to conclude that the upper bound estimate is the “true” or “most accurate” measure of rape incidents in the sample. Still, it is important to note that many of the verbatim responses that were coded as rape came from the incidents categorized as sexual assault and unwanted sexual contact by the closed-ended questions. This result raises the possibility that the NCVS’ closed-ended questions miscategorize at least some—and potentially a significant number of—rape incidents. Second and relatedly, future research may benefit from exploring how the verbatim descriptions of sexual-related incidents may be conducted in a more structured way (1) to collect more information on the incident and (2) to permit a more definitive categorization of the incident. If so, then it might be possible to explore more systematically the correspondence between the closed-ended and verbatim responses from the NCVS and, in turn, to develop a more accurate measure of sexual victimization.

Exhibit 10. Sample of verbatim responses from the NVACW Study

Verbatim responses from open-ended questions: Screen question	Verbatim responses from open-ended questions: Incident question	Type of crime code or range of type of crime code based on verbatim responses
Rape^a	I went to visit one of my friends, and we were all hanging out together, then we went to my friend's house, down into the basement. The friend that I had gone to see went upstairs and left me alone with this other friend. He told me he was going to get me, and I tried to get out but he pushed me down and raped me . I knew that he had a gun, so I was afraid to do anything.	Completed rape
Sexual assault	While walking home from work someone walked up to me and asked the time and then he walked past me 15 to 20 feet. Later he was following me I walked 3 blocks and then someone else passed me and then a few blocks more they were behind me with a gun in my back they asked for money and I had none so they took me behind a house and assaulted me there .	Sexual attack with serious assault to Completed rape
A guy tried to rape me	I invited him in (<i>who?</i>) ^b he was a friend of mine. He tried to do stuff that I didn't want to do. I hit him and kicked him out of my room.	Attempted rape
Someone tried to force himself sexually upon me	I was driving him home from a bar and I asked him to wait in the car because I had to go to the bathroom I got back in the car and said thank you and he tried to force himself on me I pushed him off and he tried again with more force I told him again to get off, he tried again with more force but didn't hurt me . I pushed him off and told him to get out of my car at that point.	Sexual attack with minor assault to Attempted rape

^a Operative words used in coding are in boldface.

^b This is the interviewer's question.

Impact of screen question-incident report method on rape estimates

The NCWSV Study and the NVACW Study provide an opportunity to explore how the screen question-incident report method used to measure rape might influence rape estimates. Recall both studies used the same sampling design, sent

Exhibit 11. Changes in incident estimates and rates from closed-ended questions in the incident survey, using the verbatims from all the incidents coded a "sexual" type of crime (n=145)

Type of incident	Frequency distribution based on closed-ended questions in incident report	Rate per 1,000 female students	Frequency distribution based on verbatim responses: Lower bound	Rate per 1,000 female students	Frequency distribution based on verbatim responses: Upper bound	Rate per 1,000 female students
Rape	20	4.51	12	2.70	42	9.48
Completed rape	9	2.03	4	0.90	13	2.93
Attempted rape	8	1.81	6	1.35	26	5.87
Verbal threat of rape	3	0.68	2	0.45	3	0.68

a cover letter to selected respondents, and had the same reference period and field period. What differed was the wording used in the screen questions and questions in the incident report. A comparison of the two studies revealed that the estimate of rape in the NCWSV Study is substantially higher than in the NVACW Study. For example, the rate of completed rapes per 1,000 female students was 9.5 times larger in the NCWSV Study compared with the NVACW Study—19.34 and 2.03, respectively. The incidence rate of attempted rape also revealed a similar pattern. The NCWSV Study had markedly higher estimates—8.8 times larger—than the NVACW Study estimates (15.97 compared with 1.81 attempted rapes per 1,000 female students) (Fisher and Cullen 1999). These results suggest, then, that even with a redesigned format, surveys using the NCVS method to measure rape victimization are likely to report far lower rates of victimization than surveys using behaviorally specific screen questions—even when these questions, as in the NCWSV Study, are also accompanied by an incident report similar to the NCVS.

Conclusion: Directions for Future Research

Measuring sexual victimization is a challenging enterprise—the “biggest methodological challenge in survey research” (Smith 1987, 185). Although at times committed by strangers in public places, sexual victimization incidents, including rapes, are most often perpetrated by someone the victim knows, in a residence, and with one or more of the parties using alcohol, drugs, or both. Discerning how much force has been used, or the extent to which consent has not been given, is a daunting methodological task; objective behaviors may be open to diverse interpretations or “constructions of reality” by the people involved in an incident and by researchers seeking to measure what has occurred. This situation is complicated by the fact that many victims deny that their victimization has crossed the line to an attempted or completed rape, and that most do not report their victimization to the police, often saying that the incident was not “serious enough” to warrant such a step (see, e.g., Fisher and Cullen 1998). Some researchers suggest that whether a victim defines an act as a rape or reports it to the police is irrelevant to whether a crime has transpired; others argue, however, that these very facts raise doubts about whether a sexual victimization has taken place (see, e.g., Gilbert 1997; Koss 1996, 1993a).

Some scholars’ interest in measuring sexual victimization has been fueled, at least in part, by their political values. Without such value-laden interest, the progress in this area might have been slower and our understanding of the extent of sexual victimization might have been commensurately limited. The risk, however, is that the politicization of the normally prosaic issue of how to measure crime/victimization pushes scholars into competing camps—one seeming to have a stake in finding as much sexual victimization as possible

and the other seeming to have a stake in finding as little sexual victimization as possible. The ensuing “advocacy battles” may at times produce good scholarship, but they also risk generating more heat than light.

The challenge, we believe, is to learn from the existing research and methodological debates and then to map out what paths the next generation of research might profitably pursue. At the risk of sounding trite, the aim is not to seek methodologies that produce higher or lower estimates of sexual victimizations but, rather, to develop measures that are the most accurate possible. The past two decades of research have been invaluable in moving us toward this goal, but the knowledge of how best to measure sexual victimization remains in a preliminary stage. From the early rape studies to the most recent ones, we have learned that “best” is not a single-item measure. Heeding Blalock’s (1970, 111) advice—“With a single measure . . . one can remain blissfully unaware of the possibility of measurement [error]”—where should future research head? We have several suggestions.

First, a major step forward in measuring sexual victimization was taken with the development by Koss and others of behaviorally specific questions to “cue” respondents to report a victimization to an interviewer. These questions use a variety of words that are meant to define key legal elements of a criminal victimization, especially force and consent. At issue, however, is whether the words used to capture these legal elements mean the same thing to all respondents—or mean the same thing to respondents as they do to the researchers who carefully craft the questions.

For example, these cueing or screen questions commonly employ such words as “have sex,” “intercourse,” “using force,” “threatening or attempting harm,” and “unwanted or uninvited” contact. At what point does having sex occur? What does intercourse mean? What constitutes force, or a threat, or an attempt? What makes an act unwanted or uninvited? More salient, on a survey questionnaire, what do these words *mean* to respondents? Is there consensus or discord on what, for example, “intercourse by force” means? For example, the term “sex”—which according to the *American Heritage Dictionary* means sexual intercourse—may have different meanings other than the meaning in the dictionary for college students. A recently published article in the *Journal of the American Medical Association* addressed the issue of what students considered as having “had

These issues are important because, to the extent that these words and phrases are differentially interpreted, they will cue very different responses by women who may have experienced the same type of incident.

sex.” Using a sample of undergraduate students at a Midwestern school, Sanders and Reinisch (1999) reported that individual attitudes varied regarding behaviors defined as having “had sex.” Thus, 59 percent of the respondents indicated that oral-genital contact did not constitute having “had sex” with a partner, and 19 percent responded similarly regarding penile-anal intercourse. These results suggest that it is essential for researchers to explicitly define for the respondents what is meant by intercourse and what types of penetration are being asked about.

These issues are important because, to the extent that these words and phrases are differentially interpreted, they will cue very different responses by women who may have experienced the same type of incident. One line of research, therefore, should be to explore what respondents believe terms such as force, attempt, threaten, and unwanted mean to them *in the context of a victimization survey*. This task could be addressed through focus groups that discuss in detail what specific words on a survey instrument mean to them. Another possibility would be to construct vignettes that experimentally vary different types of, say, force and then see how women define the “force” used in the vignette (e.g., if this act occurred to them, would they say that “force” was used against them?) (Schwartz 1998).

In our own research, we use a combination of behaviorally specific questions to cue the recall of incidents and an incident report that asks a series of more detailed questions.

Second, we have only a beginning understanding of the validity of the cueing or screen questions used in sexual victimization surveys and of how different questions might prompt women to report, at higher or lower levels, their having been sexually victimized. Take, for example, the case of measuring rape victimization. The research using incident reports shows that women eventually classified as rape victims are cued or “screen into” the incident report not only by questions designed to cue the recall of a rape victimization but also by questions designed to cue the recall of other sexual victimizations and of other nonsexual types of assault (Fisher and Cullen 1998; Percy and Mayhew 1997). Research that only uses behaviorally specific “rape” questions thus would fail to count

women whose recall is prompted by other types of cueing or screen questions. Beyond this example, our knowledge of what questions or combinations of questions are most effective in cueing the reporting of victimization incidents remains underdeveloped. Systematic studies using experimental designs are needed to investigate how the range and wording of questions affect how respondents report victimizations in interviews.

Third and relatedly, there is a need for conceptual work in deciding precisely what behaviors fall under the domain of “sexual victimization.” By analogy, the early self-report scales developed to measure delinquency usually contained only a limited number of items (e.g., six) and tended to measure petty forms of delinquency (e.g., shoplifting, joy riding, fist fights). Subsequently, scholars developed much lengthier measures (e.g., 25 to 50 items) that used multiple items to assess different types of delinquent acts, ranging from status offenses to petty thefts to serious violence, and including crimes committed in the home, in school, and in the community. Most of the measures using cueing or screen questions are relatively short and do not use a large number of items to operationalize different types of sexual victimization (e.g., different types of rapes). Again, research is needed to explore whether more comprehensive sets of cueing questions would be more effective in prompting respondents to report rape and other victimization incidents.

Fourth, in our own research, we have opted for using a combination of behaviorally specific questions to cue the recall of incidents and an incident report that asks a series of more detailed questions to determine how the incident should be classified. This approach has the advantage of using responses to detailed questions, rather than a simple “yes” response to a screen question, to categorize incidents. Still, we have only a preliminary understanding of what sources of measurement error the use of incident reports might introduce.

For example, in our NCWSV Study, nearly 1 in 5 incidents that were initially classified as a rape by the responses to the screen questions were then classified as “undetermined” because the respondents either refused to answer or answered “don’t know” to one or more questions in the incident report used to categorize incidents as a rape. It is possible, of course, that the incident report “did its job” by diverting from our count of rape victimizations those incidents that did not meet all of the legal criteria for rape (i.e., criteria measured by the questions in the incident report). It also is possible, however, that some “real” rape victims were not counted because they did not understand questions in the incident report or wearied at having to answer a second round of questions about a potentially painful event in their lives. In the absence of further research, discussions of possible measurement error associated with the use of incident reports will remain speculative.

Fifth, we also might entertain the possibility that new methods should be used to conduct surveys of sexual victimization (Schwartz 1998). Percy and Mayhew (1997) note that the 1994 British Crime Survey (BCS) included an innovative approach to measuring the victimization of women: having women use computers to read and answer questions on the BCS. When compared with the cue or screen questions normally used in the BCS, the completion of cue questions on

a computer resulted in a reporting of victimization that was 10 times higher than that normally reported in the survey to interviewers (7.5 percent versus 0.1 percent in the past year). These findings must be qualified, because the items in each measure of sexual victimization, though similar, were not identical to those in the BCS, and the responses to the computer survey were not validated by asking more detailed questions about the nature of the incident. Still, these caveats aside, the computer-assisted survey almost certainly produced more reports of victimization.

Why is this so? In general, the use of computers has been found to increase the self-reporting of a variety of sensitive behaviors, such as drug use, sexual conduct, and violence (Percy and Mayhew 1997; Turner et al. 1998). In a similar way, publicly revealing a sexual victimization incident to an interviewer may be embarrassing or otherwise discomfiting. If the perpetrator of the event is an intimate, the respondent may fear that their reporting of this information might be overheard by or otherwise “get back to” the perpetrators (Crowell and Burgess 1996). In contrast, the computer may overcome these feelings of sensitivity by furnishing the kind of anonymity that encourages more “open answers” (Percy and Mayhew 1997, 147).

As Percy and Mayhew (1997, 147) understand, however, this computer approach to measuring sexual victimization is not yet “state of the art” but only a “methodological toe in the water.” One daunting challenge will be to design computer-based surveys that respondents who answer “yes” to screen questions can then easily and reliably use to answer complex incident reports. If this is not possible, it may be necessary to combine methods, using the computer for screen questions and interviewers for the incident report. The more significant problem, however, may be that “the validity of a measure is assessed relative to a ‘true’ score. Without a true score of sexual victimization, or any valid independent criterion, it is only possible to assess the relative efficacy of different survey methods” (Percy and Mayhew 1997, 146). Of course, how to obtain a “true score” of sexual victimization has been the challenge researchers have struggled with for the past two decades. Still, through in-depth interviews with a sample of respondents, it might be possible to develop a reasonable appraisal of whether they were victimized. The respondents’ “true score” or baseline data would then be compared with the respondents’ scores on various victimization surveys, including those that have study participants use computers to convey their victimization.

Sixth and relatedly, an inherent restriction to existing strategies for measuring sexual victimization—whether it is Koss’ SES, the NCVS, or our method of trying to combine these two approaches—is that they rely on a preexisting sequence of questions that is inflexibly used with all respondents. In our research, however,

we were struck by the potential insights that could be gained from the “verbatim” responses that interviewers obtained in our NVACW Study that employed the NCVS methodology. It would have been methodologically useful at key junctures in the survey to have paused and asked respondents to explain, in their own words, what happened to them or why they were answering a question in a particular way. For example, if a respondent had screened into the incident report on a rape question but then failed to answer “yes” that she had been “forced” to have intercourse, it would have been useful for the interviewer to have had the flexibility to probe further why this discrepancy in the responses was occurring. Or, if a respondent was classified as a rape victim but then said that she was not “raped,” we would have liked to have probed why she answered in this way. More generally, a detailed description of what the incident entailed would have furnished information to clarify whether the closed-ended questions were accurately capturing the “reality” of the incident.

In this regard, then, we would propose that sexual victimization surveys be refined to include what we might call “structured qualitative” questions. On a general level, it would be important to collect data by having respondents “tell their story” of what the incident entailed (i.e., give a longer verbatim account of “what happened” to them) (see Smith 1994). On a more specific level, survey instruments could be developed that, in a more structured way, instructed interviewers at what point they should probe for and record more information. For example, if a respondent said that “force” was used, the interviewer might then use a series of probes to elicit a description of the precise kind of force the assailant employed.

These qualitative data would be useful in three regards. First, they could be used to code incidents more accurately by providing supplementary information on what respondents had experienced. Second, the results from structured qualitative surveys could be compared in experimental studies with the results from traditional surveys that measure victimization only through yes-no answers to a series of closed-ended questions. We should note that research in other fields, such as the measurement of stressful life events, has used this method of “intensely probing” respondents, within the context of a structured questionnaire, to provide verbatim response that explained “what actually happened” to you (Shrout et al. 1989). Third, verbatim accounts could shed light on the process of sexual victimization—what exactly happened, what decisions were made when and why, and what actions were taken when and why. This type of understanding could (1) help in the development of a continuum of sexual victimization, and (2) inform sexual victimization prevention programs as to intervention points (see Leidig 1992).

Seventh, sexual “victimization” can be viewed as a reality that is “socially constructed,” with participants interpreting and giving meaning to what occurred in a victimization incident (e.g., it was or was not a “rape”). Research on the measurement of sexual victimization, however, is largely predicated on the assumption that, like other illegal acts, incidents of criminal sexual victimization comprise an objective reality that can be measured. In this view, which we generally endorse, how people construct reality is best seen as a potential source of measurement error, confounding how they might answer a victimization survey. The methodological task, then, is to puncture socially constructed realities and to find out “what really happened” by developing carefully worded questions that are immune to differential interpretation by respondents. Note that this approach owes much to Koss and other feminist researchers who first developed behaviorally specific questions to measure the legal elements of sexual victimization. They understood that relying on how respondents socially constructed victimizations would lead many women who had been sexually assaulted, but were not aware of the criminal nature of this assault, *not* to report on a survey that they had been victimized criminally.

For virtually any other crime (e.g., larceny, burglary, robbery), the idea of measuring objective, rather than socially constructed, reality would raise barely a ripple of concern. With sexual victimization, however, the nagging understanding remains that how people socially construct sexual victimization incidents may be a large, not a small, source of “measurement error” in how people respond to questions on victimization surveys. We have already noted the importance of exploring what the wording of questions means to respondents and the need to probe respondents to determine what they may have experienced in a victimization incident. But we also need to know more about how a range of factors—for example, sociodemographic characteristics, individual personality traits, political consciousness, friendship support networks, and personal and vicarious experiences with previous victimization—affect how people interpret sexual victimization incidents, and then how these factors may influence responses on victimization surveys.

Developing a methodology to explore these issues is likely to be a formidable task. Again, vignette research may be a beginning strategy that is not cost prohibitive. Here, respondents (1) would be given scenarios describing incidents involving and not involving sexual victimization, (2) would be instructed to imagine that they were a participant in these incidents, and (3) would be asked to complete a survey reporting “what happened to them.” The respondents would also complete a battery of questions measuring their personal and social characteristics. The final step would be to assess if responses to the victimization questions are conditioned by these characteristics and, if so, to what degree. In

turn, these findings might be used to inform field studies that would probe the impact of these characteristics within the context of actual victimization surveys.

Eighth and finally, there is a need for longitudinal studies that explore the sexual victimization of women over the life course. With the limited exception of the NCVS (which follows households for 3 years), the major national studies of sexual victimization have used cross-sectional designs. Although valuable, these approaches have inherent limitations: the lack of a clear time bound for the reference period of the survey; the necessity to measure lifetime prevalence rates of victimization retrospectively; and the inability to explore how developmental processes, including past victimization, are potentially implicated in current victimization experiences.

In closing, we recognize that the measurement of sexual victimization will remain an imperfect endeavor. The nature of sexual victimization, and the often deeply felt reactions victims experience, undoubtedly restrict what can be measured by traditional social science survey methods. Even so, we do not believe that we have reached the limits of what we can measure and know in this area. Moreover, regardless of whether the “true” prevalence of rape annually is 1 percent or 5 percent, we understand that advocates need to move ahead and continue to establish programs to protect women against sexual victimization.

While this important task proceeds, however, researchers should not relinquish their obligation to muddle through the difficult and often uninspiring task of worrying about the wording and sequencing of questions on surveys. Survey methodologists have consistently shown that question wording and sequencing, response set ordering, and closed-ended compared with open-ended questions do influence estimates (see Lyberg et al. 1997). Although research of the measurement of sexual victimization has come a long way, it has now reached the point of illuminating not intractable truths, but the critical questions that remain to be investigated.

For their helpful comments on this manuscript, we wish to thank Brian Wiersema, Patricia Tjaden, and the Criminal Justice 2000 editors, especially Stephen Mastrofski. The results from the college women studies by Professors Fisher and Cullen were supported under award 95-WT-NX-0001 from the National Institute of Justice and 97-MU-MU-0011 from the Bureau of Justice Statistics, U.S. Department of Justice. Professor Fisher, who was a faculty member in the Department of Political Science, also thanks the Charles P. Taft Memorial Fund at the University of Cincinnati for their support during the writing of this chapter.

Notes

1. Unlike the UCR, NIBRS is incident based and includes four categories of sex offenses (forcible rape, forcible sodomy, sexual assault with an object, and forcible fondling). Note that the NIBRS definition of forcible rape differs from the UCR definition. For example, male and female victims are included in the NIBRS definition whereas only female victims are included in the UCR definition. Also, lack of consent is an explicit criteria in the definition of rape. Similar to the UCR, NIBRS includes only those crimes known to the police (see U.S. Department of Justice, Bureau of Justice Statistics and Federal Bureau of Investigation 1997).

2. In a personal communication with Jan Chaiken and Michael Rand (March 4, 2000), they wrote:

In the NCVS, all questionnaires for which any rape or sexual assault code is entered in any of the pertinent items are reviewed to determine whether the codes reflect the written entries [of the interviewers] in the [respondents'] summaries [of what happened]. Where there are clear indications that the coded entries [from the incident report] are not correct, they are edited, using guidelines developed by BJS and Bureau of the Census staffs. This procedure has proven beneficial towards improving the NCVS estimates of rape and sexual assault by removing, to the extent possible, the discrepancies existing between the coded and written entries.

3. The anal sex and digital-object questions do not contain an explicit definition of sexual penetration like the first two questions do; however, both questions are behavior specific.

4. Unlike the Kilpatrick, Edmunds, and Seymour study (1992), Tjaden and Thoennes defined anal sex within their question (see exhibit 6, question F3). The digital-object question is similar to the NWS question in terms of being behavior specific.

5. In addition to behaviorally defining penile-vaginal, oral, and anal penetration, we also provided an explicit behavioral definition for digital-object penetration. Neither the NWS nor the NVAW Survey did this. Our attempted rape screen question included threats of rape but our subsequent discussion of our classification method will clarify that this screen question was not used to ultimately classify any reported incidents.

6. In our incident report, we did not explicitly ask the respondents about being incapacitated due to alcohol or drugs at the time of the incident. Therefore, unlike those researchers who have used Koss' SES or a modified version (see DeKeseredy and Schwartz 1998), we did not use being incapacitated due to alcohol or drugs as a criterion to define rape. The NWS and the NVAW Survey also did not measure directly whether respondents were unable to give their consent to sexual activity because they were incapacitated due to alcohol or drugs. Note, however, that the introduction to our screen questions does explicitly tell the respondents that "You could be awake, or you could be asleep, unconscious, drunk, or otherwise incapacitated. Please keep this in mind as you answer the questions" (see exhibit 7).

7. We summed the total number of days from September 1, 1996, to the date of the interview across all the respondents and divided this sum by the total number of respondents (average number of days=209.43). We then summed the total number of days from September 1, 1996, to May 31, 1997 (273 days). This sum then was divided by the total number of months (30.33). The average reference period was 6.91 months (209.43/30.33).

References

Bachman, Ronet. 1998. A comparison of annual incidence rates and contextual characteristics of intimate perpetrated violence against women from the National Crime Victimization Survey (NCVS) and the National Violence Against Women Survey (NVAW). Background paper for Workshop on Building Data Systems for Monitoring and Responding to Violence Against Women, U.S. Department of Justice, National Institute of Justice and Bureau of Justice Statistics, and U.S. Department of Health and Human Services, National Center for Injury Prevention and Control and National Center for Health Statistics, 29–30 October, Arlington, Virginia.

Bachman, Ronet, and Bruce Taylor. 1994. The measurement of family violence and rape by the redesigned National Crime Victimization Survey. *Justice Quarterly* 11 (September): 499–512.

Belknap, Joanne. 1996. *The invisible woman: Gender, crime, and justice*. Belmont, California: Wadsworth Publishing Company.

Biderman, Albert D., and James P. Lynch. 1991. *Understanding crime incidence statistics: Why the UCR diverges from the NCS*. New York: Springer-Verlag.

Blalock, Herbert. 1970. Estimating measurement error using multiple indicators and several points in time. *American Sociological Review* 35 (1): 101–111.

Brownmiller, Susan. 1975. *Against our will: Men, women, and rape*. New York: Simon & Schuster.

Crowell, Nancy A., and Ann W. Burgess. 1996. *Understanding violence against women*. Washington, D.C.: National Academy Press.

DeKeseredy, Walter S. 1995. Enhancing the quality of survey data on woman abuse: Examples from a national Canadian study. *Violence Against Women* 1 (2): 158–173.

DeKeseredy, Walter S., and Katharine Kelly. 1993. The incidence and prevalence of women abuse in Canadian university and college dating relationships. *Canadian Journal of Sociology* 18 (2): 137–159.

DeKeseredy, Walter S., and Martin D. Schwartz. 1998. *Women abuse on campus*. Thousand Oaks, California: Sage Publications.

Dodge, Richard W. 1984. Comparison of the victimizations as reported on the screen questions with their final classification: 1976. In *The National Crime Survey: Working papers, Volume 2: Methodological studies*, edited by Robert G. Lehnen and Wesley G. Skogan. NCJ 90307. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Eigenberg, Helen M. 1990. The National Crime Survey and rape: The case of the missing question. *Criminology* 7 (December): 655–671.

Fisher, Bonnie S., and Francis T. Cullen. 1999. The victimization of college women: Findings from a national-level study. Draft of final report to be submitted to the U.S. Department of Justice, Bureau of Justice Statistics.

———. 1998. The sexual victimization of college women: Findings from a national-level study. Draft of final report submitted to the U.S. Department of Justice, National Institute of Justice.

Fisher, Bonnie S., Francis T. Cullen, and Michael Turner. 1999. Being pursued: Stalking victimization in a national study of college women. Paper presented at the 1999 Annual Conference of the Academy of Criminal Justice Sciences, 11 March, Orlando.

Fitzgerald, Louise F. 1996. Sexual harassment: The definition and measurement of a construct. In *Sexual harassment on college campuses: Abusing the ivory tower*, edited by Michele A. Paludi. Albany: State University of New York Press.

Geis, Gilbert. 1977. Forcible rape: An introduction. In *Forcible rape: The crime, the victim, and the offender*, edited by Duncan Chappell, Robley Geis, and Gilbert Geis. New York: Columbia University Press.

Gilbert, Neil. 1997. Advocacy research and social policy. In *Crime and justice: A review of research*, edited by Michael Tonry. Vol. 22. Chicago: University of Chicago Press.

———. 1995. *Was it rape? An examination of sexual assault statistics*. Menlo Park, California: Henry J. Kaiser Family Foundation.

———. 1992. Realities and mythologies of rape. *Society* (May/June): 4–10.

———. 1991. The phantom epidemic of sexual assault. *Public Interest* 103 (Spring): 54–65.

Goldberg-Ambrose, Carole. 1992. Unfinished business in rape law reform. *Journal of Social Issues* 48 (1): 173–185.

Goodman, Lisa A., Mary P. Koss, Louise F. Fitzgerald, Nancy Felipe Russo, and Gwendolyn Puryear Keita. 1993. Male violence against women: Current research and future directions. *American Psychologist* 48 (October): 1054–1058.

- Gordon, Margaret T., and Stephanie Riger. 1989. *The female fear*. New York: Free Press.
- Gruber, James E. 1990. Methodological problems and policy implications in sexual harassment research. *Population Research and Policy Review* 9 (3): 235–254.
- Jensen, Gary F., and Maryaltani Karpos. 1993. Managing rape: Exploratory research on the behavior of rape statistics. *Criminology* 3 (September): 363–385.
- Kanin, Eugene, J. 1957. Male aggression in dating-courtship relations. *American Journal of Sociology* 63 (September): 197–204.
- Kilpatrick, Dean G., C.N. Edmunds, and A.K. Seymour. 1992. *Rape in America: A report to the Nation*. Arlington, Virginia: National Victim Center.
- Kilpatrick, Dean G., Benjamin E. Saunders, Angelynne Amick-McMullan, Connie L. Best, Lois J. Veronen, and Heidi S. Resnick. 1989. Victim and crime factors associated with the development of crime-related post-traumatic stress disorder. *Behavior Therapy* 20 (2): 199–214.
- Kilpatrick, Dean G., Benjamin E. Saunders, Lois J. Veronen, Connie L. Best, and Judith M. Von. 1987. Criminal victimization: Lifetime prevalence, reporting to police, and psychological impact. *Crime & Delinquency* 33 (4): 479–489.
- Kirkpatrick, Clifford, and Eugene Kanin. 1957. Male sex aggression on a university campus. *American Sociological Review* 22 (February): 52–58.
- Koss, Mary P. 1996. The measurement of rape victimization in crime surveys. *Criminal Justice and Behavior* 23 (March): 55–69.
- . 1993a. Detecting the scope of rape: A review of prevalence research methods. *Journal of Interpersonal Violence* 8 (June): 198–222.
- . 1993b. Rape: Scope, impact, interventions, and public policy responses. *American Psychologist* 48 (10): 1062–1069.
- . 1992. The underdetection of rape: Methodological choices influence incidence estimates. *Journal of Social Issues* 48 (1): 61–75.
- . 1988. Hidden rape: Sexual aggression and victimization in a national sample of college students in higher education. In *Rape and sexual assault*, edited by Ann W. Burgess. New York: Garland.
- Koss, Mary P., and Sarah L. Cook. 1993. Facing the facts: Date and acquaintance rape are significant problems for women. In *Current controversies on family violence*, edited by Richard J. Gelles and Donileen R. Loseke. Newbury Park, California: Sage Publications.

Koss, Mary P., and Christine A. Gidycz. 1985. The sexual experiences survey: Reliability and validity. *Journal of Consulting and Clinical Psychology* 53 (3): 442–443.

Koss, Mary P., Christine A. Gidycz, and Nadine Wisniewski. 1987. The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students. *Journal of Counseling and Clinical Psychology* 55 (2): 162–170.

Koss, Mary P., and Christine Oros. 1982. Sexual experiences survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology* 50 (3): 455–457.

Lane, Katherine E., and Patricia A. Gwartney-Gibbs. 1985. Violence in the context of dating and sex. *Journal of Family Issues* 6 (1): 45–49.

Lehnen, Robert G., and Wesley G. Skogan, eds. 1984. *The National Crime Survey: Working papers, Volume 2: Methodological issues*. NCJ 90307. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

———. 1981. *The National Crime Survey: Working papers, Volume 1: Current and historical perspectives*. NCJ 75374. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Leidig, Marjorie W. 1992. The continuum of violence against women: Psychological and physical consequences. *Journal of College Health* 40 (January): 149–155.

Lyberg, Lars, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. 1997. *Survey measurement and process quality*. New York: John Wiley & Sons.

Lynch, James P. 1996a. Clarifying divergent estimates of rape from two national surveys. *Public Opinion Quarterly* 60 (Fall): 410–430.

———. 1996b. Understanding differences in the estimates of rape from self-report surveys. In *From data to public policy: Affirmative action, sexual harassment, domestic violence, and social welfare*, edited by Rita J. Simon. Lanham, Maryland: University Press of America.

Muehlenhard, Charlene L., and Melany A. Linton. 1987. Date rape and sexual aggression in dating situations: Incidence and risk factors. *Journal of Counseling Psychology* 34 (2): 186–196.

Muehlenhard, Charlene L., Susie C. Sympson, Joi L. Phelps, and Barrie J. Highby. 1994. Are rape statistics exaggerated? A response to criticisms of contemporary rape research. *Journal of Sex Research* 31 (2): 144–146.

Parrot, Andrea, and Laurie Bechhofer. 1991. *Acquaintance rape: The hidden crime*. New York: John Wiley & Sons.

- Percy, Andrew, and Pat Mayhew. 1997. Estimating sexual victimisation in a national crime survey: A new approach. *Studies on Crime and Crime Prevention* 6 (2): 125–150.
- Perkins, Craig A., Patsy A. Klaus, Lisa D. Bastian, and Robyn L. Cohen. 1996. *Criminal victimization in the United States, 1993*. NCJ 151657. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Rand, Michael, James P. Lynch, and David Cantor. 1997. *Criminal victimization, 1973–95*. NCJ 163069. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Rivera, George F., and Robert M. Regoli. 1987. Sexual victimization experiences of sorority women. *Social Science Research* 72 (1): 39–42.
- Roiphe, Katie. 1993. *The morning after: Sex, fear, and feminism on campus*. Boston: Little, Brown and Company.
- Russell, Diane E.H. 1984. *Sexual exploitation: Rape, child sexual abuse, and workplace harassment*. Beverly Hills: Sage Publications.
- . 1982. The prevalence and incidence of forcible rape and attempted rape of females. *Victimology: An International Journal* 7 (1–4): 81–93.
- Sanders, Stephanie A., and June M. Reinisch. 1999. Would you say you “had sex” if . . . ? *Journal of the American Medical Association* 281 (3): 275–277.
- Schwartz, Martin D. 1998. Methodological issues in the use of survey data for measuring and characterizing violence against women. Background paper for Workshop on Building Data Systems for Monitoring and Responding to Violence Against Women, U.S. Department of Justice, National Institute of Justice and Bureau of Justice Statistics, and U.S. Department of Health and Human Services, National Center for Injury Prevention and Control and National Center for Health Statistics, 29–30 October, Arlington, Virginia.
- Schwartz, Martin D., and Victoria L. Pitts. 1995. Exploring a feminist routine activities approach to explaining sexual assault. *Justice Quarterly* 12 (1): 9–31.
- Searles, Patricia, and Ronald J. Berger. 1987. The current status of rape reform legislation: An examination of State statutes. *Women’s Rights Law Reporter* 10 (Spring): 25–43.
- Shrout, Patrick E., Bruce G. Link, Bruce P. Dohrenwend, Andrew E. Skodol, Ann Stueve, and Jerrold Mirotznik. 1989. Characterizing life events as risk factors for depression: The role of fateful loss events. *Journal of Abnormal Psychology* 98 (4): 460–467.
- Smith, Michael D. 1994. Enhancing the quality of survey data on violence against women: A feminist approach. *Gender and Society* 8 (1): 109–127.

———. 1987. The incidence and prevalence of women abuse in Toronto. *Violence and Victims* 2 (3): 173–187.

Spohn, Cassia, and Julie Horney. 1992. *Rape law reform: A grass roots revolution and its impact*. New York: Plenum Press.

Taylor, Bruce M., and Michael R. Rand. 1995. The National Crime Victimization Survey redesign: New understanding of victimization dynamics and measurement. Paper presented at the 1995 Annual Meetings of the American Statistical Association, 13–17 August, Orlando.

Tjaden, Patricia. 1996. *Violence and threats of violence against women in America: Female questionnaire*. Denver: Center for Policy Research.

Tjaden, Patricia, and Nancy Thoennes. 1998a. *Prevalence, incidence, and consequences of violence against women: Findings from the National Violence Against Women Survey*. Research in Brief, NCJ 172837. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

———. 1998b. *Stalking in America: Findings from the National Violence Against Women Survey*. Research in Brief, NCJ 169592. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

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

U.S. Department of Justice, Bureau of Justice Statistics. 1997. *Criminal victimization in the United States, 1994*. NCJ 162126. Washington, D.C.

———. 1994. *Criminal victimization in the United States, 1992*. NCJ 145125. Washington, D.C.

———. 1989. *Criminal victimization in the United States, 1987*. NCJ 115524. Washington, D.C.

U.S. Department of Justice, Bureau of Justice Statistics and Federal Bureau of Investigation. 1997. *Implementing the National Incident-Based Reporting System: A project status report*. NCJ 165581. Washington, D.C.

Wyatt, G.E., and S.D. Peters. 1986. Methodological considerations in research on the prevalence of child sexual abuse. *Child Abuse and Neglect* 10:241–251.

Measurement and Analysis of Drug Problems and Drug Control Efforts

by Jonathan P. Caulkins

Drug problems are complex, and determining the best combination of drug control interventions is not always intuitive. Hence, there is a need for rigorous, even quantitative analysis of their effectiveness. This essay is a progress report on the state of the still-developing art of quantitative analysis of the effectiveness of drug control interventions.

Some limitations of existing data are first identified and discussed. They include the reliance on self-reports; the indirect relationship between available indicators and the underlying quantities of greatest interest; and an overemphasis on measures of drug use at the expense of other factors, such as externalities associated with drug control efforts. Four encouraging trends are the ongoing expansion of traditional data systems, improving information about drug markets, greater integration across data sources, and better data from other countries.

Although the relevant data are highly imperfect, they have been adequate to support initial efforts to quantify the effectiveness of a range of drug control interventions. Which interventions are most effective depends on what one defines as the objective of drug control. Available evidence concerning one objective—reducing the quantity of drugs consumed—is reviewed and found to contain key insights but

Jonathan P. Caulkins is a Professor of Operations Research and Public Policy at the H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, and is on the staff of RAND's Drug Policy Research Center.

A
B
S
T
R
A
C
T

391

also to be wanting in important respects. There is a need for better information concerning interactions between different drugs and drug markets, interactions with other domains of social policy, how interventions' effectiveness varies over the course of a drug epidemic, and how epidemics emerge and how they can be controlled in their early stages. These limitations are best viewed as a challenge, not as an excuse for basing policy on less formal or ad hoc syntheses of the literature. Drug policy is not alone in demanding creativity in the adaptation and application of quantitative analysis to evaluate effectiveness. Other policy domains in which benefit-cost or cost-effectiveness analysis is now accepted went through a similar, formative stage.

Illicit drugs impose significant costs on the United States, on source and transshipment countries and, increasingly, on other industrialized countries to the point that, in Stares' (1996) terms, drugs have become a "global habit." A variety of drug control strategies exist, the drug system is complex, and the best combination of strategies is not intuitively obvious.

When choices have important consequences that are difficult to assess, a natural response is to turn to careful analysis. This paper is, in some sense, a progress report on the status of our collective capacity to think carefully about how best to design the mix of drug control strategies.

That is an ambitious goal, and it is important to state at the outset what will and will not be covered here. The focus will be on outcomes more than on principles or morals. Some people approach drug issues from a moral perspective and believe the best drug policy is the one that advances or is consistent with a particular set of principles regardless of the consequences (cf. Wilson 1990). The focus in this paper instead is on understanding the costs and consequences of pursuing various policies.

For the most part, the discussion concerns aggregate costs and consequences. For someone who represents a particular constituency—whether defined by geographic, demographic, ethnic, or other commonalities—the best way to design drug control strategies might be different.

Likewise, this paper does not address the politics of drug policymaking. There are literally thousands of drug policymakers in the United States who represent a wide variety of interests. Sometimes the best policy from a mythical social planner's perspective is not feasible given the structure of the decisionmaking institutions. But this paper abstracts from such institutional considerations.

Not all careful thinking is quantitative, but this paper focuses on quantitative measures and arguments. Some of that focus is no doubt due to the predilections and training of the author, but there is a more fundamental reason for this bias. Drug-related behavior is sufficiently complex and sufficiently diverse so that it is rarely possible to make unambiguous qualitative statements about the nature of outcomes. If one can think of several reasons why a given policy should have one effect, one can usually think of at least one working in the opposite direction. In such circumstances, one must either be perpetually agnostic (on the one hand it might help, on the other hand it might hurt), or one must argue that one set of effects is larger than another. Comparing the magnitudes of effects is inherently a quantitative undertaking, and a premise of this paper is that if one is going to make quantitative arguments, it is best done explicitly. Furthermore, we are interested not only in the effect but also in the

effect relative to the magnitude of the intervention. That is, cost-effectiveness, not just effectiveness, is important.

Policy analysis comprises equal parts basic and applied science. As in basic science, one wants to measure, describe, and understand a problem. As in applied science, one wants to determine how to ameliorate the problem. Drug policy analysis is, in this regard, no different than other forms of policy analysis, so it makes sense to discuss these perspectives in turn. The next section discusses measuring the drug problem. The section after that examines what is known about the effectiveness of strategies for addressing the drug problem.

Existing data systems are reasonably adequate for describing patterns and trends but generally are incapable of explaining them, in part because opportunistic instead of random samples and the absence of control groups makes it difficult to tease out causal relationships.

Measuring the Drug Problem

Understanding the potential and limitations of existing data systems

Richard Hamming said “The purpose of computing is insight, not numbers.” The same could be said of data collection. Unfortunately, in the drug policy arena we have an abundance of numbers, but the glass of insight is at best half full.

We know quite a bit about drug offenders within the criminal justice system but much less about their activities on the street. We know quite a bit about how many drug users there are but little about why there are so many. In contrast, we understand why people sell drugs but know little about how many upper level dealers there are, let alone how they operate.

One could be outraged that officials leave these gaps in knowledge unfilled and publish implausible estimates of such basic quantities as the dollar value of global drug trade. (See Reuter’s [1998] debunking of the United Nations International Drug Control Programme [1997] estimate of \$400 billion per year.) One could also marvel that we know so much about a black market activity that is subject to severe sanctions. Certainly we have better estimates of the number of drug users than we do of the number of prostitutes or ivory traders.

Regardless, it is important to understand what we do know and what we do not. Fortunately, there are many thorough reviews of relevant data systems

(Executive Office of the President, Office of National Drug Control Policy [ONDCP] 1990; Ebener, Feldman, and Fitzgerald 1993; Ebener and Weidmer 1994; Executive Office of the President, ONDCP 1998b), associated estimation techniques (e.g., Hser et al. 1992), and discussions of their policy significance (e.g., Haaga and Reuter 1991; Reuter 1993; Anglin, Caulkins, and Hser 1993; Ebener, Saner, and Anglin 1995).

The goal here is not to provide another data catalogue but to identify and discuss the implications of weaknesses in current data systems. This list of weaknesses is not intended to be comprehensive. Rather, the focus is on less obvious issues that are sometimes overlooked. The existence of published data can lull an observer into thinking better information is available than is actually the case.

For example, we actually know much less about the nature and magnitude of drug control efforts than reading the national drug control budget summary might suggest (Executive Office of the President, ONDCP, 1989, 1997, 1999). This national budget is really only a Federal budget; it excludes State and local spending and, thus, more than half of all drug control spending. Furthermore, the budget is not a budget in the familiar sense of a proactive plan for dividing a pool of available resources. Rather, it is a cross-cut budget, an ex-post accounting of what portion of various agencies' activities further national drug control goals. For the Drug Enforcement Administration (DEA), this apportionment is simple, but deciding what portion of U.S. Customs activity should be attributed to drug control as opposed to interdicting violators of intellectual copyright or crop-threatening foreign insects is another matter (Murphy 1994).

On the other hand, we have a fair idea of how many retail drug sellers there are even though no official data series tracks that number. If one believes retail cocaine sales in the United States are approximately \$40 billion per year (Executive Office of the President, ONDCP 1997), that a full-time retail cocaine seller grosses approximately \$50,000 per year, and there are approximately three individuals selling for every two full-time equivalent sellers (Reuter, MacCoun, and Murphy 1990), then approximately 1.2 million people sold cocaine in the United States last year. That figure easily could be off by a factor of 2, but it is still useful. Knowing that cities such as San Diego and San Antonio have approximately 5,000 cocaine sellers helps one assess the likely impact of a crackdown that arrests 200 or 300 sellers (cf. Kleiman 1997b).

More generally, existing data systems are reasonably adequate for *describing* patterns and trends but generally are incapable of *explaining* them, in part because opportunistic instead of random samples and the absence of control groups makes it difficult to tease out causal relationships. But discovering and quantifying causal relationships is exactly what is most needed. Policymakers

would like analysts to be able to predict the consequences of various interventions, but doing so requires understanding causal linkages, something that current systems do not do well. For example, we have good descriptions of the correlation between drug use and crime but only rough estimates of how much crime rates would change if drug prices rose enough to drive down use by 10 percent.

Limitations of self-reports

The heavy reliance on self-report measures is a fundamental challenge. The appeal of surveys is obvious. They are a familiar research tool, they hold the prospect of getting a representative sample within some sampling frame, and they allow one to gather related information about categories such as employment or health insurance status. The associated estimates are even accompanied by confidence intervals that create the illusion that potential errors can be neatly bounded. However, respondents have no selfish incentive to give accurate answers and they can have real or imagined reasons not to. Drug users may fear their responses will be used against them or be ashamed of their use. Users and nonusers alike may not give accurate answers to questions such as “What is your household income?” if answering requires substantial effort, e.g., because the household has multiple sources of income or the respondent does not handle the family finances.

Nonresponse bias can be at least as problematic. If a nonrepresentative sample is interviewed, the results could be biased even if everyone interviewed gives accurate answers. For example, if truants are more likely to use drugs than other students, then school-based surveys may underestimate use by youths registered for school. Likewise, if people who travel for their job—whether as truck drivers or management consultants—use more, less, or different drugs than persons who spend more time at their principal residence, household survey findings may be biased. The particular concern with household surveys is that subpopulations that are more difficult to sample—e.g., because their residences are transient or they are suspicious of government data collection efforts—may use at higher than average rates.

A quick glance at the “Drug-Related Data” appendix to the *National Drug Control Strategy* (Executive Office of the President, ONDCP 1999) gives a sense of the magnitude of the problem. Table 2 of the 1999 *Strategy* (p. 114) reports that there were 1.5 million past-month cocaine users in 1995. Table 3 (p. 114) reports that in the same year, 3.6 million people used cocaine weekly and another 3.1 million used cocaine less than weekly. These tables are obviously inconsistent, and the second table’s estimates are well outside the 95 percent confidence interval associated with the 1.5 million estimate (U.S.

Department of Health and Human Services 1996). The explanation is that the first is based exclusively on the National Household Survey on Drug Abuse (NHSDA); the second augments NHSDA data with data from the Drug Use Forecasting (DUF) system (now called the Arrestee Drug Abuse Monitoring [ADAM] program). Similarly, NHSDA respondents can account only for approximately one-tenth of the cocaine believed to be consumed based on supply side estimates (Rydell and Everingham 1994). Thus, the only comprehensive survey of drug use in the general population misses the majority of use of the most problematic illicit drug.

Limitations of prominent systems that do not rely on self-reports

Not all data systems depend on self-reports. Notably, DUF/ADAM supplements interviews with urinalysis testing. Because arrestees consume the majority of the “expensive” drugs (cocaine, heroin, and methamphetamines), this makes DUF/ADAM a particularly useful data system. However, DUF/ADAM has limitations. Perhaps the biggest is that samples are not a random or even a representative sample of arrestees. They are not representative in part because of various sampling quotas (e.g., those arrested for drug-law violations have been undersampled in the past). More fundamentally, DUF/ADAM samples only from booked arrestees who are in the booking facility at the time of data collection. Hence, people who are arrested but not booked or not detained after booking are not included, and the probability of an individual being sampled depends on how long he or she stays in the booking facility. The magnitude of the distinction becomes clear when one looks at demographic statistics. More than half of those sampled by DUF/ADAM are black, even though less than one-third of people arrested in the United States are black.

DUF/ADAM urinalysis data are also problematic because they, like many drug-related data series, are only an indicator, not a direct or even proportional measure of the phenomena of greatest interest. This point is best illustrated by example. For many reasons, we would like to track changes in the quantity of drugs used by the criminally involved. Sometimes DUF/ADAM data are used for this purpose, but they are fundamentally limited in rather obvious ways. Ideally, if a data series does not directly measure the phenomenon of interest, one would like it at least to be proportional to that phenomenon. When proportionality holds, one can draw inferences such as, “The indicator declined by 10 percent, so a reasonable point estimate is that the underlying behavior declined by 10 percent as well.” Proportionality might hold if the behavior in question did not influence the probability of being sampled. But that is clearly violated in the case of DUF/ADAM and drug use by the criminally involved

because drug use affects the probability of being arrested. Indeed, one of the very reasons we are so interested in drug use by the criminally involved is precisely because drug use can have a criminogenic effect.

It is not even clear that drug use and DUF/ADAM data always move in the same direction. Suppose drug prices jumped. Presumably that would reduce drug use, but in the short run, the elasticity of demand may be less (in absolute value) than unity. If so, spending on the drug would go up. In as much as economic-compulsive crime is driven by the amount spent on drugs, this might increase economic-compulsive crime and resulting arrests. If the decline in use takes the form of reduced consumption per use session (e.g., because purity is lower), not less frequent use, then the probability of a drug-involved offender testing positive given arrest might not change appreciably. So, unless the price increase caused a comparable increase in the rate at which nonusers are arrested, this scenario would lead to an increase in the proportion of offenders who test positive, even though the quantity consumed by drug-involved offenders declined.

There are similar problems with the Drug Abuse Warning Network (DAWN), which monitors the number of drug-related emergency room (ER) episodes by retrospectively examining ER records in a sample of non-Federal, general care, short-stay hospitals that operate 24-hour emergency departments. For each recorded episode, DAWN gathers up to four substances of abuse, user demographics, the user's reason for using the drug and visiting the ER, the route of administration, and the source of the substance.

From an analyst's perspective, DAWN has limitations (Caulkins, Ebener, and McCaffrey 1995), such as DAWN's definition of "drug related" not being synonymous with "caused by."¹ It excludes drug-related morbidity associated with HIV and much drug-related violence. Furthermore, if intoxication leads to the assault of a nonuser, a drug user injures a victim in the course of a robbery, or a dealer injures another dealer (who is not a user) in a battle over turf, those injuries are not drug related in the DAWN sense of the term.

Likewise, the number of drug-related episodes is not proportional to the amount of drug use because the number of DAWN mentions per user or per gram can vary. For example, Newmeyer (1999) observes that between 1980 and 1998 the heroin-using population in San Francisco roughly doubled, but DAWN mentions for heroin increased tenfold. Many factors influence whether and how frequently users visit ERs, including mode of administration, use patterns, the user's socioeconomic and demographic characteristics, the location and availability of ERs, ER policy, and availability of alternative sources of care. These factors can vary by location as well as over time. For example,

although roughly 60 percent of injection drug users (IDUs) in New York City are HIV positive, a much smaller proportion in Los Angeles are infected. If IDUs with AIDS may be more likely to visit the ER than are other IDUs, there might be more DAWN episodes per IDU in New York than in Los Angeles. Hence, one must be cautious when interpreting differences in the number of ER episodes across cities as representing differences in the number of users.

Measurement error

One problem with some data systems is that accurate measures are not vital to the individuals involved in the data collection. For example, DEA maintains price information in its System to Retrieve Information from Drug Evidence (STRIDE). As discussed subsequently, knowledge of drug prices and trends is of enormous value to analysts, but STRIDE was not designed primarily to track prices (Frank 1987). This may explain why greater care has not been taken to eliminate outliers.²

Likewise, DAWN data are not collected by hospitals in order to give good medical care. ER staff are under enormous time pressure. Determining and accurately recording for subsequent analysis which drugs the patient may have used, why, in what form, and how the drugs were obtained is not always a priority. Perhaps not surprisingly, a variety of studies have found significant inaccuracies and underreporting in DAWN (Ungerleider et al. 1980; Roberts 1996). Brookoff, Campbell, and Shaw (1993) found that not one of the 82 trauma patients studied who tested positive for cocaine was recorded by DAWN—even though the hospital in question had recently been formally audited by DAWN and found to be in complete compliance with DAWN guidelines.³

Issues of definition and interpretation

Inconsistencies in definitions and interpretation can also be problematic as estimates of drug-related deaths illustrate. A common source of such estimates is the National Vital Statistics Survey (NVSS); recent figures are in the vicinity of 15,000 deaths per year. This relatively low number is often contrasted with estimates of 100,000 deaths per year due to alcohol and 400,000 deaths per year due to cigarettes, but the figures are not comparable.⁴ The alcohol and cigarette figures include chronic effects such as liver disease and lung cancer, whereas the NVSS figures include only acute effects such as overdose and poisoning. Deaths due to ill health resulting from addiction and dependency are not included, nor are deaths from HIV and homicide, two of the principal sources of drug-related mortality.

We know there are several million children living in households with a substance-dependent adult, and child protective services case loads grew along with the growth of the cocaine epidemic. However, drug policy discussions generally give short shrift to the harms borne by family members and friends of the drug abusers.

Comparable issues emerge with estimates of the social cost associated with drug use (e.g., Harwood, Fountain, and Livermore 1998). These numbers provide a sense of magnitude and can have real policy implications. For example, Tragler, Caulkins, and Feichtinger (forthcoming) argue that the preferred level of drug control spending depends on the magnitude of the associated social costs.⁵ However, it is important to recognize what these figures are not. For one, they are not the same as budgetary costs. If government interventions succeeded in reducing the social cost of drug use by \$10 billion, that would not increase taxes or the gross domestic product or any other identifiable account by \$10 billion.

Second, changes over time in the estimates reflect changes in methodology as well as changes in drug use or drug problems. These changes are not repudiations of the earlier analyses. Rather, there has been a gradual improvement in the ability to quantify components of the overall cost. (Compare Harwood, Fountain, and Livermore 1998, and Rice et al. 1990.)

Third, the studies often ascribe dollar values to morbidity and mortality in ways that weight more heavily losses to wealthy people. For example, Harwood, Fountain, and Livermore (1998) use the value of foregone earnings (the human capital approach). Hence, these methods would imply that shifting drug abuse from affluent to impoverished citizens would reduce social costs.

Focus on drug use at the expense of other drug-related problems

Another weakness of existing data is their focus on drug use. That may sound paradoxical, but it is only paradoxical if one equates drug problems with drug use, and a variety of problems stem more directly from drug distribution and drug control than from drug use (MacCoun and Reuter forthcoming). If all aspects of drug problems grew and shrank in direct proportion to the level of use, this would be a distinction without a difference, but that is not the case. Indeed, different measures of use are not always well correlated.⁶

One can speculate about the reasons for this focus. U.S. drug control goals have typically focused on drug use reduction, a point discussed further in a

later section. There has been greater emphasis on research and evaluation of demand reduction programs than of supply control programs (Reuter 1997). Drug users are the ultimate source of all drug-related problems; if no one wanted to use drugs, there would be no drug supply industry or need for drug control efforts.

Whatever the reason, this overemphasis on drug users in data collection is problematic. It focuses attention on the problems of drug users to the exclusion of other issues. We know there are several million children living in households with a substance-dependent adult, and child protective services case loads grew along with the growth of the cocaine epidemic.

However, drug policy discussions generally give short shrift to the harms borne by family members and friends of the drug abusers. This deficit may at least in part be explained by the absence of a national database tracking instances of drug-related child abuse or neglect.

Likewise, little attention is paid to the families of the approximately 400,000 people incarcerated for drug offenses. It seems likely that the number of children separated from a parent by this incarceration is in the tens of thousands, but we do not track that number or even think much about whether those separations are good or bad for the children.

More generally, little effort is made to track nondollar costs of drug control efforts. Some are important but difficult to quantify. Drug control efforts in source and transshipment countries affect those countries and our relations with them in ways that are of interest not only for altruistic reasons. The level of drug-related corruption in Mexico and its impact on Mexico's democratic institutions is of considerable importance to the United States, given that Mexico is our third-largest trading partner (in 1997), we share a 1,936-mile border, and many millions of Americans have direct family ties in Mexico. Other costs could be quantified readily, including the number of search warrants served on incorrect addresses, the value of assets seized but ultimately returned because the owner was not found guilty, the number of individuals stopped and searched under suspicion but who were not found to be in possession of any drug or in violation of any law, and so on.

We tend to measure what is easy to measure, not just what is important. That is understandable, and one can argue that some information is always better than no information. However, there is a tendency to ascribe zero value to that which is not measured and for tangible but less relevant numbers to drive out consideration of less tangible issues.

Looking at the "Drug-Related Data" appendix of the *National Drug Control Strategy* illustrates the point (Executive Office of the President, ONDCP 1999). The appendix contains 33 tables of drug-related data distributed across categories as indicated in exhibit 1.

Many aspects of the drug problem are not reflected in these data tables. They say nothing directly about impacts on family members and friends of drug users, drug-related crime or violence, morbidity and mortality caused by drugs (as opposed to that which is merely drug related), the nondollar costs of drug control efforts, and so on. We tend to measure what is easy to measure, not just what is important. That is understandable, and one can argue that some information is always better than no information. However, there is a tendency to ascribe zero value to that which is not measured and for tangible but less relevant numbers to drive out consideration of less tangible issues (Larkey and Caulkins 1991).

Recent developments in drug-related measures

Despite these problems, the existing data systems have strengths. There are multiple systems, which is valuable for triangulation. The systems are well funded. The major systems have been collecting data consistently over time. Many provide data at the local as well as national level. Furthermore, four promising trends portend better data in the future.

(1) Expansion in size, scope, and consistency of mainstream data systems

Many of the traditionally most useful data series are getting better. DAWN was among the first to benefit from this trend. In the late 1980s and early 1990s, the number of metropolitan areas covered was reduced to 21, but a national panel of facilities located throughout the remainder of the United States was added. DAWN was redesigned as a stratified random probability sample that permits calculation of site-specific and national estimates as a weighted sum of the episodes occurring in the region's sampled facilities, where the weights are recomputed each quarter.

The Monitoring the Future (MTF) survey added samples of 8th- and 10th-grade students in 1991. The expanding length of its 12th-grade data series (dating to 1975) is making it increasingly valuable.

The DUF/ADAM program is being improved by (1) making sampling procedures more consistent across sites, (2) recording the sampling probability of every

Exhibit 1. Types of data included in the *National Drug Control Strategy's* data appendix

Number of tables	Subject	Data sources
15	Prevalence as measured by the "Big Three" surveys	National Household Survey on Drug Abuse, Monitoring the Future, Drug Use Forecasting/ Arrestee Drug Abuse Monitoring
4	Number of school dropouts, use by dropouts, societal costs of dropping out of school	U.S. Census, Youth Risk Behavior Survey, Mark Cohen's studies (see his chapter in this volume)
2	Drug-related morbidity and mortality	Drug Abuse Warning Network, National Vital Statistics Survey
8	Magnitude of drug control efforts: Numbers of arrests, incarcerations, seizures, and treatments	Uniform Crime Reports, Bureau of Justice Statistics, Drug Enforcement Administration, Federal-wide Drug Seizure System, National Drug Abuse Treatment Utilization Survey, Substance Abuse and Mental Health Services Administration
4	Market aggregates: Global production, drugs available in United States, prices, and drug spending	International Narcotics Control Strategy Report, Drug Enforcement Administration, System to Retrieve Information from Drug Evidence, Office of National Drug Control Policy (Rhodes et al. 1995)

arrestee in the sample, and (3) extending coverage to (in most cases) the county containing the city in question (Rhodes 1998). In addition, the core instrument is under revision and the number of sites is scheduled to increase to 75.

Plans are under way to expand the NHSDA sample. Early NHSDAs surveyed only 4,000 to 9,000 individuals. The sample size increased to approximately 30,000 in 1991 and is expected to grow to 70,000 in 2000, allowing prevalence estimates to be produced at the State level.

(2) Emergence of information about markets

Illicit drugs are, ultimately, consumer goods, and like other goods in modern societies they are provided primarily through markets. The markets for illicit drugs have distinctive characteristics with implications for social welfare and drug policy choices (Kleiman 1992). For example, the markets are characterized by small, short-lived, vertically unintegrated, and technologically unsophisticated sole proprietorships that generate great violence and disorder. Nevertheless, drug markets are, in fact, markets.

A market-clearing equilibrium is characterized by its price and quantity, but price data have only recently been a focus of attention. Studying prices is appealing because price data are relatively abundant and can be disaggregated to the municipal or even neighborhood level (Weatherburn and Lind 1997).

Price data support a number of important observations. Perhaps the most striking is simply that drug prices are extraordinarily high. The cocaine and heroin sold in retail markets in the United States are quite literally 10 to 100 times more valuable per unit weight than gold (Caulkins and Reuter 1998).

Illicit drug prices also vary widely across market levels, between locations, over time, and from transaction to transaction. The simplest expression of the variation across market levels is that there are substantial quantity discounts for a wide variety of illicit drugs. Specifically, price per unit is proportional to transaction size raised to a negative power (Caulkins and Padman 1993).

Quantity discounts exist for many consumer goods, but the discounts are far larger for illicit drugs, and they have important implications. They help us understand how costly drug seizures at different levels are to drug suppliers. They imply most of the accounting profits accrue not to a handful of kingpins but rather are divided among a large number of low-level dealers. They imply that the effectiveness of high-level enforcement depends crucially on the manner in which price increases at one market level are translated into increases at lower levels (a point discussed further later in this chapter). They imply that one cannot estimate the dollar value of the retail market just by multiplying the quantity consumed by the price per gram; one also needs to consider the distribution of retail purchase sizes.

Data on prices also help elucidate characteristics of markets. For example, Caulkins (1995) found that drug prices within the United States increase as one moves away from the drug sources and that prices are lower in larger markets. For cocaine in particular, the data support the notion that cocaine is distributed through an urban hierarchy, in which large cities tend to be leaders, with drugs diffusing down through layers of successively smaller surrounding communities.

Price data also have raised questions not yet answered. For example, exhibit 2 shows how the inflation-adjusted price of cocaine and heroin fell by 75 percent between 1981 and 1988, even though enforcement intensity increased during that period. There are many theories but little agreement concerning why enforcement failed so dramatically to keep prices high.

Prices are not the only new source of information about drug markets. To continue the market analogy, drug suppliers are like businesses. Like business, the more sophisticated supplier organizations keep financial books. Levitt and Venkatesh (1998) used financial records from drug-dealing gangs to draw interesting inferences about their behavior, but on the whole, these records are a promising but largely untapped data resource.

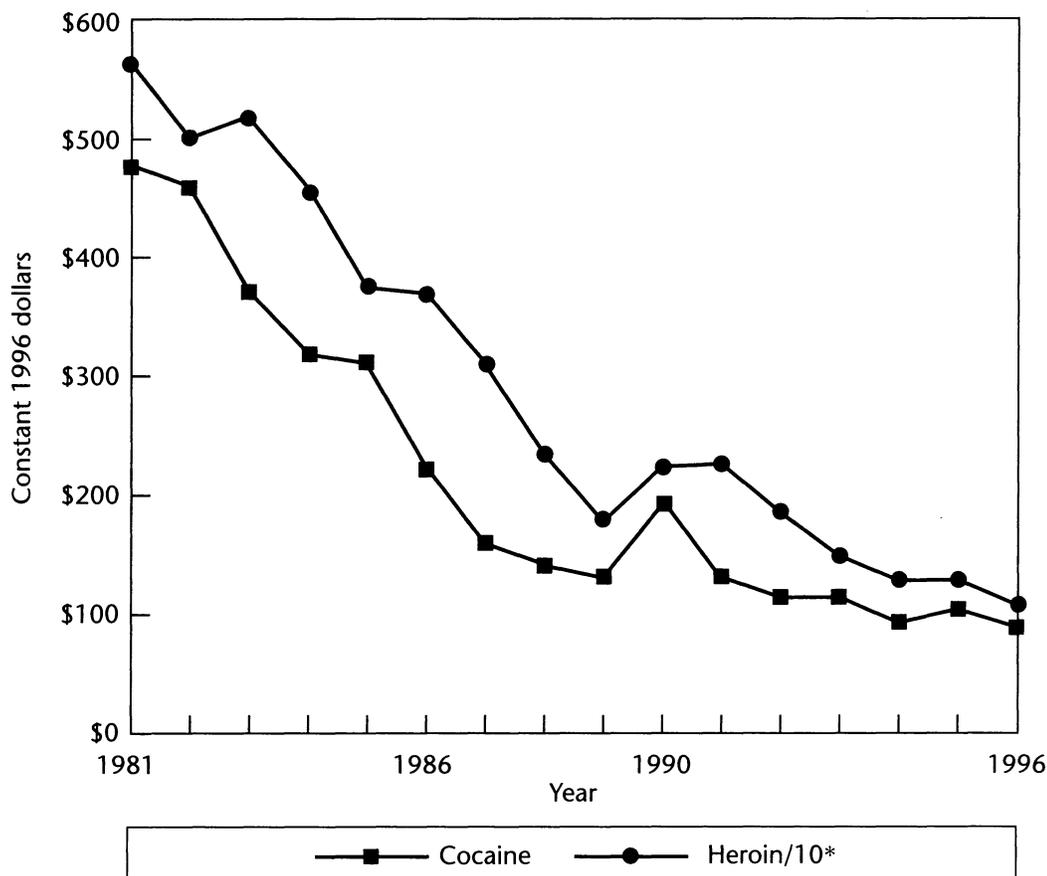
Another source of information about drug markets is ethnographic studies. Traditionally, ethnographic studies focused on drug addicts. Those addicts frequently participated in low-level selling, but selling activity was not central to the investigations. More recently, qualitative researchers (e.g., Johnson et al., 1985, 1991; Maher 1997) have been focusing explicitly on dealers and resulting data are being used in formal analyses of their organizational and business structures (Caulkins et al. 1999; Tita 1999).

(3) Data integration

Drug-related data series have not always been integrated into a coherent systems view. This disconnectedness is problematic for a variety of reasons, including the fact that it foregoes important opportunities to check the validity of numbers. As a result, numbers are repeated, trusted, and used, which are, on closer examination, simply absurd.

Max Singer (1971) was perhaps the first to draw attention to this problem with his article "The Vitality of Mythical Numbers." He noted that conventional wisdom about the number of heroin addicts and the amount of crime committed per addict were incompatible with estimates of the amount of crime that was actually committed. In particular, although it was generally assumed that addicts stole \$2 billion to \$5 billion per year in New York City, Singer showed that the total amount stolen in the city was about 10 times less, suggesting that estimates of the

Exhibit 2. Retail cocaine and heroin prices per gram



* Heroin prices divided by 10 to fit scale.

number of addicts were seriously in error. As Singer (p. 9) observed, “The main point of this article may well be to illustrate how far one can go in bounding a problem by taking numbers seriously, seeing what they imply, checking various implications against each other and against general knowledge.”

Unfortunately, this message has been slow to take hold. Thirteen years later, Peter Reuter (1984) wrote a followup piece (“The [Continued] Vitality of Mythical Numbers”), which showed that the situation had not substantially improved. Only more recently have coherent systems views been constructed (e.g., Homer 1993a, 1993b; Childress 1994a, 1994b; Dombey-Moore, Resetar, and Childress 1994) that allow us to reconcile apparently conflicting trends in indicators (Reuter, Ebener, and McCaffrey 1994) and add up all the costs in the domestic cocaine distribution industry (Caulkins and Reuter 1998). These integrative estimates have not yet banished all mythical or nonsensical numbers, but they do provide a much firmer foundation on which to base subsequent policy analyses.

(4) Expansion of international comparisons and databases

In the past, drug-related data from outside the United States were weak to nonexistent. That was understandable. It costs about the same amount to run a household survey in a country of 27 million people as it does in the United States with its population of 270 million. But the tax base for paying for data collection is obviously larger in a larger country. Also, drug problems in the United States had been much more severe than elsewhere. They remain more severe, but the gap is narrowing (Taylor and Bennett 1999).

Increasing drug use around the world is obviously an undesirable development, but inasmuch as it has spurred greater investment in data collection and reporting abroad, it has a beneficial side effect for the United States. Better data collection abroad is valuable because it provides two types of variation that are helpful for understanding relationships among variables.

First, there is greater variation in drug policy across countries (for example, within Europe) than there is across jurisdictions within the United States. Because the ultimate objective of policy analysis is to estimate the impacts of varying policy, this is of direct value (MacCoun, Reuter, and Schelling 1996; MacCoun and Reuter 1997, and forthcoming).

Second, there are variables that affect drug use, vary over time, and are difficult to control for (e.g., attitudes toward drugs). They make panel datasets much more valuable than simple time series, but, unfortunately, some important data series are not collected at the subnational level. For those, comparisons across countries can be insightful.

The United Nations International Drug Control Programme in Vienna, Austria, has long produced information on different countries' drug problems, and the quality of the data and associated publications has risen dramatically in recent years. Data on seizures, prices, purities, and so on are now available for many countries and multiple years (United Nations International Drug Control Programme 1998).

These four trends do not exhaust the list of promising developments. It is easy to get frustrated about the inadequacies of the data. But given the intrinsic difficulty of gathering information about covert activities, the progress that has been made in the past decade and the progress that can be expected in the next decade are heartening. Furthermore, even if the glass of insight is only half full, there is no reason not to make use of that half.

Solving the Problem

Having addressed the measurement of drug problems, the discussion turns next to the measurement or estimation of the effectiveness of drug control efforts. We open by considering alternative drug control goals and how the choice of goals interacts with the apparent effectiveness of drug control strategies. The bulk of the section reviews what is known about the effectiveness of interventions with respect to a particular goal, namely reducing drug use. The section closes by identifying themes for subsequent research.

Alternative goals of drug policy and their interaction with program effectiveness

To evaluate an intervention, one must first know its objective. A vigorous literature concerning the goals of drug policy has emerged since the passage of the 1988 Anti-Drug Omnibus Control Act. That act requires ONDCP to establish “long-range goals for reducing drug abuse in the United States” and “short-term measurable objectives.” There has been an ongoing expansion and evolution in the named goals, culminating in last year’s establishment of the national drug control strategy performance measurement system (Executive Office of the President, ONDCP 1998a). Other countries, including the United Kingdom, Canada, and Australia, have also established official national drug policy goals.

There is an interesting interplay among the nature of data systems, the types of goals, and the apparent effectiveness of different drug control strategies. Some data systems are more appropriate for certain goals than others. (NHSDA can measure progress on goals pertaining to overall prevalence of drug use, but other systems are better at tracking the number of certain types of heavy users.) Likewise, some data systems are more appropriate for evaluating certain control strategies than others. (City-level data on prices and search times are invaluable for evaluating enforcement interventions; individual-level data from controlled trials are most valuable for evaluating treatment and prevention programs.) Perhaps least appreciated is the interplay between the choice of goals and the apparent effectiveness of control strategies. Some goals inherently favor certain control strategies relative to others, as will be illustrated by considering some possible goals and noting at least one such bias for each.

Reducing prevalence

The first *National Drug Control Strategy* (Executive Office of the President, ONDCP 1989, 8) stressed goals that pertained to the prevalence of drug use among the general population and youths. In particular, 7 of the 10 goals related to NHSDA or MTF measures.

Setting goals in terms of overall prevalence weights all drug users equally. Because it is easier to persuade a light user to quit than it is to persuade an addict to do so, these goals inherently favor interventions that target light use. Drug testing of athletes and workers might convince some light users to quit but would presumably have less impact on unemployed street addicts. Likewise, media campaigns, tough rhetoric, and symbolic actions that stigmatize drug use might convince some recreational users to abstain. They are less likely to convince an addict to quit, and the stigmatization might even hurt addicts. It is perhaps not entirely coincidental that during the late 1980s, the number of drug users in the United States fell sharply, but the number of heavy users increased.

A case can be made for placing greater emphasis on the smaller number of most problematic users. Such an emphasis favors a different set of tactics. Inasmuch as addiction is a chronic relapsing condition, programs that discourage people from becoming heavy users are appealing. Primary prevention, which discourages initiation into drug use in the first place, is relevant for reducing both the overall prevalence and the prevalence of heavy use, but secondary prevention programs that seek to reduce escalation from light to heavy use are relevant only for the latter.

Reducing the amount of drugs used

Because heavy users consume much more per capita than light users, tracking changes in consumption is not altogether dissimilar from tracking changes in the number of heavy users, but it has the advantage of not entirely ignoring light users. In particular, focusing on consumption effectively assigns a weight to each user proportional to that user's rate of consumption.

There is evidence that treating heavy users is the most cost-effective way to reduce the quantity of cocaine consumed (Rydell and Everingham 1994), but given the high relapse rates (Anglin and Hser 1990a), it is almost certainly not the most cost-effective way to reduce the number of cocaine users. If the goal is to reduce the quantity of drugs consumed, then interrupting an addict's drug use for 3 months is valuable.⁷ Indeed, it could well be more valuable than

Setting goals in terms of overall prevalence weights all drug users equally. Because it is easier to persuade a light user to quit than it is to persuade an addict to do so, these goals inherently favor interventions that target light use. It is perhaps not entirely coincidental that during the late 1980s, the number of drug users in the United States fell sharply, but the number of heavy users increased.

convincing some light users to quit permanently.⁸ But if the goal is to reduce the number of people who are addicted to drugs, short-term interruptions are meaningless.⁹

Reducing the amount spent on drugs

Not all drug-related problems are driven by the amount of drug use. Some are driven by the amount spent on drugs. Consider drug-related crime and violence. Goldstein (1985) categorized drug-related crime and violence as psychopharmacological, economic-compulsive, and systemic. Only the first is driven directly by drug use or withdrawal. Economic-compulsive crime clearly is driven by spending on drugs. Systemic crime may be as well. Caulkins et al. (1997) estimate that about five-sixths of cocaine-related crime is economic-compulsive or systemic. Likewise, the amount of money laundering, corruption, and demoralization of law-abiding citizens by the affluence of dealers may be related more closely to the dollar value of the drug market than to the weight of drugs consumed.

Changing the goal from reducing the quantity of drugs consumed to reducing the dollar value of the drug market favors demand-side interventions relative to enforcement programs that reduce use by driving up prices. Indeed, if the (absolute value of the) elasticity of demand were less than one, then, although driving up prices would reduce drug use, it would actually increase the dollar value of drug sales. Note that not all enforcement operates principally by driving up prices. Setting a goal of reducing the dollar value of the black market would penalize high-level enforcement, including interdiction and source country control efforts, more than it would retail enforcement or partnerships with treatment because high-level enforcement reduces use primarily by driving up prices.

Harm reduction versus use reduction objectives

Another goal, one more popular outside the United States, is reducing the total damage done by drugs, not the amount of drugs used. Logically, one would like to call this social utilitarian approach *harm reduction* and contrast it with *use reduction*. The term *harm reduction* will be used in this sense here even though it has been made controversial by being used to mean other things in other contexts.

MacCoun and Reuter (forthcoming) give a useful categorization of drug problems and their relationship to use. The magnitude of the overall drug problem is some agglomeration of intoxication-based functional impairment, numbers of overdoses, amounts of drug-related crime and violence, and so on. These components of the drug problem do not always move in proportion to each other or to consumption.

If the goal is to reduce the harm caused by drugs, not the amount of drugs used, then a variety of interventions merit consideration that otherwise would not be relevant. The most famous (or infamous) is needle exchange. Needle exchanges are unlikely to have much impact on the amount of drug use. Advocates argue that they bring addicts in contact with social services and hence improve treatment outcomes. Critics argue that needle exchanges send the wrong message. But neither effect is likely to be large relative to the impact on the rate of HIV transmission. Because each new case of HIV is very costly (for society as well as the individual), if the goal were to reduce harm, needle exchange at a minimum deserves consideration. And the consensus in the scientific literature seems to be that it can actually succeed in reducing HIV transmission (e.g., Lurie et al. 1993; National Research Council [NRC] 1995).

A less discussed and perhaps less controversial example can also make the point. Inasmuch as IDUs are at risk for becoming infected with HIV and transmitting it to others, distributing condoms to IDUs might plausibly reduce the harm caused by injection drug use. It is hard to argue that doing so would have any substantial impact on the amount of drug use. Because condoms are cheap relative to the cost of treating HIV, distributing condoms to IDUs is probably cost effective if the goal is to reduce harm but worthless if the goal is to reduce use.

Harm reduction itself can be defined in a variety of ways. For example, should harm suffered by adult drug users be included in the sum of harm to be minimized? Some argue the government has neither the responsibility nor the right to protect people from themselves (Friedman and Szasz 1992). Others argue that "drug consumers may be less capable than other consumers of protecting their own interests" (Kleiman 1992). Enforcement makes users worse off by imposing sanctions, raising prices, reducing availability, and increasing variability in potency. In contrast, treatment programs help users break the cycle of addiction and avoid some of the harshest consequences of use. Hence, one is more likely to favor treatment over enforcement if harm experienced by users is part of the harm to be minimized.

MacCoun (1996) distinguishes between *microharm* reduction that seeks to reduce the average harm users experience and *macroharm* reduction that seeks to minimize aggregate societal harm. A microharm reduction perspective is particularly hostile to enforcement. A macroharm reduction perspective recognizes harm to nonusers. Thus, improving the quality of foster care and property crime control strategies (such as enhancing sanctions for burglary) might be useful from a macroharm reduction perspective, but irrelevant or even counter-productive from a microharm reduction perspective.

If one asks simply whether these various programs are effective at reducing drug use, the answer is in most cases yes. But such an answer is not very informative. Not every program that reduces drug use is worth pursuing, any more than every program that reduces pollution or improves health should be pursued. One has to ask whether the benefits are large enough to justify the costs of undertaking the program.

To summarize, there are a variety of possible goals for drug policy, and the choice of goals affects the apparent desirability of various interventions. Hence, as the discussion turns to a review of estimates of the effectiveness of interventions, it must be explicit about the goals or outcome measures. In most cases, that will be the quantity of consumption averted per program dollar, in part because that is the most common measure in the literature and in part because Rydell, Caulkins, and Everingham (1996) argue it is perhaps the single best scalar measure of the magnitude of drug problems.

Current understanding of the effectiveness of drug control interventions

Evidence concerning the effectiveness of drug control interventions is a study in contrasts. We have great faith in, but little evidence concerning, the effectiveness of prevention programs (Gorman 1995; Moskowitz 1993; Caulkins et al. 1999). In contrast, there is great unease but relatively abundant evidence concerning the efficacy of methadone maintenance (e.g., Ward, Mattick, and Hall 1994), and some evidence but great controversy concerning the effectiveness of other types of treatment (Gerstein et al. 1994; Rydell and Everingham 1994; Institute of Medicine [IOM] 1996; Crane, Rivolo, and Comfort 1997).

Paradoxically, there is less evidence about the ability of enforcement programs to control drug use even though they absorb the bulk of drug control resources (Reuter 1997). There are exceptions, including experimental studies of local police interventions (e.g., Weisburd and Green 1995; Tita 1999), model-based analyses of incarceration-oriented efforts (e.g., Caulkins et al. 1997), and model and time series-based analyses of high-level enforcement and interdiction programs (e.g., Crawford et al. 1988; Crane, Rivolo, and Comfort 1997; Yuan and Caulkins 1998). But it is striking how much greater is the ratio of research to program funding for demand control interventions than for supply control interventions.

If one asks simply whether these various programs are effective at reducing drug use, the answer is in most cases yes. But such an answer is not very

informative. Not every program that reduces drug use is worth pursuing, any more than every program that reduces pollution or improves health should be pursued. One has to ask whether the benefits are large enough to justify the costs of undertaking the program.

As Cohen explains elsewhere in this volume, there are at least two ways of thinking about whether the benefits are enough. Benefit-cost analysis seeks to monetize the benefits and compare them with program costs. Cost-effectiveness analysis seeks to compare the benefits per dollar of program cost generated by different programs. Programs that generate the greatest benefit per dollar spent (i.e., that yield the most “bang for the buck”) are viewed as being the most cost effective.

Unfortunately, although much has been written about drug control interventions and drug policy, only a tiny fraction of that literature seeks to provide quantitative estimates of programs’ effectiveness. Some of the literature is not quantitative at all, and a portion of that which is quantitative examines only process measures. Many studies that quantitatively address program outcomes seek only to establish the statistical significance of an effect or do not translate the magnitude of estimated effects into policy-relevant terms. Finding that a program has a statistically significant effect on some outcome is a necessary, but not sufficient, condition for concluding that the program is effective. And knowing that a \$10 million-per-year program reduces drug-related 911 calls for service in a medium-sized city by 5 percent is not enough to determine if it is cost effective. The results need to be translated into measures that allow comparison with other available interventions.

Theoretically, neither benefit-cost nor cost-effectiveness analysis is a good way to inform decisionmaking because both collapse a vector of outcomes into one or two aggregate measures. Ideally, one would leave estimates of the various outcomes disaggregated and apply methods of multicriteria decisionmaking. Realistically, however, this is complicated when there are multiple stakeholders, and it seems that without the discipline of needing to compute some bottom-line assessment, few studies ever even estimate quantitatively the magnitude of the effect of each program on a common and comprehensive set of outcome measures. Hence, the remainder of this section does not seek to review the entire quantitative literature concerning drug control interventions, but rather to focus on the small subset that addresses questions of benefit-cost ratios and cost-effectiveness directly.

Before proceeding, it is useful to specify a framework for categorizing drug control interventions. For present purposes, it is convenient to distinguish among three types of intervention effects:

- Reducing the quantity of drugs consumed.
- Reducing the magnitude of the drug problem per kilogram consumed.
- Displacing the problem from one location, time, or population to another.

One intervention can operate in more than one way. For example, when police shut down an outdoor street market, there might be some reduction in selling and use, some displacement of the selling to another location, and some displacement to more covert forms of dealing that impose fewer harms on neighbors per gram sold.

As mentioned, this review focuses on effectiveness at reducing drug use. Interventions can reduce consumption by reducing demand, constraining supply, or driving a wedge between supply and demand.

Demand can be suppressed by preventing people from initiating or escalating use or by treating current users, with or without assistance from enforcement. A wedge between supply and demand exists to the extent that the costs users pay to obtain and consume drugs exceed the dollar price paid to the dealer. The most-discussed example is that users expend time and effort (search time) locating a dealer and completing a transaction. Presumably, raising these nondollar costs discourages use to some extent.

Interventions can affect supply in two ways. Unanticipated interventions can disrupt supplier operations in ways that upset the market equilibrium. A market is not in equilibrium if (1) supply does not meet demand in the sense that there is a physical shortage, (2) demand does not meet supply in the sense that there is a physical surplus of goods that cannot be disposed of, or (3) the market clears but the market clearing price is changing rapidly. In the absence of ongoing exogenous shocks, one expects this situation to last only until the market has time to adjust. Ideally, the disruption takes the form of physical shortage and the market does not regenerate so the new equilibrium has no drug sales or use, but that is not the norm.

Treatment is the most thoroughly evaluated drug control intervention, and it is usually found to be cost effective.

Enforcement can also affect supply even if the intervention is fully anticipated. For example, if smugglers knew that about one-quarter of all shipments would be seized, they would ship more than they would if they thought none would be seized. So quantity seized is not a direct measure of enforcement's impact on consumption. However, presumably the smugglers would charge more per kilogram landed to make up for the losses. The higher prices

represent a shift in supply that affects retail prices and, hence, consumption. These two distinct aspects of enforcement's effect on supply are referred to here as disequilibrium and equilibrium effects.

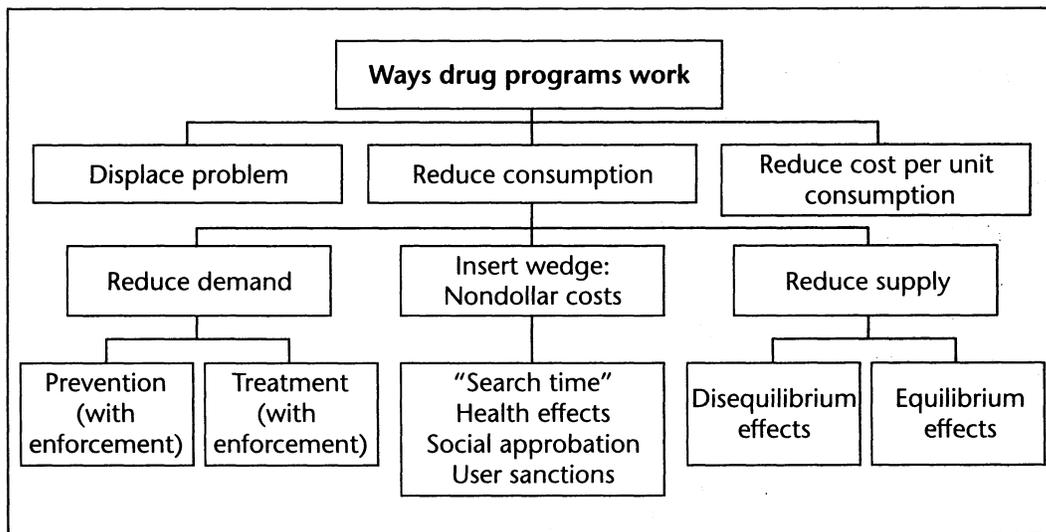
This framework for understanding drug control interventions is summarized in exhibit 3. The discussion of effectiveness uses the framework as a schematic outline.

Reducing demand through prevention

There is great confidence that drug prevention is effective and cost effective. For example, the latest national strategy (Executive Office of the President, ONDCP 1999, 54) states unequivocally, "The simplest and most cost-effective way to lower the human and societal costs of drug abuse is to prevent it in the first place."

Such a monolithic view is simplistic. There is enormous heterogeneity in programs, ranging from school-based curricula to adventure camps to community-wide events to physical fitness-based programs. Some are more effective than others (Sherman et al. 1997). Likewise, some school-based programs have been shown unambiguously to decrease illicit drug use (Ellickson and Bell 1990; Botvin et al. 1995). Yet the evidence concerning the most popular school-based program, D.A.R.E.[®], is less positive. D.A.R.E.[®] has not been shown to have any material effect on marijuana use at followup (Tobler 1997).

Exhibit 3. Framework for understanding drug control programs' effectiveness



There are only a handful of studies that actually estimate cost-effectiveness. For example, in a 1998 National Institute on Drug Abuse Monograph entitled *Cost-Benefit/Cost-Effectiveness Research of Drug Abuse Prevention: Implications for Programming and Policy* (1998), only one of the nine articles produces a specific estimate. In that article, Pentz estimates the drug-related costs averted per dollar spent on the Midwest Prevention Project. Marijuana is the only illicit drug for which she estimates outcomes and she estimates 3-year net reductions of 3.5 percent in heavy marijuana use, tapering off to 2.5-percent reductions at 5-year followup. These reductions, combined with reductions in alcohol and cigarette use and the modest cost of the intervention "indicate that comprehensive drug prevention programs, such as a multi-component community-based drug prevention program, are highly cost-beneficial and cost-effective" (p. 125).

Caulkins et al. (1999) consider a generic "best practice," school-based prevention program modeled on the evaluation data for Project ALERT and the Life Skills Program (Ellickson and Bell 1990; Botvin et al. 1995). They find there is considerable uncertainty concerning prevention's cost-effectiveness, but the range of estimates for cocaine is comparable to that previously estimated for a variety of types of drug enforcement (7 to 60 kilograms of cocaine consumption averted per \$1 million).

I am not aware of any comparable estimate for media-based prevention programs such as those at the centerpiece of the ONDCP prevention campaign launched in summer 1998. Related media-based campaigns, such as those designed to promote the use of designated drivers, have been shown to have an effect (Winsten 1999). ONDCP reports a correlation between media exposure and changes in attitudes, but there is no evidence yet concerning effects on drug use (McCaffrey 1999).

Reducing demand through treatment

Treatment is the most thoroughly evaluated drug control intervention, and it is usually found to be cost effective (e.g., Rydell and Everingham 1994; Gerstein et al. 1994). IOM (1996, 192) summarized the literature by saying "Research has shown that drug abuse treatment is both effective and cost-effective in reducing not only drug consumption but also the associated health and social consequences."

Treatment does not achieve this status by convincing most of those who enter treatment to abstain from all subsequent use. Quite the contrary. Relapse rates are very high. Indeed, many who begin treatment do not even complete the course of treatment. At first glance these observations might seem contradictory,

but Rydell and Everingham's (1994) pioneering systems analysis reconciles them by making several points.

First, focusing on relapse rates ignores the benefits of reduced use while patients are in treatment (an incapacitation effect). Rydell and Everingham show that even if every entrant to a typical cocaine treatment program relapses to heavy use and some (20 percent) use during treatment, treatment can still be cost effective on the strength of the in-treatment effect alone.

Second, the fact that many people drop out of treatment quickly increases relapse rates, but it has only a modest effect on treatment's cost-effectiveness because not many resources are consumed by patients who only stay for a day or two.

Third, careers of heavy drug use are long and costly, so even infrequent long-term successes are sufficient to make a program cost effective. There is uncertainty about specific parameter values, but suppose that the average admission to treatment costs \$2,000, the average heavy user consumes 125 grams per year, the social cost per gram of consumption is \$100, and the net present value of the residual career length for a typical entrant to treatment is 8 years. If the treatment program leads even 1 in 50 entrants to give up heavy use, the program is cost justified.¹⁰ Exactly what figures should be used in these calculations is not known, but the structural points discovered by Rydell and Everingham are clear. A treatment program can be cost effective even if it performs miserably on conventional metrics such as the proportion of entrants who are abstinent 18 months after entry.

At the same time, Rydell and Everingham provide a cautionary note. If most people relapse, then unless those individuals can be re-enrolled rapidly, there is a limit to how quickly treatment can ameliorate the drug problem. In Rydell and Everingham's model (which assumed that 13.2 percent of treatment entrants left heavy use because of that treatment), even if every heavy cocaine user received treatment once a year, cocaine use would still be cut only in half over 15 years.¹¹ Highly imperfect treatment programs, no matter how cost effective, cannot quickly eliminate an endemic drug problem.

These are very general observations. As mentioned, the treatment evaluation literature is large, so it contains many more insights than can be covered here. There are a number of excellent summaries of the treatment literature (e.g., Hubbard et al. 1989; Anglin and Hser 1990a; Gerstein and Harwood 1990; Cartwright and Kaple 1991; IOM 1996, ch. 8; McLellan et al. 1996). Indeed, the literature is so large there is even a bibliography of literature reviews of drug abuse treatment effectiveness (Prendergast, Podus, and McCormack 1998).

Complementing traditional demand reduction with the stick of enforcement

It is common to view demand reduction generally, and treatment in particular, as an alternative to enforcement programs. Indeed, simple schema such as exhibit 3 encourage such an either-or, if not competitive, view. The reality is much more complicated. Enforcement plays an important role in prevention and treatment (Moore 1990). Conversely, prevention and treatment can enhance the effectiveness of enforcement by shrinking enforcement's target, allowing it to take advantage of inherent increasing returns to intensity (cf. Caulkins et al. 1999).

One important interaction between enforcement and demand control is the use of coercive enforcement to compel people into treatment, to keep people in treatment, and/or to create additional incentives for performing well while in treatment (Anglin and Hser 1990b). The general finding in the literature is that paternalistic treatment that combines the carrot and stick is more effective than either the carrot or stick in isolation (Valliant 1997).

Civil commitment is one of the oldest forms of this cooperation between coercive and therapeutic elements. Not all civil commitment programs performed well, but some did (McGlothlin, Anglin, and Wilson 1977; Anglin and McGlothlin, 1984). As Anglin and Hser (1990a, 425) note, "Although the program results were not spectacular, outcomes at the time were as good as or better than those of other interventions for drug dependence."

Treatment in prison is another vehicle for partnership (Hiller, Knight, and Simpson 1999). There are two undeniable appeals to treating prisoners. First, they are serious offenders. Second, the State is already obligated to pay their room and board, so the marginal cost of delivering some treatment modalities (e.g., therapeutic communities) is less than it is outside of prison.

Treatment Alternatives to Street Crime (TASC) is another longstanding example. Approximately 300 programs in the United States use diversionary dispositions (deferred prosecution, community sentencing, pretrial intervention) to direct offenders into treatment (Inciardi and McBride 1991). TASC programs vary from State to State, but in at least some States, they have beneficial effects on the amount of treatment services received, rates of drug use, and HIV risk behaviors (Anglin, Longshore, and Turner 1999).

Among the better known of the carrot-and-stick partnerships are so-called drug courts, which were developed in Miami and Oakland but now exist in approximately 250 jurisdictions throughout the country. Drug courts are like TASC in the sense of suspending normal criminal justice sanctions if, and as long as,

the offender is attending and progressing through treatment. Most evaluations have been process evaluations published as reports, not in peer-reviewed academic journals, but the consensus seems to be that they can be effective at least in providing relatively close community supervision, improving treatment retention, and reducing recidivism.

The most dramatic variation on the carrot-and-stick approach is coerced (or more generally incentivized) abstinence. It suggests that what really matters is not the authority of the judge's robes but rather the incentives given for recovery, particularly the certainty and swiftness of the punishment. Indeed, replacing the judge with a clear and unbending set of rules that mandate punishment for violations could actually improve outcomes by enhancing the certainty of punishment. A variety of clinical evaluations show that frequent testing associated with immediate incentives enhances treatment success (e.g., McCarthy and Borders 1985; Higgins et al. 1993, 1994; Silverman et al. 1996).

Kleiman (1997a) argues that coerced abstinence might plausibly yield substantial reductions in consumption. He observes that the majority of cocaine used in the United States is consumed by people who are nominally under criminal justice supervision (e.g., free on probation or parole). If frequent testing with automatic and immediate sanctions could cut their use by even two-thirds, that would cut national consumption by approximately 40 percent. No other program offers the hope of such dramatic and rapid reductions in cocaine use.¹²

Driving a wedge between supply and demand

The second broad way of reducing drug use is to create a wedge between the demand and supply curves by imposing nondollar costs on users. Raising nondollar costs is appealing because it can discourage use without increasing revenues per unit sold by dealers. There are at least four categories of nondollar costs: the inconvenience of obtaining the drug, adverse physical reactions to the drug, social approbation from individuals and organizations other than the state, and punishment of users by the state. The first has received the most attention.

The search time argument was originally advanced by Moore (1973) and has been refined and applied by Kleiman (e.g., Kleiman and Smith 1990). It recognizes that users expend time and effort to locate a dealer and complete a transaction. For many years the concept was largely unevaluated, in no small part because there were no data on search times. Kleiman (1988) reported that the extreme case of effectively eliminating street markets in a city had beneficial effects, but it was not clear whether incremental expansion in enforcement could have an appreciable effect on quantities consumed.

Rocheleau and Boyum (1994) achieved a substantial breakthrough by showing that data could be collected on retail purchasing patterns generally and on search times in particular. Furthermore, they showed that even experienced heroin users expend substantial effort obtaining their supplies (an average of 35 minutes per purchase), raising hopes that search time costs were large and, hence, driving them up might be a practical way of suppressing use.

However, Caulkins (1998) argued that because the average purchase totaled approximately \$25, unless these addicts placed a high value on their time, the search time costs were still small compared with the dollar costs. Furthermore, because heavy users may know 10 to 20 alternative suppliers and new dealers can be located relatively easily (Riley 1997), it is doubtful that arresting one or even several would greatly increase search time. Quantifying these observations is highly speculative, but Caulkins (1998) estimates that arresting retail dealers of established, mass-market drugs reduces consumption by experienced users through increased search times by less than one-tenth as much as it could reduce consumption through the "risks and prices" mechanism described later in this chapter. The ratio may be even lower for enforcement directed at targets further up the distribution chain.

There are a variety of ways in which one might drive a wedge between the demand and supply for a drug. All have the considerable theoretical appeal of suppressing consumption without increasing dollar prices. None have been studied satisfactorily, and few have been studied in a quantitative way at all.

This does not mean that increasing search times is never effective. It is more likely to be effective in smaller towns, "thinner" markets, and/or with newer users who have not established alternative sources of supply. Such individuals are presently responsible for only a small portion of consumption, but if increased search times can suppress initiation, in the long run it might have a greater effect.

Another and perhaps larger nondollar cost that users pay is the physical or health costs of addiction and adverse reactions to the drugs or impurities mixed with the drugs. Theoretically, the government could seek to make drugs more dangerous, and, if reactions to the 1970s paraquat scare are any indication, doing so might reduce use. However, except for the voluntary taking of an antagonist such as naltrexone, such interventions are at best questionable ethically. Indeed, when a batch of particularly dangerous drugs hits the streets, the usual response is to warn users in order to reduce adverse physical reactions.

Social approbation and informal social controls are a third form of nondollar cost. It is generally recognized that people are constrained from committing most criminal acts not so much by the fear of criminal sanction but rather by informal social controls. Probably the most common punishments for being caught with illicit drugs are shame, ostracization, loss of trust, and so on, not arrest, because the individuals who do the catching are more likely to be parents, spouses, or children than police officers.

With the exception of encouraging the widespread use of drug testing, e.g., of athletes and employees (NRC 1994; Lemmens 1997), it is not clear whether official policy has much control over these nondollar costs of drug use. Taking a tough line on drugs (Partnership for a Drug-Free America ads, government resistance to needle exchange, draconian sentences for drug sellers, and so on) might stiffen the resolve of parents and spouses to deal "severely" with infractions. (Severely is in quotes because even the more severe informal sanctions do not appear severe when compared with the official sanctions for drug selling.) However, there is little empirical evidence concerning this conjecture.

User sanctions are the fourth category of wedges between supply and demand. Little effort has been devoted to estimating the effectiveness or cost-effectiveness of user sanctions. I once produced a rough estimate (unpublished) of the cost-effectiveness of incarcerating heavy cocaine users. The estimate was much more favorable than I had anticipated but still inferior to the results of comparable calculations for domestic enforcement and treatment.

Incarceration is not, however, the only or likely the most efficient form of user sanction. Theoretically, fines should be very cost effective because they are just a transfer, imposing no net cost on society. Realistically, collecting fines is an expensive and uncertain prospect. I am not aware of any study that quantifies how effective fines are at controlling drug use in practice or of parallel estimates for imposing community service or seizing users' cars or other assets.

In summary, there are a variety of ways in which one might drive a wedge between the demand and supply for a drug. All have the considerable theoretical appeal of suppressing consumption without increasing dollar prices. None have been studied satisfactorily, and few have been studied in a quantitative way at all. Perhaps such investigations are an unintended casualty of the predilection toward simplistic partitions of drug policy into supply and demand control programs, with minimal attention devoted to those that do not fall neatly into one box or the other.

Reducing supply through disequilibrium effects

Supply side interventions are most likely to have disequilibrium effects if they quickly affect a large proportion of supply and/or the supply chain. Because the drug supply sector in the United States for most drugs is populated by many vertically disaggregated firms, it is difficult for enforcement to remove a large proportion of the capacity of the domestic distribution network at any one time. In the language of reliability studies, the network is robust because of its many lateral linkages, independent paths, and ability to expand quickly the capacity of individual arcs. The greatest potential for disruption may exist for less common drugs and/or smaller markets. The number of suppliers of LSD is much smaller than the number of cocaine suppliers, so it might be more feasible for a moderate number of strategically coordinated enforcement operations to create a shortage of LSD than of cocaine. Likewise, it should be easier for a given level of enforcement effort to generate a shortage of cocaine in a small, isolated city (Boise) than in a large city (New York) or a small city that is close to a large city (Hartford, Connecticut).

There is sometimes greater potential with interventions in source countries because there is greater market concentration there. Perhaps the greatest success was when the combination of the Turkish opium ban, the breaking of the "French Connection" case, and Mexican opium eradication efforts substantially drove up purity-adjusted heroin prices during the mid- to late 1970s before Asian heroin filled the gap (Reuter 1985). The greatest success with cocaine was the result of a combination of U.S. efforts and the war between the Colombian government and the Medellin-based traffickers in 1989 that led to a sharp (50 to 100 percent at its peak) but short-lived (about 18-month) increase in cocaine prices (Caulkins 1994). More recently, in 1995, Peruvian interdiction of the air bridge to Colombia led to a smaller but identifiable increase in cocaine prices (Crane, Rivolo, and Comfort 1997).

These transient price increases can have meaningful effects. The heroin scarcity in the 1970s is temporally correlated with the ebbing of the heroin epidemic. ER and medical examiner mentions declined in parallel with higher cocaine prices in 1989–90 (Executive Office of the President, ONDCP 1992), and there was a one-period decline in emergency mentions in late 1995 (U.S. Department of Health and Human Services 1998). The price increases and associated benefits do not last, however, because suppliers react to market disruptions by modifying their tactics and operations (Reuter 1988). At one time or another over the past 25 years, four different regions have been the principal supplier of heroin to the United States (Mexico, South America, Southwest Asia, and Southeast Asia). Similarly, Colombia quickly replaced Mexico as the principal

supplier of marijuana to the United States in response to paraquat spraying and fears of adverse health-effects of using sprayed marijuana (Kleiman 1992).

It is important to ask whether short-term disruptions in drug markets are desirable. Moore (1990) argues that it is beneficial to create shortages that last long enough for a birth cohort to mature through its “window of vulnerability” to initiation during periods of relative scarcity. Shortages might also drive experienced users into treatment. However, short-term price spikes can increase sellers’ revenues and, perhaps, drug-related crime. (Even if the long-run price elasticity of demand is larger, in absolute value, than -1 , it is unlikely that the short-run elasticity is so large.)

Boyum (1992) argues that periodic shortages interspersed with times of relative availability could lead to greater use than would stable supplies if the demand curve is “kinked” in the sense of having an asymmetric response to price increases and decreases.

Unfortunately, the literature quantifying the magnitude of the benefits of these short-term disruptions relative to the cost of creating them is meager.

Reducing supply in the long-run equilibrium

It is believed that enforcement against suppliers can reduce consumption by driving up the price of drugs in equilibrium. This theory is embodied in so-called “risks and prices” calculations of the sort pioneered by Reuter and Kleiman (1986). The risks-and-prices paradigm recognizes that increasing enforcement risks for dealers raises their cost of doing business. Dealers could simply absorb those costs but presumably prefer to pass them along to users in the form of higher retail prices. Drug users, like consumers of other goods, respond to higher prices by reducing consumption (Caulkins and Reuter 1998).

The literature on risks and prices calculations for cocaine generates a number of insights. First, expanding domestic enforcement is probably less cost effective at reducing cocaine use and associated problems than is expanding even modestly effective treatment programs for heavy users (Rydell and Everingham 1994).

Second, some types of domestic enforcement generate greater effects through the risks-and-prices mechanism than do other types. In particular, mandatory long sentences are less cost effective than conventional sentences, and enforcement against dealers such as those prosecuted at the Federal level is more cost effective than is enforcement directed at typical dealers (Caulkins et al. 1997). (Retail enforcement may, however, be more effective at controlling the problems associated with markets.)

Third, because retail prices are much greater than those at higher market levels, the ability of enforcement directed at higher market levels to raise retail prices crucially depends on how price increases at one level are transmitted to lower levels. The additive model predicts that increasing high-level prices by \$1/gram will increase the retail price by about \$1/gram. The multiplicative model predicts that increasing high-level prices by 1 percent will increase the retail price by 1 percent (Caulkins 1990; Boyum 1992).

To illustrate why these two models have vastly different implications for the efficacy of interdiction, suppose the import and retail prices are initially X and $10X$, respectively. Would a control program that drove the import price up to $2X$ significantly reduce consumption? According to the additive model, the retail price will rise by the same amount as the import price, so it will rise only from $10X$ to $11X$ —a modest increase.¹³ According to the multiplicative model, if import prices double, retail prices will double, from $10X$ to $20X$, a large enough increase to appreciably affect consumption.

Hence, a necessary condition for high-level enforcement to suppress consumption substantially by driving up retail prices in equilibrium is for the multiplicative model to hold. Caulkins (1990, 1994) and Boyum (1992) analyze historical price data that are more consistent with the multiplicative than the additive model, but they do not test the multiplicative model directly and their data are primarily from kilogram-level transactions and below. Near the origins of the distribution chain the multiplicative model does not seem plausible because there are considerable variations in coca leaf prices that are not paralleled (even with a lag) by retail prices. No analyses have been done for intermediate market levels (between export from Colombia and the kilogram level within the United States) because of insufficient data.

Fourth, examining the cost structure of the drug distribution industry suggests that only one-quarter to one-third of the economic (not just dollar) cost of distributing drugs is directly attributable to enforcement (Caulkins and Reuter 1998). This lends empirical support to Reuter's (1983) conjecture that there are substantial structural consequences to product illegality. That is, illegality plus a modicum of enforcement is sufficient to make prices much higher than they would be if legal. Additional increments in enforcement have smaller incremental effects on price. MacCoun and Reuter (forthcoming) also find evidence for this proposition of diminishing returns to increasing enforcement intensity in their review of historical and cross-cultural analogies to the drug problem in the United States.

Assessment of the very primitive state of quantitative assessments of effectiveness

The review in the previous section is sobering. When it comes to quantitative estimates of drug control interventions' ability to reduce drug use, much more is unknown than known, and much of what is known is predicated on models and theories that are plausible but not empirically validated. With respect to the other two ways interventions operate (displacing the problem and reducing the magnitude of the problem per unit of use), even less is known. It would be easy to look at this situation and conclude that quantitative assessment of drug control interventions is a quixotic undertaking that is best abandoned.

There are two strong arguments for viewing the inadequacy of past efforts as a challenge, not a warning. First, policy decisions will be made. The choice is to have them informed by partial analysis or no analysis, and flying blind is not appealing.

Second, there are other policy arenas in which quantitative assessments seemed impossible at first but which, over time, were refined to the point of being useful and widely employed. Benefit-cost textbooks point to dams and other water projects as classic examples of problems that are amenable to benefit-cost analysis, but before the first such analysis was done, the analysis certainly would have seemed anything but routine.

Perhaps medical interventions are a closer analogy. There now is a standard approach to doing cost-utility analysis (CUA) of medical interventions (see Kamlet 1992; Gold et al. 1996), but that is a recent development. Twenty years ago, it would have been hard to imagine that such analyses would be either possible or accepted.

The medical CUA example is also instructive because it was not clear at the outset what the measure of effectiveness should be. Possibilities included cost per life saved, cost per life-year saved, and measures of social value saved per dollar spent (i.e., benefit-cost ratios). All have merits, but a consensus has emerged around cost per quality-adjusted-life-year (QALY) saved.

Similarly, there are many possible goals of drug policy and, hence, many possible measures of effectiveness. Work to date has focused on consumption averted per \$1 million spent, but that choice should be seen as a point of departure, not as the only possibility.

Emerging themes and gaps in the current literature

Over the past 10 years, significant understanding has accumulated concerning the effectiveness of drug control interventions. However, as Goethe observed, "Doubt grows with knowledge." We now know enough to know that there is a great deal we do not know. This section identifies four key areas that need further research.

(1) Interdrug effects

Polydrug use and "double-breasted selling" are common, and there are any number of reasons why interactions between drugs might modulate the effectiveness of drug control interventions. However, current understanding of these effects is rudimentary at best.

Perhaps the most basic question is whether different drugs are complements or substitutes. The classical notion of complements and substitutes pertains to cross-price-elasticities of demand.¹⁴ That is, does increasing the price of one drug increase or decrease consumption of another? If increasing the price of one drug increases consumption of a second drug, the drugs are called substitutes; otherwise they are complements.

If one thinks of different drugs as being different ways to achieve an altered state of consciousness, one might expect them to be substitutes. Evidence from earlier studies seemed to support this view among at least some pairs of psychoactive substances. DiNardo and Lemieux (1992), Model (1993), and Thies and Register (1993) found evidence that marijuana and alcohol were substitutes. Thies and Register also found some evidence that marijuana and cocaine were substitutes but had inconsistent results with respect to alcohol and cocaine. However, all of the marijuana findings arise from variables reflecting marijuana's criminalized/decriminalized state, not its price.

More recent evidence suggests that drugs are at least as likely to be complements. Saffer and Chaloupka (1995, 1999) found this for various pairings of alcohol, cocaine, and marijuana, except for marijuana and alcohol. (Heroin results were inconclusive.) Chaloupka and Laixuthai (1997) and Pacula (1998b) confirmed this finding and extended it to include marijuana and alcohol. Pacula (1998a), Chaloupka et al. (1999), and Farrelly et al. (1999) found complementarity among beer, cigarettes, and marijuana. These studies just skim the surface, however, because of weaknesses in measures of both price (particularly for marijuana) and consumption (overreliance on past-month prevalence instead of actual rates of consumption).

As MacCoun, Reuter, and Schelling (1996) point out, use of one substance can increase demand for consumption of another substance through a variety of mechanisms, not all of which are pharmacological (cf. DeSimone 1998). For example, regular use of one substance may expose one to sellers of another substance or to a peer subculture that has pro-use attitudes. It may well be the case that drugs may be short-run substitutes, particularly for those who are already established polydrug users, and long-run complements, with complementary effects dominating when one focuses on initiation.

At any rate, understanding these linkages is of the utmost importance. As Rydell, Caulkins, and Everingham (1996) note, if drugs are substitutes, it would undercut the appeal of drug control efforts that operate by driving up prices. Driving up prices might reduce use of the drug in question, but if it leads to wholesale substitution into other drugs, then overall harm could go up or down depending on the amount of substitution and the relative dangers of the different substances.

Likewise, Kleiman (1992) argues that whether one thinks relaxing marijuana enforcement is a good or a bad idea depends in no small part on one's beliefs about how it would affect the use of other drugs. An optimist might hope that alcohol and marijuana were strong substitutes, so marijuana decriminalization might reduce alcohol use and its social costs, perhaps leading to a net reduction in drug-related harm.¹⁵ A pessimist might fear that the historical correlation between the use of marijuana and other substances is causal. If so, then marijuana decriminalization might lead to substantial increases in the use of all drugs.

Interdrug effects are not confined to issues of demand. The size and nature of the market for one substance can influence the development of the market for another. There is a notion that marijuana growing in Appalachia has built on a tradition of moonshining that dates at least to the alcohol prohibition of the 1930s (Weisheit 1990). Likewise, some cocaine smugglers (e.g., Carlos Lehder) apparently got their start smuggling marijuana. The benign description of the relationship is that smuggling expertise developed with one product helped smugglers move other products. The pessimistic description is that when the United States stepped up interdiction efforts in the early 1980s, they were differentially effective against marijuana, which is bulkier and easier to detect than cocaine. Inasmuch as drug markets and drug market participants are known to adapt to variation in enforcement pressure, it is at least plausible that there was a causal connection between increased interdiction effort and the growth of cocaine smuggling.

If one analyzed each drug market in isolation, a result that increased enforcement caused increased trafficking would seem perverse. But if one recognizes the interactions across markets, it is a plausible story that can be told at various market levels. In 1973, the Knapp Commission exposed widespread police corruption in drug investigations (primarily heroin, at the time). The New York City Police Department responded by directing officers not to arrest street-level drug sellers (Kelling and Coles 1996). Over the next 10 years, New York City witnessed an explosion in street selling of cocaine and marijuana,¹⁶ although the connection may or may not be causal.

Similar stories can be told with respect to prevention and treatment. In the 1960s, overreliance on fear-arousal tactics directed at marijuana may have undercut the credibility of government warnings concerning the dangers of other substances, including heroin. Cocaine abuse by methadone maintenance patients is a problem that is almost as old as methadone maintenance itself (Chambers, Taylor, and Moffett 1972). The optimistic interpretation is that this merely reflects preexisting patterns of use. The pessimistic interpretation is that some individuals respond to methadone maintenance by switching from heroin use to cocaine use, with no net reduction in the aggregate use of expensive drugs.

It is easy to generate examples of the effectiveness of policy interventions dependent on interdrug effects. They all have the same character. It is plausible that the program is beneficial, and it is plausible that the program has no or greatly reduced effects. Which is correct depends on interdrug effects, and sufficient empirical evidence to be confident of which story is more accurate simply does not exist. Clearly, interdrug effects are a worthy topic for further research, both in the narrow sense of estimating cross-elasticities of demand and in the more general sense of improving understanding of how drug markets and control programs directed at one market interact with those of other drugs.

(2) Variation in drug control policy over time¹⁷

Musto (1999) notes that historically there have been alternating periods of greater and lesser drug use. In particular, a cycle of quiescence, rapid escalation, plateau, and gradual decline has been observed for a number of drugs including crack (Golub and Johnson 1997). These cycles are often referred to as drug “epidemics,” and the epidemic metaphor is appropriate because the rapid escalation involves “contagious transmission” of drug use from one person to another.

Because drug problems evolve over time, it seems plausible that drug control policy should as well. However, it is rare to hear someone say we should follow one policy or another now because it would be particularly effective at this

point in the epidemic or to argue that a policy once was effective but now should be scaled back. Instead, most such opinions are voiced without explicit qualification about the times for which the policy is sensible.

Initial efforts to enrich understanding on this dimension have taken the form of rather abstract modeling exercises. One thread in the literature has emphasized interventions with local street markets. Based on a variety of empirical reports (e.g., Kleiman 1988), Caulkins (1990, 1993) developed a dynamic model that describes how a local drug market might respond to intensive enforcement operations. The model generates suggestions for how such crackdowns should be managed, but the model itself is purely descriptive. Baveja et al. (1993, 1997), Naik and colleagues (1996), and Kort et al. (1998) extended it by adopting a prescriptive approach. The general finding is that the simple strategy of either accommodating a market or using the maximum available enforcement until the market has collapsed is optimal in most instances. However, this line of research considers only a single type of drug control—namely local crackdowns—and its effects on a specific local market.

A second, distinct thread has sought to address what might be called strategic policy questions from a national perspective. These questions include determining the best division of resources among competing drug control programs (such as prevention, treatment, and various types of enforcement) and whether enforcement should be directed at users or sellers.

Again the early work was primarily descriptive, addressing effectiveness indirectly (Schlenger 1973; Levin, Roberts, and Hirsch 1975; Gardiner and Shreckengost 1987; Homer 1993). The first explicit efforts to develop prescriptive models were made by Rydell (1997) and a group of researchers at Austria's Vienna University of Technology (e.g., Dawid and Feichtinger, 1996; Gragnani, Rinaldi, and Feichtinger 1997).

The image of a mature market is one in which the density of market participants is great enough to support "professional" transactions. In contrast, in a social network market, most sales are transacted between individuals who have reasons for contact other than consummating the transaction. For example, the seller and buyer may be friends, coworkers, neighbors, or schoolmates who share the same routine activities. Street markets in which buyers and sellers may not even know each other are clearly mature markets.

Tragler, Caulkins, and Feichtinger (forthcoming) consider an intertemporal decision model where the government wants to minimize the sum of social costs caused by drug use and expenditures on two controls: price-raising enforcement against dealers and treatment. According to this model, it is usually best to rely primarily on enforcement at first in order to keep prices high and suppress initiation. Enforcement spending should increase as the number of users grows, but not nearly as fast in percentage terms as treatment spending. Hence, treatment should receive a larger share of control resources when a drug problem is mature than when it is first growing.

Behrens et al. (1999, forthcoming) is a complementary effort that focuses on prevention and treatment. They extend Everingham and Rydell's (1994) model of cocaine use to make initiation increasing in the number of light users and decreasing in the number of heavy users. The insights suggested by this model include:

- Prevention is most appropriate when there are relatively few heavy users, e.g. in the beginning of an epidemic. Treatment is more effective later.
- The transition period when it is optimal to use both prevention and treatment is very brief.
- Total social costs increase dramatically if control is delayed.

These studies are just initial efforts. In some, certain controls are not treated, and all of them use crude aggregations that disguise prevailing heterogeneities. Hence, these modeling efforts need to be refined and extended, and they need to be complemented by both more empirical and more qualitative studies.

(3) Detection and control of an emerging market

One finding of the foregoing modeling is that detecting the onset of a drug epidemic quickly is valuable. This stresses the importance of having a good understanding of the early stages of drug epidemics. Likewise, the European Monitoring Centre for Drugs and Drug-Addiction (EMCDDA) has identified that a better understanding of initiation into drug use is one of its priority questions. However, little is known about how to describe and analyze the initiation of a drug epidemic compared with what is known about how to analyze mature markets of illicit drugs. The lack of understanding is easy to explain. Mature markets (such as the cocaine market in the United States) are easier to observe and study than are markets that are embedded in social networks (such as the cocaine market in some European cities and MDMA [Ecstasy] markets in the United States). Likewise, aggregate models in which market participants are

not individually distinguished are easier to analyze than models that explicitly recognize the structure of social networks. These models are just beginning to be applied to drug distribution (cf. Carley 1990 and Zeggelink 1995).

The image of a mature market is one in which the density of market participants (both sellers and buyers) is great enough to support "professional" transactions. In contrast, in a social network market, most sales are transacted between individuals who have reasons for contact other than consummating the transaction. For example, the seller and buyer may be friends, coworkers, neighbors, or schoolmates who share the same routine activities. Street markets in which buyers and sellers may not even know each other are clearly mature markets. So are typical crack house and beeper sales.

The term mature reflects a hypothesis that drug use initially spreads within social networks and only emerges into professional markets when the market reaches some critical size.¹⁸ It is an empirical question whether the majority of drug epidemics is characterized by a transition from a social network to a mature market. What is beyond question, though, is that we have very limited capacity to predict epidemics early enough to take action to prevent the rapid spread of use during the infectious stage of the epidemic. For example, the United States fully appreciated that cocaine was a severe problem around 1984, but by then, initiation had already grown from an average of 40,000 per year in the mid-1960s to an average of 1.4 million per year in the 6 years preceding 1984 (Johnson et al. 1996). Likewise, we know little about how the effectiveness of different interventions depends on the character of the market. It seems plausible that efforts to drive up search time would be more effective in the early stages of market development, but this remains a conjecture.

(4) Interaction with other policy areas

Drug policy is often construed as being either self-contained or a subset of crime control policy. An alternative view is that drug policy should be a subset of medical or public health policy. It is not clear that any of these is accurate inasmuch as they underappreciate the extent to which drug policy affects outcomes in other policy areas and, conversely, policies and outcomes in other areas affect drug-related problems (Boyum and Reuter, forthcoming).

Perhaps the most concrete example comes from prevention. There is a literature on drug prevention, but it recognizes that programs designed or funded to prevent one type of delinquent behavior often affect an array of such behaviors (Karoly et al. 1998), including violence, gang participation, teen pregnancy, and dropping out of school.

As another example, drug enforcement has civil rights implications, the implicit if not explicit use of profiling being just one concern. Racial disproportionality in incarceration is most severe for drug offenses (Blumstein 1993) and the rapid expansion in incarceration for drug offenses has played an important role in the expansion in incarceration generally, particularly for minorities. This expansion has implications for everything from labor force participation to voting rates to demographic outcomes, including family structure.

Likewise, source and transit zone interdiction activities interact with foreign policy. The interaction with efforts to counter insurgencies and build democratic institutions has long been recognized. (See, for example, Steinitz 1985, Garcia Arganaras 1997.) Historically, the lament was that drug control objectives were subordinated to other foreign policy objectives (a notorious example being the Air America operations in Southeast Asia). More recently, the tables have been turned. For example, the certification process created by the Anti-Drug Abuse Act of 1986 requires the administration to identify foreign countries as cooperating or not in drug control. Although the sanction provisions have rarely had a direct effect on U.S. foreign assistance, the entire process clearly has implications (largely negative) for U.S. foreign relations (Falco 1995; Drug Strategies 1998).

Conversely, phenomena outside drug policy can affect drug use and drug control efforts. For example, the movement toward managed care and the growth in the population with no health insurance has affected financing for drug treatment. Likewise, Stares (1996) argues that globalization of commerce and transportation have substantially enhanced smugglers' ability to deliver drugs, the North American Free Trade Agreement being a particularly relevant example for the United States. The spread of HIV among IDUs has affected drug use (Caulkins and Kaplan 1991), and (especially outside the United States) HIV/AIDS policy has affected drug policy.

Among the most complicated intersections of drug and social policy are questions of actions the government does or does not take to help the more vulnerable individuals in our society. Parental substance abuse is involved in a large proportion of cases of child abuse and neglect. Very often, alcohol is involved, but often, illicit substances are involved as well. Impulses toward zero-tolerance and family preservation are not the only things that conflict in such complicated situations.

The question of providing income support to substance users, particularly the substance dependent, is another example. This issue came to the fore when the number of Supplemental Security Income (SSI) recipients qualifying because of drug addiction or alcoholism (DA&A) grew sharply in the early 1990s.¹⁹ On

the one hand, addiction is clearly a condition that can restrict income, and SSI is specifically intended to provide income to the needy and disabled. On the other hand, there were reports that taxpayer money was being wasted on drug purchases (Wright 1995) and evidence this use was manifesting in adverse health outcomes (Shaner et al. 1995; Satel 1995; Satel et al. 1997). As an added complication, at least in theory, providing income support to addicts might reduce economic-compulsive crime even if it increased their use, and the bits of available evidence Reuter and MacCoun (1996) review are consistent with that hypothesis.

It is not hard to develop a list of interactions between drug use and drug policy and outcomes and policies in other domains that are of first- not second-order importance. Yet there is only the most limited capacity to quantify these interactions. To develop coherent and effective policy, more than a mere laundry list of potential issues is necessary.

Conclusions

When it comes to measuring and analyzing the extent of drug problems and the effectiveness of drug control efforts, there is an abundance of numbers, but the glass of insight is at best half full. The gaps in insight are understandable; it is difficult to study covert activities. But some of the gaps can be closed. Four encouraging trends discussed here are the expansion of traditional data systems, improved modeling of drug prices and drug markets, better integration of data systems, and improvements in data systems in other countries.

Analysts do not need to wait, however, until all possible information is available before beginning to assess the effectiveness and cost-effectiveness of drug control interventions. Indeed they cannot; those determined to wait for perfect information are determined to be irrelevant.

Pioneering work has begun to evaluate a variety of interventions' ability to control drug use. Much needs to be done to refine these estimates. Yet some of the most interesting, emerging questions pertain to complex interactions across borders, across policy domains, between substances, and over time. The cutting edge of research needs not only to refine existing estimates and improve precision but also to identify new perspectives and link previously self-contained analyses.

This has implications for data collection and analysis. There will be a premium on integrating databases, innovative one-time studies, and collecting evidence that directly informs decisions, not merely monitoring the size of the problem. In short, existing monitoring systems need to be complemented by analysis that

turns them into decision support systems in order to have the greatest and most beneficial impact on policy.

This paper draws heavily on the ideas of colleagues Mark Kleiman, Rob MacCoun, and Peter Reuter and on work supported in part by the National Science Foundation under grant no. SES-9122244. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Notes

1. The DAWN recording manual lists four criteria for determining whether an ER episode is drug related (U.S. Department of Health and Human Services, National Institute on Drug Abuse 1992): (1) The patient must be treated in the hospital's emergency department. (2) The patient's presenting problem(s) must be induced by or related to drug abuse. (3) The case must involve the nonmedical use of a legal drug or any use of an illegal drug. (4) The reason for taking the substance was for psychic effects, dependence, or a suicide attempt or gesture.

2. For example, in earlier analysis (Caulkins 1994, 47-49), I found that on the order of 1-2 percent of STRIDE's price observations were implausible. The most egregious included payments of one-tenth of 1 cent for 23 grams of 84-percent pure cocaine (San Diego, December 12, 1987) and \$6,800 for 1 milligram of cocaine (Albany, February 2, 1987).

3. This is, of course, not proof of measurement error. It may be that DAWN's definition of "drug related" excluded every instance in which a patient had cocaine in his or her system. If so, however, it would suggest that gross differences between DAWN counts and the number of cocaine users seeking emergency medical treatment are systemic and not merely confined to facilities with unusual reporting practices.

4. I owe this point to Mark Kleiman.

5. For a dissenting opinion on the importance of these numbers, see Reuter (1999).

6. For example, annual prevalence of cocaine use fell during the 1980s even as the prevalence of heavy use grew (Everingham and Rydell 1994).

7. Consider an addict who consumes 120 (pure) grams of cocaine per year. If the average social cost per gram of cocaine consumed is \$100, then a 3-month interruption in that career might be worth on the order of \$3,000 to society. That is more than the average admission to treatment costs (Rydell and Everingham 1994).

8. Consider a light user who spends \$20 per week on cocaine. That is roughly \$1,000 per year. At \$100 per (pure) gram, that is the equivalent of consuming 10 (pure) grams a year. If the light user in question would not have persisted in use for more than 3 years, then

completely eliminating that subsequent career of use would avert less consumption that would imposing a 3-month interruption on a typical heavy user who consumes at a rate of 120 pure grams per year.

9. The interruption is meaningless if one views being addicted as determined by an accumulated history of use, not current intoxication. That is, if people who are in recovery are still addicted, then an addict who has been abstinent for 3 months is in that sense still addicted.

10. The cost per unit of consumption and length of the residual career of heavy use in this example are more representative of cocaine and heroin than of marijuana.

11. Furthermore, Rydell and Everingham's analysis did not consider the possibility that such an expansion in treatment might have an adverse feedback effect on initiation (cf. Behrens et al. forthcoming). On the other hand, two-thirds of the exits from heavy use took the form of deescalation to light use.

12. Widespread availability of a substitute such as methadone is probably the only other program that can make a comparable claim concerning heroin.

13. An increase to 11X is implied by a very literal interpretation of the additive model. Those who have employed the spirit of the additive model in the past (e.g., Reuter and Kleiman 1986) have recognized that the increase would likely be slightly greater because there are additional inventory carrying costs incurred by those who purchase the more expensive drugs for resale.

14. Rosalie Pacula contributed substantially to this section.

15. Harwood, Fountain, and Livermore (1998) estimate that the social costs associated with alcohol substantially exceed those associated with all illicit drugs combined, let alone marijuana by itself.

16. I owe this observation to Bruce Johnson.

17. This section draws heavily on joint work of Gustav Feichtinger, Doris Behrens, Gernot Tragler, and others at the Vienna University of Technology in Austria.

18. Number and value of transactions may be more relevant measures of size than the number of users, given the nature of marijuana markets.

19. Reuter and MacCoun (1996) cite an increase from 20,000 in 1990 to 80,000 in 1994, or 250,000 if one includes those with DA&A as a secondary diagnosis.

References

Anglin, M. Douglas, Jonathan P. Caulkins, and Yih-Ing Hser. 1993. Prevalence estimation: Policy needs, current capacity, and future potential. *Journal of Drug Issues* 23:345-360.

Anglin, M.D., and Y. Hser. 1990a. Treatment of drug abuse. In *Drugs and crime*, edited by Michael Tonry and James Q. Wilson. Vol. 13 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

———. 1990b. Legal coercion and drug abuse treatment: Research findings and social policy implications. In *Handbook of drug control in the United States*, edited by James A. Inciardi. Westport, Connecticut: Greenwood Press.

Anglin, M. Douglas, Douglas Longshore, and Susan Turner. 1999. Treatment alternatives to street crime: An evaluation of five programs. *Criminal Justice and Behavior* 26:168–195.

Anglin, M. Douglas, and William H. McGlothlin. 1984. Outcome of narcotics addict treatment in California. In *Drug abuse treatment evaluation: Strategies, progress, and prospects*, edited by F.M. Tims and J.P. Ludford. National Institute on Drug Abuse Monograph 51. Rockville, Maryland: U.S. Department of Health and Human Services, National Institute on Drug Abuse.

Baveja, Alok, Rajan Batta, Jonathan P. Caulkins, and Mark H. Karwan. 1993. Modeling the response of illicit drug markets to local enforcement. *Socio-Economic Planning Sciences* 27 (2): 73–89.

Baveja, Alok, Jonathan P. Caulkins, Wensheng Liu, Rajan Batta, and Mark H. Karwan. 1997. When haste makes sense: Cracking down on street markets for illicit drugs. *Socio-Economic Planning Sciences* 31 (4): 293–306.

Behrens, Doris A., Jonathan P. Caulkins, Gernot Tragler, and Gustav Feichtinger. Forthcoming. Optimal control of drug epidemics: Prevent and treat—but not at the same time. *Management Science*.

———. 1997. Controlling the cocaine epidemic in the United States: Prevention from light use vs. treatment of heavy use. Working paper 214, Vienna University of Technology, Austria.

Behrens, D.A., J.P. Caulkins, G. Tragler, J. Haunschmied, and G. Feichtinger. 1999. A dynamic model of drug initiation: Implications for treatment and drug control. *Mathematical Biosciences* 159:1–20.

Belenko, Stephen. Forthcoming. Ten years of drug courts: A review of current research. *National Drug Court Institute Review*.

Blumstein, A. 1993. Racial disproportionality of U.S. prison populations revisited. *University of Colorado Law Review* 64:743–760.

Botvin, G.J., E. Baker, L. Dusenbury, E.M. Botvin, and T. Diaz. 1995. Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. *Journal of the American Medical Association* 273:1106–1112.

Boyum, D. 1992. Reflections on economic theory and drug enforcement. Ph.D. diss., Harvard University, Cambridge.

Boyum, D., and P. Reuter. Forthcoming. *Reflections on drug policy and social policy*. Cambridge: Harvard University Press.

Brookoff, Daniel, Elizabeth A. Campbell, and Leslie M. Shaw. 1993. The underreporting of cocaine-related trauma: Drug Abuse Warning Network reports vs. hospital toxicology tests. *American Journal of Public Health* 83 (3): 369–371.

Carley, K. 1990. Group stability: A socio-cognitive approach. *Advances in Group Processes* 7:1–44.

Cartwright, W.S., and J.M. Kaple. 1991. *Economic costs, cost-effectiveness, financing, and community-based drug treatment*. Rockville, Maryland: U.S. Department of Health and Human Services, National Institute on Drug Abuse.

Caulkins, Jonathan P. 1998. The cost-effectiveness of civil remedies: The case of drug control interventions. In *Civil remedies and crime prevention*, edited by Lorraine Green Mazerolle and Janice Roehl. Vol. 9 of *Crime prevention studies*. Monsey, New York: Criminal Justice Press.

———. 1995. Domestic geographic variation in illicit drug prices. *Journal of Urban Economics* 37:38–56.

———. 1994. *Developing price series for cocaine*. Document no. MR-317-DPRC. Santa Monica, California: RAND.

———. 1993. Local drug markets' response to focused police enforcement. *Operations Research* 41 (5): 848–863.

———. 1990. The distribution and consumption of illicit drugs: Some mathematical models and their policy implications. Ph.D. diss., Massachusetts Institute of Technology, Cambridge.

Caulkins, J.P., G. Crawford, and P. Reuter. 1993. Simulation of adaptive response: A model of drug interdiction. *Mathematical and Computer Modeling* 17:37–52.

Caulkins, Jonathan P., Patricia Ebener, and Daniel McCaffrey. 1995. Describing DAWN's dominion. *Contemporary Drug Problems* 22:547–567.

Caulkins, Jonathan P., Bruce Johnson, Angela Taylor, and Lowell Taylor. 1999. What drug dealers tell us about their costs of doing business. *Journal of Drug Issues* 29 (2): 323–340.

Caulkins, J.P., and E.H. Kaplan. 1991. AIDS' impact on the number of intravenous drug users. *Interfaces* 21:50–63.

Caulkins, Jonathan P., and Rema Padman. 1993. Quantity discounts and quality premia for illicit drugs. *Journal of the American Statistical Association* 88:748–757.

Caulkins, J.P., and P. Reuter. 1998. What price data tell us about drug markets. *Journal of Drug Issues* 28:593–612.

———. 1997. Setting goals for drug policy: Harm reduction or use reduction. *Addiction* 92:1143–1150.

Caulkins, J.P., C.P. Rydell, S.S. Everingham, J. Chiesa, and S. Bushway. 1999. *An ounce of prevention, a pound of uncertainty: The cost-effectiveness of school-based drug prevention programs*. Santa Monica, California: RAND.

Caulkins, J.P., C.P. Rydell, W.L. Schwabe, and J. Chiesa. 1997. *Mandatory minimum drug sentences: Throwing away the key or the taxpayers' money?* Santa Monica, California: RAND.

Chaloupka, F.J., and A. Laixuthai. 1997. Do youths substitute alcohol and marijuana? Some econometric evidence. *Eastern Economic Journal* 23 (3): 253–276.

Chaloupka, F.J., R.L. Pacula, M. Farrelly, L. Johnston, P. O'Malley, and J. Bray. 1999. Do higher cigarette prices encourage youth to use marijuana? Working paper 6938, National Bureau of Economic Research, Cambridge, Massachusetts.

Chambers, C.D., W.J. Taylor, and A.D. Moffett. 1972. The incidence of cocaine abuse among methadone maintenance patients. *International Journal on Addiction* 7:427–441.

Childress, Michael. 1994a. *A systems description of the heroin trade*. Santa Monica, California: RAND.

———. 1994b. *A systems description of the marijuana trade*. Santa Monica, California: RAND.

Crane, B.D., A.R. Rivolo, and G.C. Comfort. 1997. *An empirical examination of counterdrug program effectiveness*. Alexandria, Virginia: Institute for Defense Analyses.

Crawford, G.B., and P. Reuter, with K. Isaacson and P. Murphy. 1988. *Simulation of adaptive response: A model of drug interdiction*. Santa Monica, California: RAND.

Dawid, H., and G. Feichtinger. 1996. Optimal allocation of drug control efforts: A differential game analysis. *Journal of Optimization Theory and Applications* 91:279–297.

DeSimone, J. 1998. Is marijuana a gateway drug? *Eastern Economic Journal* 24 (2): 149–164.

DiNardo, J., and T. Lemieux. 1992. Alcohol, marijuana, and American youth: The unintended effects of government regulation. Working paper 4212, National Bureau of Economic Research, Cambridge, Massachusetts.

Dombey-Moore, Bonnie, Susan Resetar, and Michael Childress. 1994. *A systems description of the cocaine trade*. Santa Monica, California: RAND.

Drug Strategies. 1998. *Passing judgement: The U.S. drug certification process*. Washington, D.C.

Ebener, Patricia A., Eva Feldman, and Nora Fitzgerald. 1993. *Federal databases for use in drug policy research: A catalogue for data users*. Santa Monica, California: RAND.

Ebener, Patricia A., Hilary Saner, and M. Douglas Anglin. 1995. *Building a data and analysis infrastructure to support substance abuse policy decisionmaking: A strategic plan*. Santa Monica, California: RAND.

Ebener, Patricia A., and Beverly Weidmer. 1994. Compendium of existing surveys and data sets on alcohol and other drug use and consequences in California. Report submitted to the State of California, Department of Alcohol and Drug Programs. Santa Monica, California: RAND.

Ellickson, Phyllis L., and Robert M. Bell. 1990. Drug prevention in junior high: A multi-site longitudinal test. *Science* 247 (March):16-248.

Everingham, Susan S., and C. Peter Rydell. 1994. *Modeling the demand for cocaine*. MR-332-ONDCP/A/DPRC. Santa Monica, California: RAND.

Executive Office of the President, Office of National Drug Control Policy. 1999. *The National Drug Control Strategy*. Washington, D.C.

———. 1998a. *Performance measures of effectiveness: A system for assessing the performance of the national drug control strategy*. Washington, D.C.

———. 1998b. *Report of the Drug Control Research, Data, and Evaluation Committee*. Washington, D.C.

———. 1997. *The National Drug Control Strategy*. Washington, D.C.

———. 1992. *The National Drug Control Strategy*. Washington, D.C.

———. 1990. *Leading drug indicators*. Washington, D.C.

———. 1989. *The National Drug Control Strategy*. Washington, D.C.

Falco, Mathea. 1995. Passing grades. *Foreign Affairs* 74:15-20.

Farrell, G., K. Mansur, and M. Tullis. 1996. Cocaine and heroin in Europe 1983-1993: A cross-national comparison of trafficking and prices. *British Journal of Criminology* 36:255-281.

Farrelly, M.C., J.W. Bray, G.A. Zarkin, B.W. Wendling, and R.L. Pacula. 1999. The effects of prices and policies on the demand for marijuana: Evidence from the National Household Surveys on Drug Abuse. Working paper 6940, National Bureau of Economic Research, Cambridge, Massachusetts.

Frank, Richard S. 1987. Drugs of abuse: Data collection systems of DEA and recent trends. *Journal of Analytical Toxicology* 11 (November/December): 237–241.

Friedman, Milton, and Thomas Szasz. 1992. *On liberty and drugs: Essays on the free market and Prohibition*. Washington, D.C.: Drug Policy Foundation Press.

Garcia Arganaras, F. 1997. Harm reduction at the supply side of the drug war: The case of Bolivia. In *Harm reduction: A new direction for drug policies and programs*, edited by Patricia G. Erickson, Diane M. Riley, Yuet W. Cheung, and Patrick A. O'Hare. Toronto: University of Toronto Press.

Gardiner, L.K., and R.C. Shreckengost. 1987. A system dynamics model for estimating heroin imports into the United States. *System Dynamics Review* 3:8–27.

Gerstein, Dean R., and Henrick J. Harwood, editors. 1990. *Treating drug problems: A study of the evolution, effectiveness, and financing of public and private drug treatment systems*. Vol. 1. Washington, D.C.: National Academy Press.

Gerstein, D.R., R.A. Johnson, H.J. Harwood, D. Fountain, N. Suter, and K. Malloy. 1994. *Evaluating recovery services: The California drug and alcohol treatment assessment*. Chicago: National Opinion Research Center; and Fairfax, Virginia: Lewin-VHI.

Gold, M.R., J.E. Siegel, L.B. Russell, and M.C. Weinstein. 1996. *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Goldstein, P.J. 1985. The drugs/violence nexus: A tripartite conceptual framework. *Journal of Drug Issues* 15 (4): 493–506.

Golub, Andrew Lang, and Bruce D. Johnson. 1997. *Crack's decline: Some surprises across U.S. cities*. Research in Brief, NCJ 165707. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Gorman, D.M. 1998. The irrelevance of evidence in the development of school-based drug prevention policy, 1986–1996. *Evaluation Review* 22:118–146.

———. 1995. On the difference between statistical and practical significance in school-based drug abuse prevention. *Drugs: Education, Prevention, and Policy* 2 (3): 275–283.

Graggani, A., S. Rinaldi, and G. Feichtinger. 1997. Dynamics of drug consumption: A theoretical model. *Socio-Economic Planning Sciences* 31 (2): 127–137.

Grossman, M., and F.J. Chaloupka. 1998. The demand for cocaine by young adults: A rational addiction approach. *Journal of Health Economics* 17:427–474.

Haaga, John G., and Peter Reuter. 1991. *Improving data for Federal drug policy decisions*. Santa Monica, California: RAND.

Harwood, H., D. Fountain, and G. Livermore. 1998. *The economic costs of alcohol and drug abuse in the United States, 1992*. Washington, D.C.: U.S. Department of Health and Human Services.

Higgins S.T., A.J. Budney, W.K. Bickel, J.R. Hughes, F. Foerg, and G. Badger. 1993. Achieving cocaine abstinence with a behavioral approach. *American Journal of Psychiatry* 150:763–769.

Higgins S.T., A.J. Budney, W.K. Bickel, J.R. Hughes, F. Foerg, R. Donham, and G. Badger. 1994. Incentives improve outcome in outpatient behavioral treatment of cocaine dependence. *Archives of General Psychiatry* 51:568–576.

Hiller, Matthew L., Kevin Knight, and Dwayne D. Simpson. 1999. Prison-based substance abuse treatment, residential aftercare and recidivism. *Addiction* 94 (6): 833–842.

Homer, Jack B. 1993. A system dynamics model for cocaine prevalence estimation and trend projection. *Journal of Drug Issues* 23:251–279.

———. 1993. Projecting the impact of law enforcement on cocaine prevalence: A system dynamics approach. *Journal of Drug Issues* 23:281–295.

Hser, Yih-Ing, M. Douglas Anglin, Thomas D. Wickens, Mary-Lynn Brecht, and Jack Homer. 1992. Techniques for the estimation of illicit drug-use prevalence: An overview of relevant issues. NCJ 133786. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Hubbard, R.L., M.E. Marsden, J.V. Rachal, H.J. Harwood, E.R. Cavanaugh, and H.M. Ginzburg. 1989. *Drug abuse treatment: A national study of effectiveness*. Chapel Hill: University of North Carolina Press.

Inciardi, James A., and D.C. McBride. 1991. *Treatment Alternatives to Street Crime (TASC): History, experiences, and issues*. Rockville, Maryland: U.S. Department of Health and Human Services, National Institute on Drug Abuse.

Institute of Medicine. 1996. *Pathways of addiction: Opportunities in drug abuse research*. Washington, D.C.: National Academy Press.

Johnson, B.D., P. Goldstein, E. Preble, J. Schmeidler, D. Lipton, B. Spunt, and T. Miller. 1985. *Taking care of business: The economics of crime by heroin abusers*. Lexington, Massachusetts: D.C. Heath.

Johnson, B.D., A. Hamid, and H. Sanabria. 1991. Emerging models of crack distribution. In *Drugs and crime: A reader*, edited by T. Mieczkowski, Boston: Allyn and Bacon.

Johnson, R.A., D.R. Gerstein, R. Ghadialy, W. Choy, and J. Gfroerer. 1996. *Trends in the incidence of drug use in the United States, 1919–1992*. Washington, D.C.: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Kamlet, Mark. 1992. *The comparative benefits modeling project: A framework for cost-utility analysis of governmental health care programs*. Washington, D.C.: U.S. Department of Health and Human Services.

Karoly, Lynn A., Peter W. Greenwood, Susan S. Everingham, Jill Hoube, M. Rebecca Kilburn, C. Peter Rydell, Matthew Sanders, and James Chiesa. 1998. *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions*. MR-898-TCWF. Santa Monica, California: RAND.

Kelling, G.L., and C.M. Coles. 1996. *Fixing broken windows: Restoring order and reducing crime in our communities*. New York: Free Press.

Kennedy, Michael, Peter Reuter, and Kevin Jack Riley. 1994. *A simple economic model of cocaine production*. Santa Monica, California: RAND.

Kleiman, Mark A.R. 1997a. Coerced abstinence: A neo-paternalistic drug policy initiative. In *The new paternalism*, edited by L.A. Mead. Washington, D.C.: Brookings Institution Press.

———. 1997b. The problem of replacement and the logic of drug law enforcement. *Drug Policy Analysis Bulletin* 3.

———. 1992. *Against excess: Drug policy for results*. New York: Basic Books.

———. 1988. Crackdowns: The effects of intensive enforcement on retail heroin dealing. In *Street-level drug enforcement: Examining the issues*, edited by Marcia R. Chaiken. NCJ 115403. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Kleiman, M.A.R., and K.D. Smith. 1990. State and local drug enforcement in search of a strategy. In *Drugs and crime*, edited by Michael Tonry and James Q. Wilson. Vol. 13 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

Kort, P.M., G. Feichtinger, R.F. Hartl, and J.L. Haunschmied. 1998. Optimal enforcement policies (crackdowns) on an illicit drug market. *Optimal Control Applications & Methods* 19:169–184.

Larkey, P.D., and J.P. Caulkins. 1991. The persistence of performance evaluation systems. *Proceedings of the 1991 National Public Management Research Conference*. Syracuse, New York: Syracuse University, Maxwell School of Citizenship and Public Affairs.

Lemmens, Trudo. 1997. Harm reduction, doping, and the clashing values of athletic sport. In *Harm reduction: A new direction for drug policies and programs*, edited by Patricia G. Erickson, Diane M. Riley, Yuet W. Cheung, and Patrick A. O'Hare. Toronto: University of Toronto Press.

Levin, G., E.B. Roberts, and G.B. Hirsch. 1975. *The persistent poppy: A computer-aided search for heroin policy*. Cambridge, Massachusetts: Ballinger.

Levitt, Steven, and Sudhir Alladi Venkatesh. 1998. *An economic analysis of a drug-selling gang's finances*. Cambridge, Massachusetts: National Bureau of Economic Research.

Lurie, P., A.L. Reingold, B. Bowser, D. Chen, J. Foley, J. Guldish, J.G. Kahn, S. Lane, and J. Sorensen. 1993. *The public health impact of needle exchange programs in the United States and abroad: Summary, conclusions, and recommendations*. San Francisco: University of California, Berkeley, School of Public Health, and University of California at San Francisco, Institute of Health Policy Studies.

MacCoun, Robert J. 1996. *The psychology of harm reduction: Comparing alternative strategies for modifying high-risk behavior*. Wellness Lecture Series, vol. 6. Woodland Hills: California Wellness Foundation and the University of California.

MacCoun, Robert J., and Peter Reuter. Forthcoming. *Beyond the drug war: Learning from other times, places, and vices*. New York: Cambridge University Press.

———. 1997. Interpreting Dutch cannabis policy: Reasoning by analogy in the legalization debate. *Science* 278:47–52.

MacCoun, Robert J., Peter Reuter, and Thomas Schelling. 1996. Assessing alternative drug control regimes. *Journal of Policy Analysis and Management* 15:330–352.

Maher, Lisa. 1997. *Sexed work: Gender, race, and resistance in a Brooklyn drug market*. New York: Cambridge University Press.

McCaffrey, Barry R. 1999. A report to the Nation: The youth anti-drug media campaign. Testimony before the House Appropriations Subcommittee on Treasury, Postal Service, and General Government, 106th Cong. 25 March.

McCarthy, J., and O.T. Borders. 1985. Limit setting on drug abuse in methadone maintenance patients. *American Journal of Psychiatry* 142:1419–1423.

McGlothlin, W.H., M.D. Anglin, and B.D. Wilson. 1977. *An evaluation of the California Civil Addict Program*. Rockville, Maryland: U.S. Department of Health and Human Services, National Institute on Drug Abuse.

McLellan, A.T., G.E. Woody, D. Metzger, J. McKay, J. Durell, A. Alterman, and C.P. O'Brien. 1996. Evaluating the effectiveness of addiction treatments: Reasonable expectations, appropriate comparisons. *Milbank Quarterly* 74:51–85.

Model, Karyn E. 1993. The effect of marijuana decriminalization on hospital emergency room drug episodes: 1975–1978. *Journal of the American Statistical Association* 88:737–747.

Moore, Mark H. 1990. Supply reduction and drug law enforcement. In *Drugs and crime*, edited by M. Tonry and J.Q. Wilson. Vol. 13 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

———. 1973. Policies to achieve discrimination on the effective price of heroin. *American Economic Review* 63 (2): 270–277.

Moskowitz, Joel M. 1993. Why reports of outcome evaluations are often biased or uninterpretable. *Evaluation and Program Planning* 16:1–9.

Murphy, Patrick. 1994. *Keeping score: The frailties of the Federal drug budget*. Santa Monica, California: RAND.

Musto, D.F. 1999. *The American disease*. New Haven: Yale University Press.

Naik, A.V., A. Baveja, R. Batta, and J.P. Caulkins. 1996. Scheduling crackdowns on illicit drug markets. *European Journal of Operational Research* 88:231–250.

National Research Council. 1995. *Preventing HIV transmission: The role of sterile needles and bleach*. Washington D.C.: National Academy Press.

———. 1994. *Under the influence? Drugs and the American workplace*. Washington, D.C.: National Academy Press.

———. 1991. *Improving information for social policy decision: The uses of microsimulation modeling, volume I: Review and recommendations*. Panel to Evaluate Microsimulation Models for Social Welfare Programs. Washington, D.C.: National Academy Press.

Newmeyer, J. 1999. Eighteen years of AIDS and IDUs. *MidCity Numbers: A Monthly Bulletin of AIDS-Related Statistics* 12:2–3.

Pacula, R.L. 1998a. Adolescent alcohol and marijuana consumption: Is there really a gateway effect? Working paper 6348, National Bureau of Economic Research, Cambridge, Massachusetts.

———. 1998b. Does increasing the beer tax reduce marijuana consumption? *Journal of Health Economics* 17:557–586.

Pentz, Mary Ann. 1998. Costs, benefits, and cost-effectiveness of comprehensive drug abuse prevention. In *Cost-benefit/cost-effectiveness research of drug abuse prevention: Implications for programming and policy*. Research Monograph 176. Washington, D.C.: U.S. Department of Health and Human Services, National Institute on Drug Abuse.

Prendergast, M.L., M.D. Anglin, T.H. Maugh II, and Y. Hser. 1995. The effectiveness of treatment for drug abuse. Los Angeles: UCLA Drug Abuse Research Center. Paper prepared for the U.S. Department of Health and Human Services, National Institute on Drug Abuse, Washington, D.C.

Prendergast, Michael, Deborah Podus, and Kai McCormack. 1998. Bibliography of literature reviews on drug abuse treatment effectiveness. *Journal of Substance Abuse Treatment* 15:267–270.

Reuter, Peter. 1999. Are calculations of the economic costs of drug abuse either possible or useful? *Addiction* 94:635–638.

———. 1998. Review of United Nations International Drug Control Program's World Drug Report. *Journal of Policy Analysis and Management* 17:730–734.

———. 1997. Why can't we make prohibition work better? Some consequences of ignoring the unattractive. *Proceedings of the American Philosophical Society* 141:262–275.

———. 1993. Prevalence estimation and policy formulation. *Journal of Drug Issues* 23:167–184.

———. 1988. Quantity illusions and paradoxes of drug interdiction: Federal intervention into vice policy. *Law and Contemporary Problems* 51:233–252.

———. 1985. Eternal hope: America's international narcotics efforts. *Public Interest* 79:13–94.

———. 1984. The (continued) vitality of mythical numbers. *Public Interest* 78:135–147.

———. 1983. *Disorganized crime: The economics of the visible hand*. Cambridge: MIT Press.

Reuter, P., and J.P. Caulkins. 1995. Redefining the goals of drug policy: Report of a working group. *American Journal of Public Health* 85:1059–1063.

Reuter, P., P. Ebener, and D. McCaffrey. 1994. Patterns of drug use. In *When drug addicts have children: Reorienting child welfare's response*, edited by Douglas Besharov. Washington, D.C.: American Enterprise Institute Press.

- Reuter, P., and M.A.R. Kleiman. 1986. Risks and prices: An economic analysis of drug enforcement. In *Crime and justice: An annual review of research*, edited by M. Tonry and N. Morris. Vol. 7. Chicago: University of Chicago Press.
- Reuter, P., and R. MacCoun. 1996. Harm reduction and social policy: Should addicts be paid? *Drug and Alcohol Review* 15:225–230.
- Reuter, P., R. MacCoun, and P. Murphy. 1990. *Money from crime: A study of the economics of drug dealing in Washington, D.C.* Santa Monica, California: RAND.
- Rhodes, W. 1998. ADAM program development focus: Sampling issues. In *ADAM 1997 annual report on adult and juvenile arrestees*. Research Report, NCJ 171672. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.
- Rhodes, W., P. Scheiman, T. Pittayathikhun, L. Collins, and V. Tsarfaty. 1995. *What America's users spend on illegal drugs, 1988–1993*. Washington, D.C.: Executive Office of the President, Office of National Drug Control Policy.
- Rice, D.P., S. Kelman, L.S. Miller, and S. Dunmeyer. 1990. *The economic costs of alcohol and drug abuse and mental illness: 1985*. San Francisco: University of California, Institute for Health and Aging.
- Riley, Kevin Jack. 1997. *Crack, powder cocaine, and heroin: Drug purchase and use patterns in six U.S. cities*. Research Report, NCJ 167265. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.
- Roberts, C.D. 1996. Data quality of the Drug Abuse Warning Network. *American Journal of Drug and Alcohol Abuse* 22 (3): 389–401.
- Rocheleau, Ann Marie, and David Boyum. 1994. *Measuring heroin availability in three cities*. Washington, D.C.: Executive Office of the President, Office of National Drug Control Policy.
- Rydell, C. Peter. 1997. Dynamic analysis of drug control effectiveness. Mimeo.
- . 1997. Variation over time in the marginal cost-effectiveness of drug control interventions. Mimeo.
- Rydell, C. Peter, Jonathan P. Caulkins, and Susan Everingham. 1996. Enforcement or treatment: Modeling the relative efficacy of alternatives for controlling cocaine. *Operations Research* 44 (5): 687–695.
- Rydell, C.P., and S.S. Everingham. 1994. *Controlling cocaine: Supply versus demand programs*. Santa Monica, California: RAND.

Saffer, H., and F.J. Chaloupka. 1999. Demographic differentials in the demand for alcohol and illicit drugs. In *The economic analysis of substance use and abuse: An integration of econometric and behavioral economic research*, edited by Chaloupka, F.J., M. Grossman, W.K. Bickel, and H. Saffer. Chicago: University of Chicago Press.

———.1995. The demand for illicit drugs. Working paper number 5238, National Bureau of Economic Research, Cambridge, Massachusetts.

Satel, Sally. 1995. When disability benefits make patients sicker. *New England Journal of Medicine* 333:794–796.

Satel, S., P. Reuter, D. Hartley, R. Rosenheck, and J. Mintz. 1997. Influence of retroactive disability payments on recipients' compliance with substance abuse treatment. *Psychiatric Services* 48:796–799.

Schlenger, W.E. 1973. A systems approach to drug user services. *Behavioral Science* 18:137–147.

Shaner, A., T.A. Eckman, L.J. Roberts, J.N. Wilkins, D.E. Tucker, J.W. Tsuang, and J. Mintz. 1995. Disability income, cocaine use, and repeated hospitalization among schizophrenic cocaine abusers—A government-sponsored revolving door? *New England Journal of Medicine* 333:777–783.

Sherman, L.W., D.C. Gottfredson, D.L. MacKenzie, J. Eck, P. Reuter, and S.D. Bushway. 1997. *Preventing crime: What works, what doesn't, what's promising*. Research Report, NCJ 165366. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Silverman K., S.T. Higgins, R.K. Brooner, I.D. Montoya, E.J. Cone, C.R. Schuster, and K.L. Preston. 1996. Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. *Archives of General Psychiatry* 53:409–415.

Singer, Max. 1971. The vitality of mythical numbers. *Public Interest* 23.

Stares, Paul B. 1996. *Global habit: The drug problem in a borderless world*. Washington, D.C.: Brookings Institution Press.

Steinitz, Mark S. 1985. Insurgents, terrorists and the drug trade. *Washington Quarterly* (Fall): 141–153.

Taylor, B., and T. Bennett. 1999. *Comparing drug use rates of detained arrestees in the United States and England*. Research Report, NCJ 175052. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Thies, C.F., and C.A. Register. 1993. Decriminalization of marijuana and the demand for alcohol, marijuana and cocaine. *Social Science Journal* 30:385–399.

- Tita, G. 1999. An ecological study of urban street gang formation and their effect on crime. Ph.D. diss., Carnegie Mellon University, Pittsburgh.
- Tobler, N.S. 1997. Meta-analysis of adolescent drug prevention programs: Results of the 1993 Meta-analysis. In *Meta-analysis of drug abuse prevention programs*, edited by William J. Bukoski. NIDA Research Monograph 170. Washington, D.C.: U.S. Department of Health and Human Services, National Institute on Drug Abuse.
- Tragler, Gernot, Jonathan P. Caulkins, and Gustav Feichtinger. Forthcoming. Optimal dynamic allocation of treatment and enforcement in illicit drug control. *Operations Research*.
- . 1997. The impact of enforcement and treatment on illicit drug consumption. Working paper 212, Department of Operations Research and Systems Theory, Vienna University of Technology, Austria.
- Ungerleider, J. Thomas, G.D. Lundberg, I. Sunshine, and C.B. Walberg. 1980. The Drug Abuse Warning Network (DAWN) program. *Archives of General Psychiatry* 37:106–109.
- United Nations International Drug Control Programme. 1998. *Supply of and trafficking in narcotic drugs and psychotropic substances, 1996*. Vienna, Austria: United Nations.
- . 1997. *World drug report*. New York: Oxford University Press.
- U.S. Department of Health and Human Services. 1998. *Mid-year 1997 preliminary emergency department data from the Drug Abuse Warning Network*. Washington, D.C.
- U.S. Department of Health and Human Services, National Institute on Drug Abuse. 1992. *Annual emergency room data, 1991: Data from the Drug Abuse Warning Network*. Series 1, no. 11–A. Washington, D.C.
- U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration. 1996. *National Household Survey on Drug Abuse: Population estimates 1995*. Washington, D.C.
- Valliant, G. 1997. Poverty and paternalism: A psychiatric viewpoint. *The new paternalism*, edited by Lawrence M. Mead. Washington, D.C.: Brookings Institution Press.
- Ward, J., R. Mattick, and W. Hall. 1994. The effectiveness of methadone maintenance treatment: An overview. *Drug and Alcohol Review* 13:327–336.
- Weatherburn, Don, and Bronwyn Lind. 1997. On the epidemiology of offender populations. *Australian Journal of Psychology* 49 (3): 169–175.
- Weisburd, D. and L. Green. 1995. Defining the drug market: The case of the Jersey City DMA system. In *Drugs and crime: Evaluating public policy initiatives*, edited by D.L. MacKenzie and C.D. Uchida. Newbury Park, California: Sage Publications.

Weisheit, Ralph A. 1990. Cash crop: A study of illicit marijuana growers. Report submitted to the U.S. Department of Justice, National Institute of Justice, Washington, D.C.

Wilson, James Q. 1990. Against the legalization of drugs. *Commentary* 89: 21–28.

Winsten, Jay. 1999. Personal communication, April.

Wright, C. 1995. SSI: The black hole of the welfare state. *Policy Analysis* 224. Washington, D.C.: Cato Institute.

Yuan, Y., and J.P. Caulkins. 1998. The effect of variation in high-level domestic drug enforcement on variation in drug prices. *Socio-Economic Planning Sciences* 32.

Zeggelink, E. 1995. Evolving friendship networks: An individual-oriented approach implementing similarity. *Social Networks* 17:83–110.

Fear of Crime in the United States: Avenues for Research and Policy

by Mark Warr

Fear of crime affects far more people in the United States than crime itself, and there are sound reasons for treating crime and fear of crime as distinct social problems. After assessing the state of knowledge on fear, this chapter considers whether public fear of crime can and ought to be controlled, and the moral and practical implications of doing so. The discussion draws on the literatures of risk perception and risk communication, as well as research on the etiology of fear and public beliefs about crime. A final objective of the chapter is to identify the most pressing unanswered questions about fear confronting investigators today.

451

Mark Warr is a Professor in the Department of Sociology at the University of Texas at Austin.

Criminal events capture the attention of the general public in a way that few other events can (cf. Skogan and Maxfield 1981). One reason is that crimes receive extraordinary emphasis in the mass media, from news coverage to feature films to television dramas to crime fiction (Graber 1980; Skogan and Maxfield 1981; Warr 1994). But even without this sort of amplification, crimes are intrinsically interesting events. As condensed and emblematic accounts of human conflict, they raise profound questions about the nature and sources of human motivation, the misfortune of fellow humans, the ability of the state to maintain social order, and, ultimately, the presence or absence of justice in human affairs.

There is another, perhaps more crucial, reason that crimes generate such acute public interest. Criminal events, at their most elemental level, are *frightening* events. They are reminders to all that the world is not a safe place, that danger can strike at any time or location, and that life, in the end, is tenuous and precious.

Judging from the attention it elicited from criminologists, public fear of crime was regarded as a trivial consequence of crime through most of the history of criminology. None of the great names of 19th-century criminology gave the matter much attention, and the situation changed relatively little during the first half of the 20th century. Many investigators, it seems, adopted the commonsensical but questionable notion that fear is directly proportional to objective risk, and assumed that strategies to control crime are ipso facto strategies to control fear. Although the serious personal consequences of criminal victimization were apparent to criminologists, no one allowed that fear *alone* could be debilitating.

Some three decades ago, however, the President's Commission on Law Enforcement and Administration of Justice (1967, 3) offered this brief but trenchant observation: "The most damaging of the effects of violent crime is fear, and that fear must not be belittled." That statement prefigured a fundamental shift in the way that criminologists think about the consequences of crime, one that was to heavily influence the course of criminological research in years to come. To fully understand the social consequences of crime, criminologists came to realize, investigators cannot focus merely on those who become direct victims of crime. Important as these individuals surely are, researchers must also concentrate on those who suffer forms of *indirect* victimization (Conklin 1971, 1975), the most egregious of which is fear of crime.

The wisdom and awful implications of this insight were quickly borne out by survey research demonstrating that fear of crime in the United States is far more prevalent than actual victimization (often by orders of magnitude), and that Americans react to this fear through a variety of precautionary behaviors so pervasive and normative that they form a significant and defining element of American culture (Warr 1994).

Since the days of the President's Commission, hundreds of studies of fear of crime have been conducted, and the topic regularly appears in the journals of the field. For reasons that remain elusive, however, the study of fear seems to have stalled at a rudimentary phase of development, a situation that is in danger of turning into outright stagnation. Investigators continue to revisit the same well-worn issues, and, even after three decades, the meaning of the term "fear" remains a matter of controversy.

This chapter has three principal purposes. One is to identify the most pressing unanswered questions about fear of crime, giving proper recognition to existing lines of research and traditions in the field. Another goal is to consider the merits and prospects for controlling public fear of crime, recognizing that fear has beneficial as well as deleterious consequences, that individuals can be too unafraid as well as too afraid, and that fear depends in part on subjective factors for which there is no objective standard or valuation. The logistic and ethical complexities of controlling fear have thus far deterred researchers from any protracted discussion of the matter, but it is far too important an issue to ignore or defer. A final purpose of the chapter is to offer a brief recounting of the history of research on fear of crime to those unfamiliar with the field.

The Nature of Fear

Despite decades of research and debate, investigators have yet to settle on a definition of fear of crime. Over the years, the phrase has been equated with a variety of emotional states, attitudes, or perceptions (including mistrust of others, anxiety, perceived risk, fear of strangers, or concern about deteriorating neighborhoods or declining national morality). Even those whose work is otherwise laudable seem to have trouble defining fear of crime. Ferraro and LaGrange, for example, initially defined fear as "negative emotional reactions generated by crime or symbols associated with crime" (1987, 73). Under that definition, however, it would be difficult to distinguish fear from sadness, anger, despair, or resignation.

Much of the confusion over the meaning of fear seems to arise from a failure to recognize elementary distinctions between perception, cognition, and emotion. Notwithstanding the claims of some, fear is not a perception of the environment (an awareness or experience of sensory stimuli), but a reaction to the perceived environment. Although fear may result from the cognitive processing or evaluation of perceptual information (e.g., a judgment that an approaching male is armed, or that a sound signals danger), fear is not itself a belief, attitude, or evaluation. On the contrary, fear is an emotion, a feeling of alarm or dread caused by an awareness or expectation of danger (see Sluckin 1979). This

affective state is ordinarily (though not invariably) associated with certain physiological changes, including increased heart rate, rapid breathing, sweating, decreased salivation, and increased galvanic skin response (Thomson 1979; Mayes 1979).

Were it not such a serious matter, the disarray among criminologists over the meaning of fear might be amusing. Whatever confusion criminologists may suffer, however, the concept of fear is routinely and profitably used in psychology and the life sciences, with considerably less dispute as to its meaning. In everyday life, the emotion of fear is a common experience for most human beings, for whom it is no more mysterious than anger, joy, or despair. For their part, criminologists continue to exhibit a tendency to isolate or compartmentalize “fear of crime,” to assume that it differs in some fundamental way from other ordinary fears, such as fear of traffic accidents, fear of falling, or fear of disease. But there is no evidence that fear of crime is qualitatively different from other forms of fear. What differentiates one from another is merely the object or stimulus of fear.

One common source of confusion when it comes to defining fear of crime occurs when investigators equate fear of crime with the perceived risk of victimization (i.e., the subjective probability of victimization). However, there are compelling reasons (among them predictive accuracy, convergent validity, and logical necessity) to believe that perceived risk is a proximate cause of fear—not fear itself (see Warr and Stafford 1983; Warr 1984, 1985, 1991, 1994; Ferraro 1995). And there is corroborating evidence that measures of fear and measures of perceived risk do not measure the same phenomenon and do not behave similarly with respect to other variables (Rountree and Land 1996; Ferraro 1995). In short, fear is not perceived risk; by all indications, it is its consequence.

Fear of crime may be aroused by an immediate danger, as when an individual is confronted by an armed attacker or is verbally threatened with harm. This type of intense, immediate experience appears to be what some have in mind when they speak of fear of crime. As sentient and symbolic beings, however, humans have the ability to anticipate or contemplate events that lie in the future or are not immediately apparent. Hence people may experience fear merely in anticipation of possible threats or in reaction to environmental cues (e.g., darkness, graffiti) that imply danger. Psychologists commonly use the terms fear and anxiety to differentiate reactions to immediate threats (fear) from reactions to future or past events (anxiety). This terminological clarity has not been adopted in research on fear of crime, but it appears that most measures of fear are designed to capture anxiety rather than fear of victimization. This practice evidently rests on the assumption that anxiety about future victimization is much more common among the general public than fear associated with actual

encounters with crime, a reasonable assumption (see Warr 1994). Hereafter, I will draw the distinction between fear and anxiety where it is heuristically useful or otherwise appropriate.

By its very nature, the term fear seems to imply a deleterious emotional or psychological condition. Unlike love, pleasure, or happiness, fear is not a state that people (thrill-seekers aside) ordinarily pursue. To assume that fear is therefore dysfunctional for an organism, however, is to commit a serious error. On the contrary, the presence of fear in virtually all animals is no accident. Without fear, prey animals would walk amid predators, and humans would stroll across busy freeways, knowingly ingest toxic substances, or leave their infants unprotected. From an evolutionary point of view, animals that lacked fear would be unlikely to live long enough to reproduce, suggesting that fear is a potent natural selection factor (Russell 1979; Mayes 1979).

Fear, then, is not intrinsically bad. It is when fear is out of proportion to objective risk that it becomes dysfunctional. We will return to that issue later as we consider the control of fear.

Fear of crime can be characterized according to a number of properties, including intensity (the English language recognizes many degrees of fear: terror, worry, alarm, apprehension, dread), prevalence (the proportion of a population that experiences fear during some reference period), and duration, both among individuals and within social units (e.g., communities, cities, nations). Because actual criminal events or exposure to immediate signs of danger are usually brief, episodes of fear (strictly defined) are usually brief as well. Anxiety, on the other hand, can become a chronic or obsessive condition (Sluckin 1979).

The prevalence and power of altruistic fear are illustrated by the enormous public reaction that often attends crimes committed against children.

When individuals face an ostensibly dangerous environment, they may naturally experience fear for their own personal safety. At the same time, they may also fear for other individuals (e.g., children, spouses, friends) whose safety they value. It is essential, therefore, to distinguish personal fear (fear for oneself) from altruistic fear (fear for others). The prevalence and power of altruistic fear are illustrated by the enormous public reaction that often attends crimes committed against children (e.g., Polly Klaas, Columbine High School). Such reactions surely reflect not only distress for the victim, but also parents' profound concern for the safety of their own children.

One of the strongest indictments that can be leveled against research on fear of crime is the continuing failure of investigators to collect systematic data on altruistic fear, or even to recognize its existence. It is entirely possible that altruistic fear is as prevalent as personal fear (perhaps more so) and has consequences that are distinct from or amplify those arising from personal fear. Research on altruistic fear could also provide insights into the sociometry of fear in social organizations. For example, in family households, do wives fear for their husbands as much as husbands do for wives? Do they share equal fear for their children? How does the age or sex of children affect their parents' fear?

The Measurement of Fear

Fear of crime can be measured by soliciting self-reports from individuals or by monitoring physiological processes associated with fear. The emotion of fear is ordinarily accompanied by certain involuntary physiological changes, and these can be used as indicia to measure the presence or intensity of fear (Thomson 1979; Mayes 1979). One of the potential advantages of physiological measures of fear is that it enables the measurement or monitoring of fear *as it is occurring*, that is, in real time in natural settings. Because fear is often a fleeting emotion and may occur at inopportune times or places (e.g., late at night in a downtown parking lot), this is no minor advantage. Another related benefit of physiological measures of fear is that they eliminate many of the problems associated with self-reports, including errors in recall, demand effects, or reluctance to disclose emotions.

Physiological measures of fear have certain limitations, however. They cannot directly reveal the source of fear, i.e., the persons, things, or events to which the subject is reacting. Furthermore, they cannot distinguish fear of crime from other forms of fear (e.g., fear of accidents or threatening weather). These limitations may present few problems in controlled laboratory experiments (when, for example, subjects are presented with dangerous or innocuous scenes) because the cues or stimuli of interest can be isolated and confounding cues eliminated or controlled. However, the number and variety of cues that appear in natural settings suggest that physiological measures of fear may be of limited value in nonexperimental research. Moreover, the physiological changes commonly associated with fear can accompany other emotional states as well (Thomson 1979; Mayes 1979). Thus, for example, there appears to be little physiological basis for distinguishing between persons who react to a violent threat with anger and those who react with fear. Still another problem is that feelings of fear and physiological reactions appear to be more strongly coupled under some circumstances (when, for example, fear is intense) than others (Mayes 1979).

Despite these possible pitfalls, there is a pressing need to explore the uses of physiological measures of fear, because the payoff in knowledge is potentially great. Consider some of the questions that might be answered using a continuous, unobtrusive measure of fear:

- Does fear follow a reliable daily or weekly periodicity?
- What microenvironments—blocks, businesses (e.g., bars), neighborhoods—are most conducive to fear?
- How is fear affected by the presence or absence of companions or bystanders?
- Do certain kinds of persons (minority members, the homeless) evoke fear among some individuals?
- Is being alone in public places more frightening than being with strangers?
- In which types of routine activities—school, work, shopping, or home—is fear most pronounced?
- Does carrying a weapon outside the home reduce or actually exacerbate fear?
- What are the less obvious precautionary behaviors that people undertake in response to fear (e.g., scheduling habits, monitoring the whereabouts of others)?

Survey measures of fear

Survey research on fear of crime is extensive, but a truly bewildering variety of questions have been used by investigators over the years to measure fear of crime (see Ferraro 1995; Ferraro and LaGrange 1987; DuBow, McCabe, and Kaplan 1979). Much of this diversity stems from variation in the context stipulated in survey questions. Some questions ask about fear during the day; others, about fear at night. Some pertain to fear at home, whereas others question respondents about their fear in their own neighborhood or in their city. Still others ask respondents about their fear when alone or with others. Such sensitivity to context among researchers is commendable, but it is of little value unless such contextual variables are systematically varied and their effects evaluated. Unfortunately, that is rarely the case.

One item, however, has become the *de facto* standard for measuring fear of crime: “Is there anywhere near where you live—that is, within a mile—where you would be afraid to walk alone at night?” The item has become conventional not because it was chosen by social scientists, but because it has been routinely

used by the Gallup Organization and the National Opinion Research Center (NORC) to measure fear since the 1960s. During the past three decades, approximately 40 to 50 percent of Americans surveyed each year have responded affirmatively to this question (for a review, see Warr 1995a).

The Gallup/NORC item has been criticized (e.g., Ferraro 1995) on many grounds: it is hypothetical (how afraid would you be), is limited to nighttime, does not mention crime, and only crudely measures intensity. In fairness, the measured prevalence of fear obtained with this item is not radically different from that measured in other national surveys (see Warr 1995a), and the routine use of the item permits longitudinal comparisons of fear, if only in relative terms.

Much deeper issues are raised by questions of this kind, however. Almost two decades ago, Warr and Stafford (1983) asked residents of Seattle to report their everyday fear, not of "crime" in general, but of a variety of specific offenses ranging from violent crimes like homicide, rape, and robbery to various property and public order offenses. Even today, the rank order of offenses that emerged from their analysis remains startling to many. Murder, for example, was low on the list of fears, while residential burglary outranked all other offenses on fear. Warr and Stafford demonstrated that these findings were not anomalous or even counterintuitive. Contrary to common assumption, they showed, fear is not determined simply by the perceived seriousness of an offense. Instead, the degree of fear attached to particular crimes is a multiplicative function of the perceived seriousness and perceived risk of the offenses. To generate strong fear, an offense must be perceived as both serious *and* likely to occur. Residential burglary is the most feared crime in the United States because it is viewed as both relatively serious and rather likely. Murder, on the other hand, is perceived to be very serious but unlikely to occur.

Since the publication of Warr and Stafford's findings, only scattered offense-specific data on fear have been gathered (see Warr 1995a; Ferraro 1995; Haghghi and Sorensen 1996). These data generally corroborate the hierarchy of fears observed by Warr and Stafford (insofar as they use comparable offenses), but fear continues to be monitored primarily through generalized, omnibus measures of the sort used by Gallup and NORC. As a consequence, important questions about fear remain unanswered today. For example, when respondents report fear of "crime" in social surveys, what specific offenses do they have in mind? Are those offenses similar across individuals? The answer is that they almost surely are not; fear of rape, for example, is very pronounced among women but presumably not among men (see Warr 1985). On a different question, are particular precautionary behaviors, such as spatial avoidance and time-shifting (carrying out the same activity at an ostensibly safer time), linked to fears of particular crimes?

Another limitation of current survey data is that there are no time-series data on fear of individual offenses. Is rape feared today more than in the 1980s? Do offense-specific fears follow offense-specific trends in the incidence of crime, or do they respond to some “master” offense? Geographic variation in fear should also be considered. Do residents of large cities fear the same offenses as suburbanites or small-town residents? How far does the fear inspired by a particular incident spread, and how does it vary with the nature of the offense? Without systematic, offense-specific data on fear of crime, questions of this sort cannot be answered.

None of this is necessarily to discount the value of omnibus measures of fear. It is not unreasonable to assume that individuals can report an overall assessment of their fear about “crime” as a category of risk, any more than it is possible to list and measure all conceivable offenses that individuals might fear. Omnibus measures, in short, are useful as a complement to, but not a substitute for, offense-specific measures of fear.

Behavioral indicators of fear

Nearly all those who have investigated the emotion of fear agree that fear reveals itself through behavior, from the myriad responses of nonhuman species (distress cries, freezing, defecation, tonic immobility, or feigning death) to the complex and sometimes subtle avoidance behaviors of humans (Sluckin 1979). A major problem with behavioral indicators of fear in humans, however, is the difficulty of ascertaining exactly what people are *not* doing (or are doing) out of fear, and convincingly linking it back to fear. Is it obvious that a person with fear of heights is intentionally avoiding tall buildings, bridges, or amusement rides? Is it evident that a person who fears drowning showers rather than bathes for that reason?

However difficult it may be to establish, the link between fear and behavior underscores one of the great ironies of fear: Those most profoundly affected by fear—fear of flying, fear of automobile accidents—may rarely experience it because they have taken extraordinary measures to avoid the source of their fear (Kenny 1963). In the end, then, behavior may be the best indicator of fear, but the behaviors through which fear makes itself known are not always easily identifiable or detectable.

Transient public episodes of fear

When it comes to measuring fear of crime, what are the appropriate units of analysis (individuals, neighborhoods, cities, nations)? And what is the appropriate time interval for measurement (hourly, daily, monthly, yearly)? The answers depend, of course, on the question to be answered, but the conventional method

for measuring fear—annual surveys employing national samples—is apt to overlook crucial aspects of fear.

Consider an example. No close observer of American society will fail to notice that certain horrific criminal events seize the attention of the general public and become matters of almost universal discussion, speculation, and concern. Although some gain national attention, most remain matters of more local concern, affecting a particular city or portion of a city. In my own city of Austin, Texas, the gruesome murder of four teenage girls in a yogurt store created something approaching mass hysteria in the city and remained the lead story on local television news for many weeks thereafter. Landlords reported that it was difficult to lease apartments in the part of town where the crime occurred, and nearby business owners reported significant declines in revenues.

The annual national surveys ordinarily used to measure fear of crime are too coarse, both spatially and temporally, to capture these sorts of events. Put another way, the scale of such surveys simply does not match the scale of certain events that ought to be measured. Because of this, little is known about the natural history of localized urban “panics,” even though they are perhaps the most common social outbreaks of fear. How long do events of this kind ordinarily last? Does fear decay gradually or subside suddenly? Once initial media attention wanes, does fear decline as well? Does the arrest of a suspect affect the course of fear? Does fear decay at different rates among various segments of the population (young and old, male and female)? Even after fear fades, do events of this kind become part of the collective memory and lore attached to districts of a city? (“The south side is dangerous—remember that girl who was killed?”) Put differently, is there a permanent “residue” of fear that remains behind after such events?

Filling such gaps in existing knowledge will require surveys of social units much smaller than the nation as a whole. And because the events that incite such incidents are unpredictable and require pre-event (i.e., baseline) measures and repeated postevent measures, the only feasible research strategy is to implement routine surveys in selected jurisdictions. The best design would be a series of small, periodic (e.g., monthly) sample surveys in perhaps a dozen cities over several years. Capturing one transient episode is not enough to adequately describe such events because the duration and intensity of events is likely to depend on characteristics of the crime. (What age and sex were the victim and offender? Was the event provoked?)

Why is it important to measure smaller scale events of this sort? Apart from determining the frequency and geographic dispersion of such events, initial public reactions to criminal events often seem to be based on local media

reports that are sketchy, hasty, and so short on critical details (Did the victim know the suspect? Is the suspect still at large? Has this happened before?) that an intelligent evaluation of the event is all but impossible. In such situations, some individuals will assume the worst and act accordingly. Understanding the features of criminal events that determine the intensity and duration of public panics might lead to more judicious reporting on crime and less unnecessary fear.

Regulating Public Fear of Crime

Social scientists are inclined to approach fear of crime by asking the sorts of questions they raise about other human phenomena: What are its causes? What are its consequences? What are its contemporary and historical parameters (incidence, prevalence, social distribution)? Policymakers may be interested in these questions as well, but ultimately they must confront other immediate issues, some empirical and some normative. Can this phenomenon be controlled, and, if so, at what expense? Assuming it can be controlled, *should* it be controlled? Are there harms as well as benefits associated with intervention? These questions will frame our discussion of this complicated issue.

Should fear be controlled?

Imagine for a moment that we possessed a magic dial through which we could control or regulate fear of crime in the United States. Turn the dial to the left and fear immediately decreases proportionately; turn it to the right, and fear rises proportionately.

With the public interest in mind, no doubt our first inclination would be to reduce fear substantially by rotating the dial far to the left. Suppose, however, that the risks of crime are in fact real and substantial. Were we to greatly reduce fear of crime with our device, we would concomitantly increase the chances that individuals would fail to take necessary precautions to protect their own safety (or the safety of others), and thereby increase the risk of victimization.

Reducing fear, in other words, is not necessarily an unqualified good or cost-free; by relieving fear, we stand the chance of increasing public injury. On the other hand, were we to turn the dial too far to the right, people would engage in needless precautions and unnecessarily constrain their own lives. At the extreme is a "fortress society" in which citizens withdraw from public life altogether and everyday social intercourse is sharply curtailed.

Which way should the dial be turned, then, or should it in fact be touched at all? In the real world, of course, there is no magic dial or any direct way to

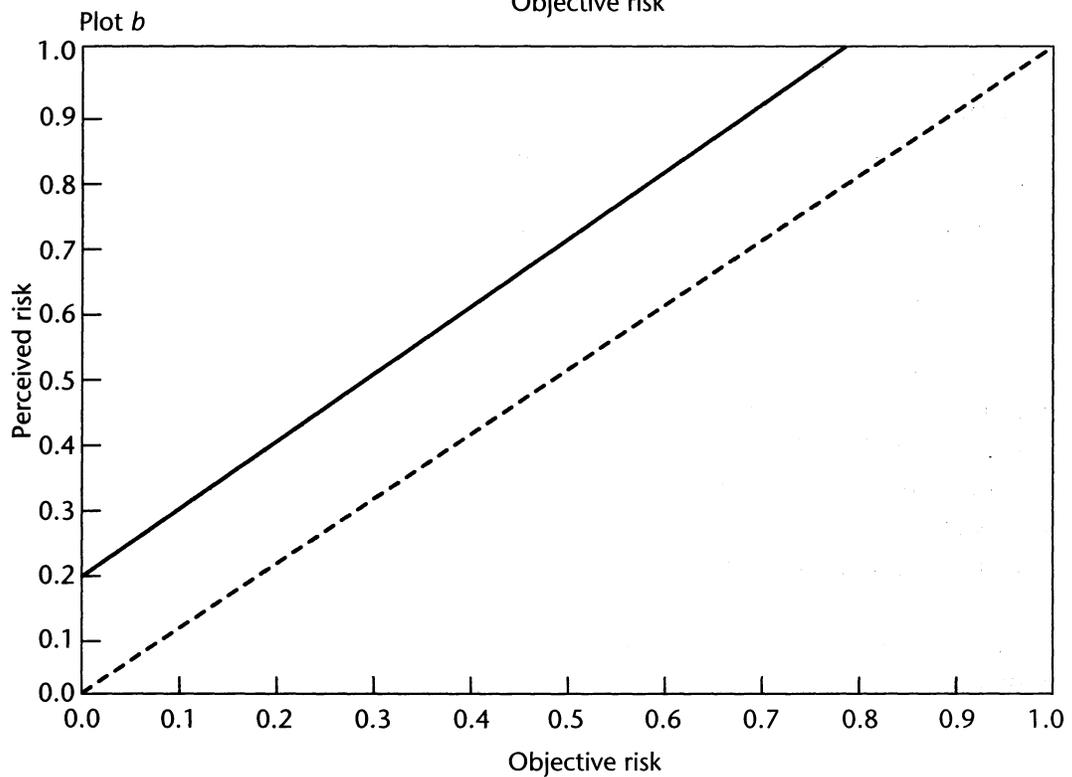
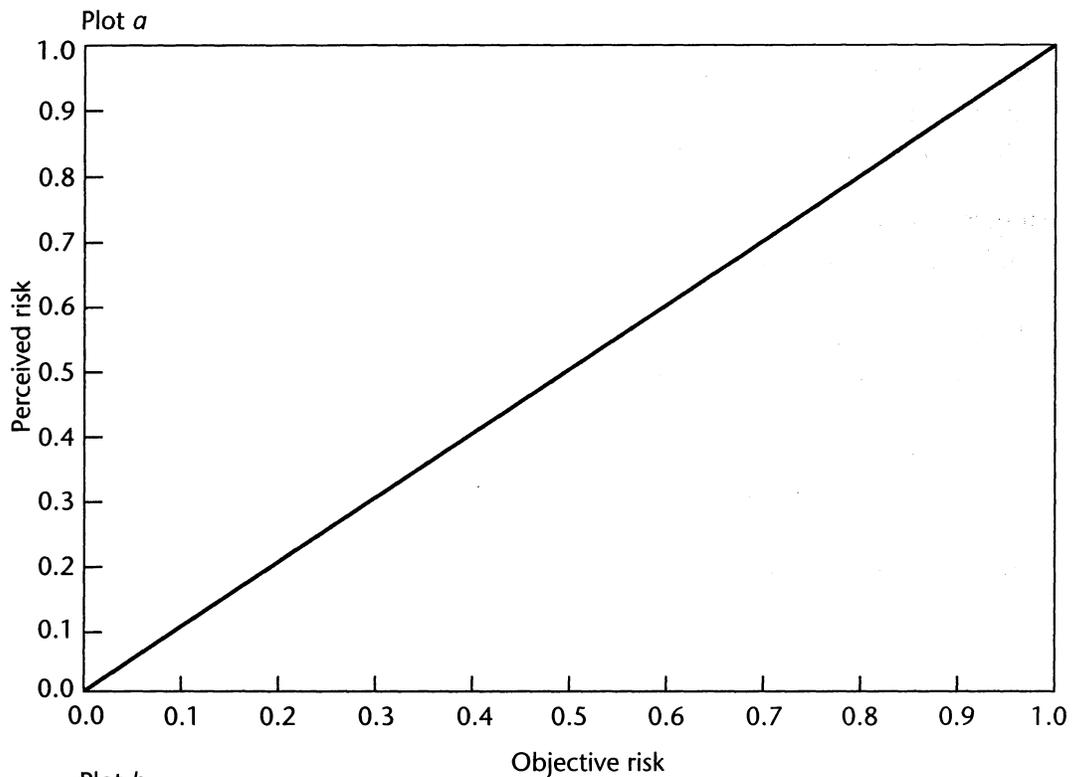
manipulate an emotion like fear (unless one were to propose dispensing sedatives or other pharmacological agents to the public). Instead, we must attempt to control fear by controlling its causes. As noted earlier, research by Warr and others consistently suggests that the proximate cause of fear for any one crime (that is, ignoring differences in the seriousness of crimes) is the perceived risk of that crime. Altering fear, then, requires altering perceptions of risk.

Reformulating the question somewhat, then, which way should the “perceived risk” dial be turned? The answer ultimately depends on the relation between perceived risk and objective risk. Plot *a* in exhibit 1 illustrates a situation in which perceived risk precisely matches objective risk; any increase or decrease in the latter (over time or place or across crimes) is always matched by a change in the former. In plot *b*, however, perceived risk always exceeds objective risk by a fixed amount; the public consistently overestimates the risk of victimization. In plot *c*, exactly the opposite is true; objective risk is greater than the public realizes. In the final plot, *d*, the relation between objective and perceived risk is more complex. When objective risk is low, the public overestimates risk; when objective risk is high, the public underestimates risk.

What are the policy options implicit in these examples? If the world truly operated as in plot *a*, the policy choice would be abundantly clear: Do not touch the perceived risk dial; public perceptions are accurate and existing levels of fear are justified. As noted earlier, many criminologists tacitly assume a tight connection between perception and reality when it comes to crime, and thus feel free (indeed, obliged) to set aside perceptions in favor of concentrating on crime reduction itself. The same seems to be true of politicians, who often advocate crime reduction without questioning whether public beliefs about crime are accurate or in need of alteration. Such approaches rest on an implicit but untested assumption, to wit, that any reduction in objective risk will be noticed and appreciated by the public.

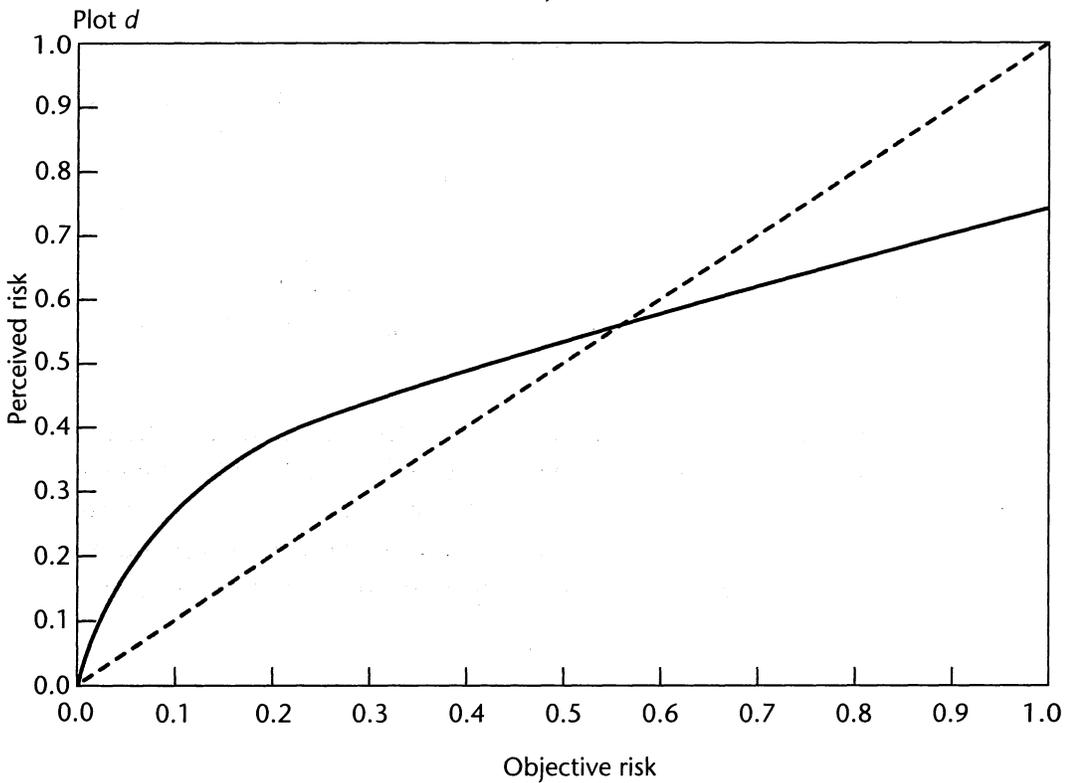
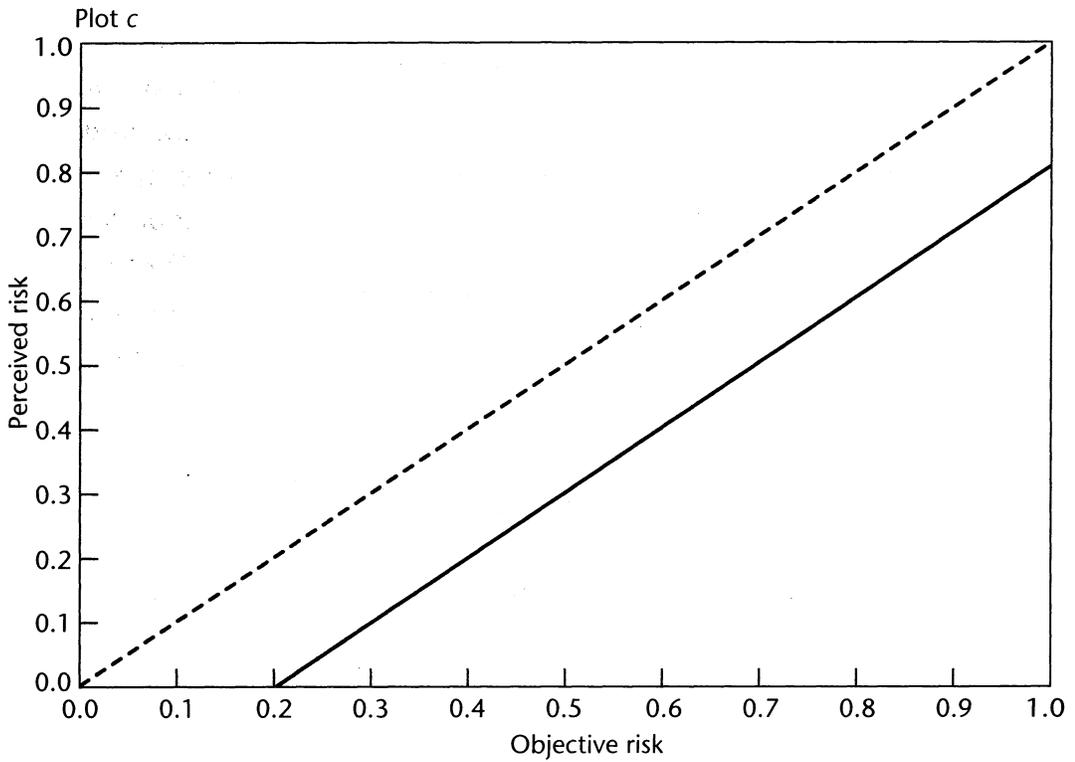
If the world were as in plot *b*—the public overestimates risk and is needlessly afraid—then one would want to turn the dial down and reduce what is clearly unnecessary fear. This is surely the most desirable and morally unequivocal situation because it results in a gain in the public good without any increased exposure to danger. This situation would be difficult to ignore once it is recognized, and it is every social engineer’s dream. If, conversely, plot *c* were true, we would be obligated to *raise* perceived risk so that the public would be suitably afraid. This is a morally defensible but politically difficult task, one in which the objective is literally to frighten the public. Though it might seem extreme, it is not without precedent; public campaigns against smoking, teenage pregnancy, AIDS, lead paint, dietary salt, and prescription drug risks are but a few examples (e.g., Slovic, Fischhoff, and Lichtenstein 1982;

Exhibit 1. Perceived risk versus objective risk



continued

Exhibit 1 (continued)



Fischhoff, Bostrom, and Quadrel 1997). Plot *d* presents a more complicated case, but the policy implication is no different from the preceding instances: take steps to bring objective and perceived risk into congruence.

One might perhaps want to entertain other policy options, of course. If fear reduction were the only consideration, one might argue that no change is required for plot *c*; in other words, ignorance is bliss. As we have seen, however, ignorance in this case is not the absence of harm, and there is no morally defensible theory that would justify that strategy. Similarly, the situation depicted in plot *b* would in all likelihood produce the lowest possible rate of criminal victimization, but at a personal and social cost that would be difficult to justify.

In the end, then, the question comes down to this: Which of these plots describes the real world? The weight of contemporary evidence suggests that the general public probably exaggerates the risk of serious criminal victimization in a way that resembles plot *d*.

What is this evidence? A small but persuasive body of studies in cognitive psychology (see Lichtenstein et al. 1978; Slovic, Fischhoff, and Lichtenstein 1979, 1980, 1982, 1987) indicates that individuals tend to significantly exaggerate the risk of rare lethal events (that is, causes of death like tornadoes, homicide, floods, fire, accidents, or botulism), while underestimating the risk of common lethal events (e.g., deaths due to heart disease, diabetes, or cancer). Slovic, Fischhoff, and Lichtenstein (1980, 1982) attribute this tendency to a common error of judgment arising from the availability heuristic (Tversky and Kahneman 1982) or the tendency to judge the frequency of events by the ease with which they can be recalled or imagined.

Why would individuals readily imagine or remember what are actually rare causes of death? Slovic, Fischhoff, and Lichtenstein (1980, 1982) cite evidence from Combs and Slovic (1979) showing that public perceptions concerning the frequency of causes of death closely match the frequency with which those causes are reported in newspapers. Newspaper accounts, in turn, are glaringly at odds with reality:

[M]any of the statistically frequent causes of death (e.g., diabetes, emphysema, various forms of cancer) were rarely reported by either paper during the period under study. In addition, violent, often catastrophic, events such

Findings suggest that the public is likely to exaggerate the frequency of rare, serious crimes and underestimate the frequency of more common, less serious ones.

as tornadoes, fires, drownings, homicides, motor vehicle accidents, and all accidents were reported much more frequently than less dramatic causes of death having similar (or even greater) statistical frequencies. For example, diseases take about 16 times as many lives as accidents, but there were more than 3 times as many articles about accidents, noting almost 7 times as many deaths. Among the more frequent events, homicides were the most heavily reported category in proportion to actual frequency.

Although diseases claim almost 100 times as many lives as do homicides, there were about 3 times as many articles about homicides as about disease deaths. Furthermore, homicide articles tended to be more than twice as long as articles reporting disease and accident deaths. (Slovic, Fischhoff, and Lichtenstein 1982, 468)

These investigators did not insist on a causal connection between media reports and public perceptions, but they suggested that the pattern of errors in the two is much too similar to be coincidental.

How do perceptions about one class of hazards (causes of death) relate to another (the risk of criminal victimization)? Because the most serious crimes (homicide, rape, robbery) are also the rarest forms of crime, the preceding findings suggest that the public is likely to exaggerate the frequency of rare, serious crimes and underestimate the frequency of more common, less serious ones. In the early 1980s, Warr (1980; see also Bordley 1982) presented direct evidence of this phenomenon, showing that the objective and perceived incidence of offenses in four cities were related by a power function ($y=aX^b$). That is, people tended to systematically overestimate the frequency of rare offenses while underestimating the frequency of common ones. Public perceptions

were remarkably accurate as to the relative frequency of different crimes (for example, people recognize that homicide is less common than burglary), but considerably less accurate as to absolute frequencies.

The mass media are thus a powerful amplifying mechanism when it comes to crime; information known only to a few can within hours or days become known to thousands or millions.

Aside from these findings, there is another reason to suspect that the true relation between perceived and objective risk probably resembles plot *d* (or perhaps *b*). When the general public is asked where they obtain most of their information about crime, the resounding answer is the mass media, especially news coverage of crime. Graber (1980), for example, reported that 95 percent of respondents in her survey identified the media as their primary source of information on crime, although 38 percent cited other sources as well (conversations or, more rarely,

personal experience). Skogan and Maxfield (1981) found that more than three-quarters of respondents in the three cities they surveyed reported watching or reading a crime story on the previous day (44 percent had read a newspaper crime story, 45 percent had watched a crime story on television, and 24 percent had done both). The mass media are thus a powerful amplifying mechanism when it comes to crime; information known only to a few can within hours or days become known to thousands or millions.

The media are most likely to report precisely those crimes that are least likely to occur.

What is the image of crime presented in the mass media? A number of forms of distortion in news coverage of crime have been identified and documented, distortions that tend to exaggerate the frequency and the seriousness of crime. In the real world, for example, crimes occur in inverse proportion to their seriousness; the more serious the crime, the more rarely it occurs (e.g., Erickson and Gibbs 1979). Thus, in the United States, burglaries occur by the millions, robberies by the hundreds of thousands, and homicides by the thousands. In news coverage of crime, however, the emphasis is on “newsworthiness,” and a key element of newsworthiness is seriousness; the more serious a crime, the more likely it is to be reported. By using seriousness as a criterion, however, the media are most likely to report precisely those crimes that are least likely to occur (Skogan and Maxfield 1981; Sherizen 1978; Sheley and Ashkins 1981; Roshier 1973), or exactly the same pattern outlined previously for lethal events.

Among other things, this “mirror image” depiction of crime means that the media place extraordinary emphasis on violent crime. Skogan and Maxfield (1980) reported that homicides and attempted homicides constituted one-half of all newspaper crime stories in the cities they examined, even though homicides are but a minute fraction of all offenses. Furthermore, the number of homicide stories reported in city newspapers, they found, did not closely match the actual homicide rates of the cities examined, suggesting that the amount of space devoted to crime has more to do with the “newshole” allocated to crime by editors than with the true crime rate.

News coverage of crime has been criticized on other grounds as well, including the practice of using crime news as “filler” when other news is slow, the use of crime news (“If it bleeds, it leads”) to attract larger audiences, and the tendency to report trends in crime using numbers rather than rates, thereby ignoring changes in population (see Graber 1980; Warr 1980, 1994, 1995b).

The fact that the media present a distorted image of crime, of course, is no guarantee that the public believes or heeds what it sees, hears, and reads. And public

perceptions (or exaggerations) of the incidence of crimes do not necessarily translate into estimates of personal risk (Slovic, Fischhoff, and Lichtenstein 1982). Still, the evidence on public perceptions of crime and media distortion of crime news is strikingly corroborative, and it is difficult to believe that the media have little or no effect on perceptions, especially when the public cites the media as their primary source of information on crime and spends so much time watching, reading, and listening to the media (Skogan and Maxfield 1981).

Given the gravity of fear as a social problem and the presumptive role of the mass media, it is truly astonishing to find that there is virtually no systematic research that assesses the impact of the media on public perceptions of crime or fear of crime. To be sure, there is known to be a positive correlation between fear of crime and the number of hours spent watching television (Skogan and Maxfield 1981), but the causal direction is unclear and the correlation may well be spurious with respect to age and other viewer characteristics.

In the end, the causal influence of media crime coverage cannot be established without simultaneous measurements of (1) media content, (2) public exposure to that content, and (3) the postexposure effects of media communications. Such research is difficult to conduct in natural settings because of the enormous quantity and variety of media and interpersonal messages on crime to which the public is exposed (e.g., Graber 1980). Remarkably, a great deal of Federal money is spent today to document the extent of crime and violence on television, but rarely is there any accompanying research on the effects of such televised violence on those who are exposed to it. It is, in essence, a research design with no dependent variable.

One study that approaches an ideal design and points the way for future research was conducted by Heath (1984). She questioned samples of newspaper readers in 36 cities and examined their fear of victimization in light of the characteristics of the newspapers they read. Heath found that fear was higher among readers of newspapers that emphasized local crimes and crimes that were sensational (bizarre or violent) or random (apparently unprovoked). However, reports of sensational or random crimes evidently reduced fear if those crimes were not local. Apparently, readers were reassured by learning that such crimes were occurring to other people in other places.

Can fear be regulated?

The discussion thus far leads to a tentative conclusion that the public exaggerates the risk of serious criminal victimization. It is worthwhile to reiterate that this situation, if true, is the most desirable problem to remediate; fear can legitimately be reduced without any attendant increase in the risk of victimization.

Let us turn now to the second question posed earlier. Assuming that fear ought to be regulated, can it be regulated, and, if so, how?

There are two general approaches to this problem. One might be called self-corrective, meaning that it focuses on means to alter the way that crime is presently depicted in the mass media. The other approach could be described as counteractive, meaning that it attempts to discount or replace messages promulgated through the mass media.

Let us begin with the former approach, bearing in mind that media news coverage is the public's primary source of information about crime. Assuming that such coverage substantially affects public perceptions of crime, how might news coverage of crime be changed for the better?

Consider the characteristics of everyday news coverage. Crime is ordinarily reported in the form of isolated, discrete incidents ("three young adults were injured today in a standoff with police") or occasionally as counts ("thus far this year, 13 robberies have been reported to the police"). (See, for example, Graber 1980.) These reported events are not a complete or exhaustive list of all crimes. Instead, they are selected from a much larger pool of crime events available for reporting. What is more, the selection process is not random or representative, but quite the opposite. Not only are violent crimes disproportionately emphasized (particularly homicides), but crimes may be selected merely because they are odd or unusual, involve prominent persons or public figures, or fit a preestablished journalistic theme like "crimes against the elderly" or "careless tourists" (Skogan and Maxfield 1981; Gordon and Heath 1981; Sherizen 1978; Ericson, Baranek, and Chan 1987; Fishman 1978, 1981).

To imagine the consequences of such reporting practices, consider this question: Could an individual reliably estimate the magnitude and causes of population growth in a city through isolated, incomplete, and non-representative interviews with those who have left or have moved in?

From a public information perspective, what often seems to be missing in news coverage of crime is not raw information on criminal events (on the contrary) but an informed perspective about crime risks. Only occasionally are criminal events presented as population-based rates, from which one could estimate personal risk or risk to loved ones. And rarely are such rates placed in any sort of seasonal, historical, demographic, or geographic context. From the point of view

Trying to detect patterns or achieve valid inferences about crime from isolated, sporadic news accounts is an exercise in futility.

of readers or viewers, trying to detect patterns or achieve valid inferences about crime from isolated, sporadic news accounts is an exercise in futility.

What is seriously lacking in news reporting, and might be of greatest benefit to the public, is information about the risk of criminal victimization relative to other aversive or benchmark life events. To illustrate, what is the probability that I will be robbed this year compared with the chances that I will be involved in a serious automobile collision, eat contaminated food in a restaurant, contract an infectious disease at work or school, or suffer a heart attack? How does my age, sex, racial/ethnic identity, or location affect my chances? Many Americans, including journalists, might be surprised to learn that they are more likely to be a victim of suicide than homicide, that automobiles kill more individuals than all violent crime, or that, as a group, children face greater danger from their parents than from strangers.

The didactic value of what I have called informed perspective can be illustrated by comparing two possible television news accounts of the same hypothetical event. The first reads:

A homicide occurred late last evening at 223 East Lansing. The victim, a 23-year-old male, was stabbed twice and died shortly thereafter at Our Lady General Hospital. According to the police, no arrest has been made.

Now add these words:

Fewer than 1 in 10,000 persons living in our city are victims of homicide each year. Most, as in this case, are young males who die in alcohol-related arguments with persons they know. The number of homicides thus far this year—27—is no higher than average for the previous 5 years, and two-thirds of those homicides occurred within the same three census tracts of the city. For more information, contact the Metro Police Department at 366-8942.

Journalists may object that the latter version is too long and dull, but a crime worth reporting is surely worth properly contextualizing. In defense of journalists, police crime reports often are sketchy or incomplete, and there is the pressure of deadlines. On the other hand, it is not the police who are eager to release incomplete reports, and reporters' professions about public safety are sometimes little more than transparent ruses to be the first to release the story. In any event, audiences are likely to fill in sketchy or missing information by assuming the worst, which is all the more reason to place reports of crime within a larger factual context.

Counteractive measures

Aside from (or in addition to) changing media crime-reporting practices, information on crime can be promulgated through alternative channels. Messages about crime can be disseminated through a variety of means, including pamphlets; billboards, transit ads, and other signage; magazine and newspaper ads; Web sites; and oral presentations at public gatherings. Such a strategy might seem inconsequential compared with the awesome power of television and newspapers, but messages of this type have figured heavily in public campaigns about smoking, heart disease, and other health risks (see generally, National Research Council 1989).

When it comes to crime, at least two public agencies are a natural choice for conveying such messages. One is the municipal police department. Most modern police departments have a public information office to dispense information on crime to persons (often journalists) on request. There is no major logical or logistical jump in moving from a reactive function of this type to a more proactive version of public information. The logical connection between crime and the police makes the police a perfect agent for crime communications, and, notwithstanding occasional scandals, the police enjoy enormous public support in the United States (e.g., Warr 1995a). I witnessed this function of the police when they embarked on a door-to-door campaign to inform residents of one Austin neighborhood that several rapes had recently occurred there. Where the costs of distribution pose a problem to police departments, there is no lack of civic organizations and volunteers willing to place pamphlets or booklets in mailboxes or on front porches.

Slovic, Fischhoff, and Lichtenstein (1982, 484) have also argued that a proper setting for communicating risks is the school:

Informing people, whether by warning labels, package inserts, or extensive media programs, is but part of the larger problem of helping people cope with the risks and uncertainties of modern life. We believe that some of the responsibility lies with our schools. Public school curricula should include material designed to teach people that the world in which they live is probabilistic, not deterministic, and to help them learn judgment and decision strategies for dealing with that world. [T]hese strategies are as necessary for navigating in a world of uncertain information as geometry and trigonometry are to navigating among physical objects.

Although modern schools are swamped with demands for their time and suggestions about their curriculums, a case can be made that the risks of crime are of sufficient size and gravity that at least some time ought to be devoted to them, if only to alleviate unnecessary fear throughout life. In occasional

lectures to schools and churches, I have been struck by the almost desperate hunger of many people for objective information on crime and its risks.

No matter who the messenger is, what is the content that messages about crime should convey? In recent years, an entirely new field of research in science has emerged known as risk communication (cf. National Research Council 1989). Concerned with the methods, problems, and efficacy of communicating risk to the general public, the field has concentrated largely on new technological risks (nuclear power, pesticides, toxic waste disposal), medical/health risks (recombinant DNA, smoking, seat belt use, high cholesterol, alcohol abuse, cancer) and both natural and manmade disasters (hurricanes, floods, aircraft crashes, lightning, tornadoes, earthquakes). No one, to my knowledge, has grappled with risk communication about crime, but the lessons of this field are nonetheless useful and enlightening.

One lesson brings to mind the primary obligation of the physician: First do no harm. The effect of a communication can only be ascertained empirically. Untested messages can have unintended consequences, and therein lies the danger:

Poor risk communications may cause more damage than the risks they are intended to control. They can lead to wrong decisions by omitting key information or failing to contradict misconceptions. They can create confusion by prompting inappropriate assumptions or emphasizing irrelevant information and produce conflict by eroding the audience's faith in the communicator. They can cause recipients to be unduly alarmed or complacent or to undertake ineffective actions. Because communicators' intuitions about recipients' perceptions cannot be trusted, there is no substitute for empirical validation. (Bostrom et al. 1994, 796)

Fischhoff, Bostrom, and Quadrel (1997, 993) similarly observed:

Effective risk communication requires careful empirical research. Poor risk communication can often cause more public health (and economic) damage than the risk it attempts to describe. One should no more release an untested communication than an untested medical device.

In short, risk communications must be pretested before they are disseminated.

Fischhoff (1989) argues that an essential prerequisite for designing risk communication is the need to know what the public does not know, a matter that ordinarily requires empirical research and cannot simply be presumed. In that connection, Fischhoff, Bostrom, and Quadrel (1997) offer an ingenious idea for altering public perceptions of risk. In keeping with evidence that people are

often highly overconfident about the information or beliefs they hold, these investigators suggest that one function of communications is to give people “the appropriate degree of confidence in their beliefs,” especially in cases “where people confidently hold incorrect beliefs that could lead to inappropriate actions” (1997, 997). As an example of misplaced confidence, they cite an investigation showing that a majority of teenagers are not aware that a single beer affects driving ability as much as a shot of vodka and that mistaken teenagers are usually very confident about their incorrect information. Studies of this kind are particularly useful for identifying misconceptions that need to be targeted in messages.

How should communications about risk be constructed? What elements should they contain? Information about risk is often highly technical, but technical terms and examples should be avoided (Covello, von Winterfeldt, and Slovic 1987; Fischhoff 1989). Risks can be presented in everyday terms using alternative examples. For example, the proportion of Americans who are murdered each year in the United States (fewer than 1 in 10,000 per year) is roughly the same as 1 day in 27 years, or 1 inch in 833 feet, or 1 gallon in a home swimming pool. By contrast, the crude probability that a household will be burgled in the United States is about 1 in 10–20 per year, or roughly the chance of drawing two consecutive cards of the same suit from a fresh poker deck. Cross-hazard comparisons (where the risk of, say, murder is compared with the risk of an auto accident, disease, or lightning strike) can be useful for illustrating risk as well, but they can be difficult to interpret or understand if not properly constructed (Slovic, Fischhoff, and Lichtenstein 1982).

Information in communications must also be relevant to the audience:

Poorly chosen information can . . . be seen as wasting their time (indicating insensitivity to their situation), . . . can take up the place (in the media or school) that could be filled with pertinent information (imposing an opportunity cost), and . . . can lead them to misunderstand the extent of their knowledge. (Fischhoff, Bostrom, and Quadrel 1997, 997)

Communicators must also ensure that events or risks discussed in the communication have the same meaning to recipients as to themselves (e.g., Do respondents know what a burglary is? Do they confuse it with robbery?), and they need to be honest and straightforward about the limitations of their own information (Fischhoff 1994; Fischhoff, Bostrom, and Quadrel 1997). The public is often skeptical of experts and government officials, and honesty about the accuracy of their estimates can help to offset mistrust.

It is impossible to fully survey the literature on risk communication here, but it is useful to offer two summaries of effective communications by prominent

investigators in the field. These are not merely seat-of-the pants recommendations but careful statements based on extensive research and experience. First, in an appendix to a National Academy of Sciences conference on risk assessment, Covello, von Winterfeldt, and Slovic (1987, 117–118) offer the following advice on risk communication:

- Use simple, graphic, and concrete material, avoiding technical or specialized language wherever possible.
- Compare risks within a carefully defined context that is relevant to the target audience.
- Avoid comparisons of risk that may appear to the audience to be non-comparable because of different qualitative characteristics—for example, the risk of smoking compared to that of living near a nuclear power plant.
- Understand and recognize qualitative concerns, such as concerns about catastrophic potential, dread, equity, and controllability.
- Identify and explain strengths and limitations of different risk measures, and present (whenever possible) alternative indexes of risk—for example, measured or expected fatalities or incidences of diseases for the entire population and for the most- and least-exposed individuals.
- Identify, acknowledge, and explain uncertainties in risk estimates.
- Provide opportunities for people to learn how to interpret risk information.
- Relate on a personal level—that is, when people ask personal questions such as “Can I drink the water?” respond in a personal way without minimizing risks and uncertainties.
- Recognize the power of subtle changes in the way that information is presented and use such knowledge responsibly.
- Understand and recognize that health and environmental debates often involve much broader considerations, including political values and ideologies.

In addition, Fischhoff, Bostrom, and Quadrel (1997, 998) offer this advice:

Once information has been selected, it must be presented in a comprehensible way. That means taking into account the terms that recipients use for understanding individual concepts and the mental models that they use for

integrating those concepts. It also means building on the results of research on text comprehension. That research shows, for example, that comprehension improves when text has a clear structure and, in particular, when the structure conforms to recipients' intuitive representation of a topic; that critical information is more likely to be remembered when it appears at the highest level of a clear hierarchy; and that readers benefit from 'adjunct aids,' such as highlighting, advanced organizers (showing what to expect), and summaries. Such aids might even be better than full text for understanding, retaining, and being able to look up information.

Fear of Crime and the Perceived Seriousness of Offenses

The discussion thus far has concentrated on public perceptions of risk and the rationales and mechanisms for altering such perceptions. Although altering perceptions of risk is surely a defensible strategy, it is not the only means for reducing fear. Earlier we saw that the fear evoked by different crimes depends not only on their perceived risk but also on their perceived seriousness. Crimes vary enormously in their perceived seriousness, from homicide to trespassing and rape to shoplifting, and there is also variation (though to a much smaller degree) among individuals in the perceived seriousness of any particular crime (Wolfgang et al. 1985; Warr 1993, 1994).

The importance of seriousness when it comes to generating fear suggests that fear can be regulated or controlled by altering the perceived seriousness of crimes. At first glance, that suggestion may seem preposterous. Do we reduce fear of homicide by convincing people that homicide is not a serious crime? Of course not. However, one of the elements that enters into peoples' judgments of seriousness is the perceived harmfulness of a crime (see Warr 1993), and there is evidence that people sometimes attach greater consequences to criminal events than is warranted.

Some years ago, Warr (1985) uncovered strong correlations between fears of certain crimes. Examined closely, these configurations of offenses often consisted of crimes that can occur contemporaneously or in sequence (e.g., burglary and rape, robbery and homicide). The strong correlations between fear of these offenses suggested that the general public often views these offenses as crimes that ordinarily accompany one another in the same event (what Warr called "perceptually contemporaneous offenses").

Perception and reality are sometimes at odds, however, when it comes to criminal events. For example, rape and homicide appear to be perceptually

contemporaneous offenses for many women (“If I’m raped, he’ll probably kill me.”) (Warr 1985). But rape rarely results in death. In 1996, for example, fewer than 1 in 1,000 rapes and attempted rapes known to the police in the United States resulted in death (U.S. Department of Justice, Federal Bureau of Investigation 1997). Because many rapes (but not homicides) fail to come to the attention of the police, that figure actually overstates the risk. Although rape is a serious crime, it is not ordinarily a lethal event.

Similarly, many people seem to associate residential burglary with violent attack or injury, but burglaries normally occur when no one is home, evidently because burglars do not want to be injured by armed homeowners (Miethe and McCorkle 1998). The point is that educating the public about the likely course and consequences of victimization might well prove to be an effective and morally defensible strategy for reducing fear. As in the earlier discussion of perceived risk, it appears to be a “win win” situation; fear can be reduced without deleterious side effects.

Much of the public also appears to be unaware that the outcome of criminal events often depends in part on the actions of victims. In robberies, for example, there is a strong correlation between resistance and the probability of injury, and law enforcement agencies often recommend against resistance in such situations. Risk communications notifying the public of these facts and discouraging resistance might be effective in reducing fear as well as unnecessary injury to those who become victims.

Fear and Cues to Danger

What has been said thus far about controlling fear bears more directly on anxiety about crime (concern about future victimization) than fear of crime in the strict sense (reactions to immediate threats). In everyday life, fear of crime (strictly defined) is most likely to occur as people navigate their environment away from home—walking to school, grocery shopping, traveling to work, going out for entertainment, running errands—and encounter signs of danger in the environment.

What are such signs of danger? Using a factorial survey design, Warr (1991) identified several cues to danger that affect people in public places. One particularly potent cue is darkness; by its very nature, darkness obscures potential threats that may lurk in the vicinity. Another cue to danger is novelty; unfamiliar environments are more frightening than familiar ones (a phenomenon not limited to humans; see Russell 1979). Still another cue is the presence of bystanders or companions. The presence of other people in the immediate

vicinity ordinarily acts to alleviate the fear that individuals would otherwise feel if they were alone. This calming effect does not operate, however, if those “others” are perceived to be dangerous persons. Warr found that young males are frightening to many individuals, and few sights are more alarming to the public than a group of young males.

In addition to these cues, a number of investigators have examined various “signs of incivility” that can provoke fear (cf. Ferraro 1995). These include physical features of neighborhoods like graffiti, broken windows, trash and litter, stripped cars, or abandoned buildings, and social cues like beggars or homeless persons, raucous groups of young people, drug sellers or users, and prostitutes. Empirical evidence regarding the potency of such cues in producing fear is generally supportive (LaGrange, Ferraro, and Supancic 1992), although largely indirect, and investigators rarely control for objective crime rates when examining the effects of incivilities.

Manipulating environmental cues to danger offers a concrete and potentially powerful means for regulating public fear of crime. This is perhaps more feasible than altering established practices of news coverage or deemphasizing crime in popular entertainment. In many cases, the costs of adopting this strategy are likely to be minor—painting over graffiti, picking up litter and waste, or improving lighting. The efficacy of such measures could be readily measured by having impartial audiences judge the perceived safety of an area before and after implementation. The results are likely to show what architects and shopkeepers have known for centuries: that a clean, orderly, and aesthetically pleasing environment draws crowds and creates its own social definition of place.

Manipulating streets and neighborhoods to reduce fear, however, is not altogether uncontroversial. For example, should a genuinely dangerous neighborhood be made to appear less frightening, even if it remains dangerous? The risk, of course, is that innocent citizens will fail to recognize—indeed, be lured to—a location that is deceptively dangerous. On the other hand, the physical improvement of a neighborhood may help to restore community control and actually reduce the risk of criminal victimization. That is the sort of process envisioned by Wilson and Kelling (1982) in their famous broken windows hypothesis, but the evidence for it remains uncertain (e.g., Skogan 1990).

There is no easy answer to the question, but commercial and residential areas that appear frightening and dangerous are surely doomed in the long run. Perhaps the best answer to the problem is to upgrade law enforcement efforts, if only temporarily, in areas that are undergoing improvement, so that changes in apparent safety are accompanied by changes in *real* safety as well, and dangerous places do not become an invitation to unwary visitors. Beyond this, the

sorts of cosmetic but symbolically important measures discussed previously are surely justified—indeed, almost imperative—in areas that have experienced real declines in objective risk. To look dangerous and actually *be* safe is a tragedy that demands attention.

The Rationality of Fear

All strategies for controlling public fear of crime presuppose an answer to this question: How much fear is justified or appropriate? The premise of this chapter is that fear is justified when perceived risk is congruent with objective risk. In research on fear of crime, however, the issue has often been framed by reference to the “rationality” of fear. Early investigators often expounded on the rationality of public fear of crime, particularly when applied to specific population groups who were viewed as inordinately afraid (women, the elderly).

When applied to fear of crime, however, the concept of rationality is an ill-advised notion. As it is most often used with respect to fear, the concept of rationality implies a high degree of correspondence between some subjective phenomenon (perceived risk, fear) and an objective standard or counterpart. Such a comparison is possible when it comes to risk because risk has both a subjective and objective component; one can estimate objective risk for many hazards and ask people to report their perceptions as well.

But the same is not true when it comes to fear, because fear is not simply a function of perceived risk. As we have repeatedly observed, fear also depends on the perceived seriousness of crimes, which in turn depends on the value that individuals place on persons and property. In general, people tend to judge the seriousness of crimes in a similar fashion (Wolfgang et al. 1985), but in real life there is ample room for variation. What price is to be put on a lost wedding album, a recording of a deceased parent, a lifetime collection of art, or the life of a favorite uncle? Ultimately, such valuations are wholly subjective and personal, and economists and insurers aside, attempts to quantify or objectify such matters are likely to prove futile.

Another reason for concentrating attention on perceived risk rather than fear itself is that the same level of perceived risk often produces different levels of fear among different people, especially between men and women and young and old (Warr 1984). The reasons for this seem to have a lot to do with perceptually contemporaneous offenses. Among women, for example, the threat of rape often carries over to other crimes. What for men is the perceived risk of robbery would for many women be the perceived risk of robbery, plus rape, plus additional injury (Warr 1985; Ferraro 1996). Yet even when two individuals

react differently to the same perceived risk of the same hazard, it would be difficult to characterize one person's fear as more "rational" than the other. Such differences in fear are likely to stem, once again, from the value placed on persons and property.

In the end, social scientists may legitimately judge the perceptions or information that underlie people's reactions to crime and gauge the gap between perception and reality. They are in no position, however, to tell people the value they should place on the elements of their lives and how much they should fear their loss or destruction.

There are entire industries in the United States that rely on fear of crime to sell products and services.

The Selling of Fear

Earlier we noted some distortions that arise in media news coverage of crime. Though it is beyond dispute that crime is sometimes used by newspapers and networks to attract readers or viewers, it is probably fair to say that media misrepresentations of crime are often inadvertent rather than intentional. The individual and social consequences of fear are so substantial, after all, that it is difficult to believe that any organization or individual would deliberately increase fear merely for reasons of self-interest.

In fact, however, there are entire industries in the United States that rely on fear of crime to sell products and services, from home security systems, anti-auto theft devices, and travelers checks to personal security devices (sprays, alarms, and other weapons), property insurance, and cellular phones. Some firms are responsible and circumspect in the claims they make for their products. Others deliberately exaggerate or dramatize the risks of criminal victimization in an effort to frighten potential purchasers into buying products, some of which are of questionable utility.

I personally experienced the extent of such practices a few years ago when I was building a new home and was contacted by a national home security service that wished to meet with me and explain their services. After politely listening to the agent, I told him that I wanted some time to think the matter over. On hearing this, the agent blithely told me that the last customer who had postponed just a few days came to deeply regret it because his daughter was raped by an intruder during the time he took to reach a decision. I did not believe this unlikely story, and perhaps the salesman did not expect me to. But some would believe it and would be pressured into a purchase by force of their own fear.

In a recent commercial that repeatedly aired on national television, a young woman is driving on a lonely road at night, when a man in a pickup truck suddenly appears from behind and pursues her. The narrator asks the viewer what he or she would do if pursued by a such a “human predator” and offers advice on how to get out of the situation. The logical connection between this staged incident and the company itself (a major petroleum company) was tangential at best. It seemed that the company was deliberately provoking fear, particularly among younger women, in an effort to trumpet its concern for public safety and garner public trust and gratitude.

There is no law, of course, against using fear of crime as a sales tool, and the rule of *caveat emptor* applies to crime prevention as much as any other realm of commerce. But there is something deeply cynical about exploiting people’s concern for their safety (and their loved ones) for monetary reasons. If only as a research question, it would be intriguing to know whether certain segments of the population—the aged, those who live alone (widows and widowers), students, young women—are targeted by such industries for special attention and the degree to which fraudulent claims are used to sell products and services. To be sure, one of the strange ironies of life is that, even if they *are* fraudulent and unnecessary, such products may actually function to reduce fear among those who decide to invest in them.

If fear is useful as a sales device, it also has value to politicians, who are sometimes quick to exploit it as a political tool. By some accounts, the 1968 election campaign of Richard Nixon, with its emphasis on law and order, was the first to capitalize on crime and fear of crime for political advantage. In the Bush/Dukakis presidential contest, the infamous Willie Horton commercials appeared to play a pivotal role. Today crime continues to figure heavily in local and national political campaigns, and there appears to be little prospect for change. In a just world, the cynical exploitation of fear for political purposes would be appreciated for what it is. Yet, if nothing else, the eagerness of political figures to capitalize on public fear of crime is testimony to its central place in modern life.

The Consequences of Fear: The Big Picture

In the final analysis, what makes fear of crime so important as a social problem is the depth and breadth of its consequences for our society. Over the years, investigators have identified many behavioral precautions associated with fear of crime. These range from relatively trivial and nearly universal behaviors (e.g., turning on lights and locking doors when leaving home) to more personally and socially consequential actions (not leaving the house at night or going out alone) (cf. Skogan and Maxfield 1981; Warr 1994).

What is often missing in research on fear of crime, however, are studies of the large-scale social consequences of fear. To illustrate, it appears that the ecology of American cities is regulated to a considerable degree by fear of crime. According to survey data, the single most common reaction to fear of crime in the United States is spatial avoidance; that is, staying away from places that are perceived to be dangerous (Warr 1994).

The ecology of American cities is regulated to a considerable degree by fear of crime.

In surveys of Seattle and Dallas, for example, 63 percent and 77 percent of respondents, respectively, reported that they “avoided certain places in the city,” and when Dallas residents were asked to identify the most dangerous areas of their city, more than four of five reported that they did not go near or through those areas regularly. Along with spatial avoidance per se, fear of crime also seems to affect the routes that people take when they travel, the form of transportation they employ, and the times they choose to leave their residence (see DuBow, McCabe, and Kaplan 1979; Warr 1994).

Such habits of avoidance must inevitably affect commerce, road use, leisure activities, and social interaction. Retail businesses that are located in putatively dangerous areas are likely to suffer a shortage of customers, and reputedly dangerous neighborhoods are likely to find themselves socially isolated (Conklin 1975; Skogan 1990). Remarkably, however, there is no systematic evidence on the financial impact of fear of crime on retail business, nor evidence on the degree to which fear isolates neighborhoods from ordinary social intercourse. The same is true when it comes to leisure activities. The impact of fear on interstate and intercity tourism is an obvious topic for research, but aside from occasional journalistic accounts (as in the infamous murders of tourists in Florida in the early 1990s), there is little research on the economic consequences of fear on tourism. Additionally, public use of facilities such as parks, beaches, campsites, and other recreational areas is surely affected by fear, but the nature and magnitude of this effect remains unknown.

There is another potential consequence of fear. Some commentators have remarked on the apparent tendency of Americans to spend increasing amounts of time, including their own leisure time, in their own homes, in what amounts to a general withdrawal from the outside world. The trend is sometimes described in humorous terms (like “couch potato”) and supported by reference to sale of items such as big-screen TVs, home theaters, and hot tubs. Assuming that this trend is indeed under way, what are its causes? One cause, of course, may be public fear of crime and the avoidance behavior it engenders. Although survey data show little change in the prevalence of fear in recent decades, a significant national increase in fear did occur in the late 1960s (Warr 1995a). Even a constant crude prevalence

rate of fear can produce changes in behavior if those changes stem from cumulative exposure to fear. If this process is in fact under way, its scale and depth are sobering: A "free" society increasingly retreats to its dwellings in a form of asylum from an ostensibly dangerous world.

The asylum argument touches on a major longstanding controversy concerning fear of crime. Is fear ultimately a disintegrative force in a society? Does it disrupt normal social intercourse, making citizens afraid to greet or talk to one another, and undermine the civility and trust that makes civic life possible? When substantial portions of the American public are in fact afraid to leave their house at night, when they are afraid to travel on foot or traverse certain sections of their city, it is difficult to deny the power of fear to tear the social fabric asunder. Nowhere is that more evident now than in Mexico City, where fear of crime, by disrupting transportation, recreation, and commerce, has threatened the city's ability to function as a coherent system. It has also apparently undermined the legitimacy of political leaders and law enforcement itself (Newshour with Jim Lehrer 1999).

What is so often overlooked in discussions of fear, however, is the apparent ability of fear to create or increase social cohesion. As the sociologist Emile Durkheim noted long ago, crime integrates communities by drawing them together in the face of danger. Today, many millions of Americans participate in community crime watch programs, cooperative police/community associations, "bring back the night" marches and rallies, and other forms of communal protection. Whether such integrative forces are capable of counteracting the disintegrative effect of fear is difficult to say. But even if such activities compensate for a decline in face-to-face interaction in everyday life, they may do little to repair what seems to be a loss of confidence in social and political institutions. If the first priority of government is to protect its own citizens, widespread public fear of crime can only be construed as a failure of government to meet that responsibility.

Conclusion

Fear is a natural and commonplace emotion. Under many circumstances, it is a beneficial, even life-saving emotion. Under the wrong circumstances, it is an emotion that can unnecessarily constrain behavior, restrict freedom and personal opportunity, and threaten the foundation of communities.

What differentiates fear of crime from some other hazards of life is that it often rests on highly uncertain information about risk. Most citizens have little scientific foundation for their beliefs about crime. In daily life, they are constantly

confronted with information about crime from sources that may not appreciate nor care about the (in)accuracy of that information and that may use crime to entertain, sell, advertise, exploit, or win votes. In the end, most citizens are left to reason as best they can about the risks of crime. Because the consequences of victimization can be catastrophic for themselves and those they love, many are likely to err on the side of caution, worrying about and guarding against crime more than is necessary or defensible.

Given the ubiquity of messages about crime in our society and the costs of inaccurate information, it is incumbent on criminal justice officials to provide the public with reliable information about crime, including information about the risk of victimization for different criminal offenses, the sources and likelihood of error in those estimates, the nature of victimization events (including the risk of injury associated with those events), and, where known, the personal, social, and temporal/spatial characteristics that increase or reduce risk. Without information of this kind, citizens will remain uninformed about the risks of crime. In that condition they will indeed become victims, if only to those for whom crime and fear of crime are merely tools to entertain, titillate, or sell.

Crime, after all, is not like some virulent new disease whose risks and epidemiology are poorly understood. The risks associated with many criminal offenses are understood with a degree of certitude that would startle many casual observers, and such information was developed largely at public expense. The problem today is not the absence of knowledge itself, but rather the failure of criminologists and public officials to demystify crime for the general public and to present a reasoned and understandable version of the facts of crime. The gap that remains between the state of knowledge and public awareness is not merely unfortunate, it is dangerous.

References

- Baumer, Terry L. 1978. Research on fear of crime in the United States. *Victimology* 3 (3-4): 254-264.
- Biderman, A., L. Johnson, J. McIntyre, and A. Weir. 1967. *Report on a pilot study in the District of Columbia on victimization and attitudes toward law enforcement*. President's Commission on Law Enforcement and Administration of Justice, Field Surveys I. Washington, D.C.: U.S. Government Printing Office.
- Bordley, Robert F. 1982. Public perceptions of crime: A derivation of Warr's power function from the Bayesian odds relations. *Social Forces* 61:134-143.
- Bostrom, Ann, Cynthia J. Atman, Baruch Fischhoff, and M. Granger Morgan. 1994. Evaluating risk communications: Completing and correcting mental models of hazardous processes, part II. *Risk Analysis* 14:789-798.

- Clemente, Frank, and Michael Kleiman. 1977. Fear of crime in the United States: A multivariate analysis. *Social Forces* 56:519–531.
- Chiricos, Ted, Michael Hogan, and Marc Gertz. 1997. Racial composition of neighborhood and fear of crime. *Criminology* 35:107–128.
- Conklin, John. 1975. *The impact of crime*. New York: Macmillan.
- . 1971. Dimensions of community response to the crime problem. *Social Problems* 18:373–385.
- Combs, Barbara, and Paul Slovic. 1979. Newspaper coverage of causes of death. *Journalism Quarterly* 56:837–843.
- Covello, Vincent T., Detlof von Winterfeldt, and Paul Slovic. 1987. Communicating scientific information about health and environmental risks: Problems and opportunities from a social and behavioral perspective. In *Risk communication*, edited by Clarence J. Davies, Vincent T. Covello, and Frederick W. Allen. Washington, D.C.: Conservative Foundation.
- DuBow, Frederick, Edward McCabe, and Gail Kaplan. 1979. *Reactions to crime: A critical review of the literature*. Washington, D.C.: U.S. Government Printing Office.
- Erickson, Maynard L., and Jack P. Gibbs. 1979. Community tolerance and measures of delinquency. *Journal of Research in Crime and Delinquency* 17:55–79.
- Ericson, R.V., P.M. Baranek, and J.B.L. Chan. 1987. *Visualizing deviance: A study of news organizations*. Toronto: University of Toronto Press.
- Ferraro, Kenneth F. 1996. Women's fear of victimization: Shadow of sexual assault? *Social Forces* 75:667–690.
- . 1995. *Fear of crime: Interpreting victimization risk*. Albany: State University of New York Press.
- Ferraro, Kenneth F., and Randy LaGrange. 1992. Are older people most afraid of crime? Reconsidering age differences in fear of victimization. *Journal of Gerontology: Social Sciences* 47:S233–S244.
- . 1988. Are older people afraid of crime? *Journal of Aging Studies* 2:277–287.
- . 1987. The measurement of fear of crime. *Sociological Inquiry* 57:70–101.
- Figgie, H.E. 1980. *The Figgie report on fear of crime, Part 1: The general public*. Willoughby, Ohio: ATO.
- Fischhoff, Baruch. 1994. What forecasts (seem to) mean. *International Journal of Forecasting* 10:387–403.

———. 1989. Risk: A guide to controversy. Appendix C of *Improving risk communication*, National Research Council, Committee on Risk Perception and Communication. Washington, D.C.: National Academy Press.

Fischhoff, Baruch, Ann Bostrom, and Marilyn Jacobs Quadrel. 1997. Risk perception and communication. In *The methods of public health*, edited by Roger Detels, Walter W. Holland, James McEwen, and Gilbert S. Omenn. 3d ed. Vol. 2 of the *Oxford textbook of public health*. New York: Oxford University Press.

Fischhoff, Baruch, Sarah Lichtenstein, Paul Slovic, Stephen L. Derby, and Ralph L. Keeney. 1981. *Acceptable risk*. Cambridge: Cambridge University Press.

Fisher, B., and J. Nasar. 1995. Fear spots in relation to microlevel physical cues: Exploring the overlooked. *Journal of Research in Crime and Delinquency* 32:214–239.

Fishman, M. 1981. Police news: Constructing an image of crime. *Urban Life* 9:371–394.

———. 1978. Crime waves as ideology. *Social Problems* 25:531–543.

Furstenburg, Frank F., Jr. 1971. Public reaction to crime in the streets. *American Scholar* 40:601–610.

Gallup, George. 1985. *The Gallup report, report no. 239*. Princeton, New Jersey: Gallup Poll.

———. 1983. *The Gallup report, report no. 210*. Princeton, New Jersey: Gallup Poll.

Garofalo, James. 1977. *Public opinion about crime: The attitudes of victims and nonvictims in selected cities*. NCJ 41336. Washington, D.C.: U.S. Department of Justice, National Criminal Justice Information and Statistics Service.

Godbey, G., A. Patterson, and L. Brown. 1979. *The relationship of crime and fear of crime among the aged to leisure behavior and use of public leisure services*. Washington, D.C.: Andrus Foundation.

Gordon, M., and Linda Heath. 1981. The news business, crime, and fear. In *Reactions to crime*, edited by Dan A. Lewis. Beverly Hills: Sage Publications.

Graber, Doris A. 1980. *Crime news and the public*. New York: Praeger.

Haghighi, Bahram, and Jon Sorensen. 1996. America's fear of crime. In *Americans view crime and justice: A national public opinion survey*, edited by Timothy J. Flanagan and Dennis R. Longmire. Thousand Oaks, California: Sage Publications.

Heath, Linda. 1984. Impact of newspaper crime reports on fear of crime: A multimethodological investigation. *Journal of Personality and Social Psychology* 47:263–276.

- Hindelang, Michael J., Michael R. Gottfredson, and James Garofalo. 1978. *Victims of personal crime: An empirical foundation for a theory of personal victimization*. Cambridge, Massachusetts: Ballinger.
- Jaycox, Victoria. 1978. The elderly's fear of crime: Rational or irrational? *Victimology* 3:329–334.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kenny, Anthony. 1963. *Action, emotion, and will*. London: Routledge & Kegan Paul.
- LaGrange, Randy L., and Kenneth F. Ferraro. 1989. Assessing age and gender differences in perceived risk and fear of crime. *Criminology* 27:697–719.
- . 1987. The elderly's fear of crime: A critical examination of the research. *Research on Aging* 9:372–391.
- LaGrange, Randy L., Kenneth F. Ferraro, and Michael Supancic. 1992. Perceived risk and fear of crime: Role of social and physical incivilities. *Journal of Research in Crime and Delinquency* 29:311–334.
- Lee, Gary R. 1982. Sex differences in fear of crime among older people. *Research on Aging* 4:284–298.
- Lewis, Dan A., and Michael G. Maxfield. 1980. Fear in the neighborhoods: An investigation of the impact of crime. *Journal of Research in Crime and Delinquency* 17:160–189.
- Lichtenstein, Sarah, Paul Slovic, Baruch Fischhoff, Mark Layman, and Barbara Combs. 1978. Judged frequency of lethal events. *Journal of Experimental Psychology* 4:551–578.
- Liska, Allen E., Andrew Sanchirico, and Mark D. Reed. 1988. Fear of crime and constrained behavior: Specifying and estimating a reciprocal effects model. *Social Forces* 66:827–837.
- Lundgren, Regina G. 1994. *Risk communication: A handbook for communicating environmental, safety, and health risks*. Columbus, Ohio: Battelle Press.
- Mayes, Andrew. 1979. The physiology of fear and anxiety. In *Fear in animals and man*, edited by W. Sluckin. New York: Van Nostrand Reinhold.
- Miethe, Terance D., and Richard McCorkle. 1998. *Crime profiles: The anatomy of dangerous persons, places, and situations*. Los Angeles: Roxbury Publishing Company.
- National Research Council, Committee on Risk Perception and Communication. 1989. *Improving risk communication*. Washington, D.C.: National Academy Press.

Newshour with Jim Lehrer. 1999. Public Broadcasting Service, 6 April.

Ortega, Suzanne L., and Jessie L. Myles. 1987. Race and gender effects on the fear of crime: An interactive model with age. *Criminology* 25:133–152.

Plous, Scott. 1993. *The psychology of judgment and decision making*. Philadelphia: Temple University Press.

President's Commission on Law Enforcement and Administration of Justice. 1967. *The challenge of crime in a free society*. Washington, D.C.: U.S. Government Printing Office.

Pyle, G.F. 1980. Systematic sociospatial variation in perceptions of crime location and severity. In *Crime: A spatial perspective*, edited by D.E. Georges-Abeyie and K.D. Harris. New York: Columbia University Press.

Reiss, Albert J., Jr. 1967. *Studies in crime and law enforcement in major metropolitan areas*. President's Commission on Law Enforcement and Administration of Justice, Field Surveys III, part 1. Washington, D.C.: U.S. Government Printing Office.

Roshier, B. 1973. The selection of crime news by the press. In *The manufacture of news*, edited by S. Cohen and J. Young. Beverly Hills: Sage Publications.

Rountree, Pamela W., and Kenneth C. Land. 1996. Perceived risk versus fear of crime: Empirical evidence of conceptually distinct reactions in survey data. *Social Forces* 74:1353–1376.

Russell, P.A. 1979. Fear-evoking stimuli. In *Fear in animals and man*, edited by W. Sluckin. New York: Van Nostrand Reinhold.

Sheley, Joseph S., and C.D. Ashkins. 1981. Crime, crime news, and crime views. *Public Opinion Quarterly* 45:492–506.

Sherizen, S. 1978. Social creation of crime news: All the news fitted to print. In *Deviance and mass media*, edited by C. Winick. Beverly Hills: Sage Publications.

Short, James F. 1984. Toward the social transformation of risk analysis. *American Sociological Review* 49:711–725.

Skogan, Wesley G. 1990. *Disorder and decline: Crime and the spiral of decay in American neighborhoods*. New York: Free Press.

———. 1977. Public policy and fear of crime in large American cities. In *Public law and public policy*, edited by J.A. Gardiner. New York: Praeger.

Skogan, Wesley G., and Michael G. Maxfield. 1981. *Coping with crime: Individual and neighborhood reactions*. Beverly Hills: Sage Publications.

Slovic, Paul. 1987. Perception of risk. *Science* 236:280–285.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1987. Behavioral decision theory perspectives on protective behavior. In *Taking care: Understanding and encouraging self-protective behavior*, edited by Neil D. Weinstein. Cambridge: Cambridge University Press.

———. 1982. Facts versus fears: Understanding perceived risk. In *Judgment under uncertainty; Heuristics and biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky. Cambridge: Cambridge University Press.

———. 1980. Facts and fears: Understanding perceived risk. In *Societal risk assessment: How safe is safe enough?*, edited by Richard C. Schwing and Walter A. Albers, Jr. New York: Plenum Press.

———. 1979. Rating the risks. *Environment* 21:14–20, 36–39.

Sluckin, W. 1979. *Fear in animals and man*. New York: Van Nostrand Reinhold.

Smith, C.J., and G.E. Patterson. 1980. Cognitive mapping and the subjective geography of crime. In *Crime: A spatial perspective*, edited by D.E. Georges-Abeyie and K.D. Harris. New York: Columbia University Press.

Stafford, Mark C., and Omer R. Galle. 1984. Victimization rates, exposure to risk, and fear of crime. *Criminology* 22:173–185.

Taub, Richard P., D. Garth Taylor, and Jan D. Dunham. 1984. *Paths of neighborhood change: Race and crime in urban America*. Chicago: University of Chicago Press.

Thomson, Robert. 1979. The concept of fear. In *Fear in animals and man*, edited by W. Sluckin. New York: Van Nostrand Reinhold.

Tversky, Amos, and Daniel Kahneman. 1982. Availability: A heuristic for judging frequency and probability. In *Judgment under uncertainty; Heuristics and biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky. Cambridge: Cambridge University Press.

U.S. Department of Justice, Federal Bureau of Investigation. 1997. *Crime in the United States, 1996*. Uniform Crime Reports. Washington, D.C.

Warr, Mark. 1995a. Poll trends: Public opinion on crime and punishment. *Public Opinion Quarterly* 59:296–310.

———. 1995b. Public perceptions of crime and punishment. In *Criminology: A contemporary handbook*, edited by Joseph F. Sheley. 2d ed. Belmont, California: Wadsworth Publishing Company.

———. 1994. Public perceptions and reactions to violent offending and victimization. In *Consequences and control*, edited by Albert J. Reiss, Jr., and Jeffrey A. Roth. Vol. 4 of *Understanding and preventing violence*. Washington, D.C.: National Academy Press.

———. 1992. Altruistic fear of victimization in households. *Social Science Quarterly* 73:723–736.

———. 1990. Dangerous situations: Social context and fear of victimization. *Social Forces* 68:891–907.

———. 1989. What is the perceived seriousness of crimes? *Criminology* 27:795–821.

———. 1988. The hierarchy of fear: A comparison of two cities. Unpublished manuscript.

———. 1987. Fear of victimization and sensitivity to risk. *Journal of Quantitative Criminology* 3:29–46.

———. 1985. Fear of rape among urban women. *Social Problems* 32:238–250.

———. 1984. Fear of victimization: Why are women and the elderly more afraid? *Social Science Quarterly* 65:681–702.

———. 1982. The accuracy of public beliefs about crime: Further evidence. *Criminology* 20:185–204.

———. 1981. Which norms of justice? A commentary on Hamilton and Rytina. *American Journal of Sociology* 85:433–435.

———. 1980. The accuracy of public beliefs about crime. *Social Forces* 59:456–470.

Warr, Mark, and Mark C. Stafford. 1983. Fear of victimization: A look at the proximate causes. *Social Forces* 61:1033–1043.

Wilson, James Q., and George L. Kelling. 1982. Broken windows. *Atlantic Monthly* 249:29–38.

Wolfgang, Marvin E., Robert M. Figlio, Paul E. Tracy, and Simon I. Singer. 1985. *The National Survey of Crime Severity*. Washington, D.C.: U.S. Government Printing Office.

Vlek, Charles, and Pieter-Jan Stallen. 1980. Rational and personal aspects of risk. *Acta Psychologica* 45:273–300.

Measurement and Explanation in the Comparative Study of American Police Organizations

by Edward R. Maguire and Craig D. Uchida

This essay serves as a roadmap for theory and research on American police organizations. Organizational scholarship in policing has not progressed in an orderly or cumulative fashion. Some of the classic works in the study of police organizations remain well read but infrequently replicated or improved upon. Current research on police organizations is beginning to build on foundations established more than three decades ago.

The essay explores trends in the measurement and explanation of police organizations since their emergence in the early 19th century. The discussion spans the gamut of measurement and explanation, from data collection and statistical analysis methods to scholarly theory and public policy on policing. The essay demonstrates that paying careful attention to sound measurement and explanation is vital for research, theory, and practice.

Edward R. Maguire, Ph.D., is Associate Professor of Administration of Justice at George Mason University in Fairfax, Virginia. Craig D. Uchida, Ph.D., is President of 21st Century Solutions, a crime and public policy consulting firm, in Silver Spring, Maryland.

A
B
S
T
R
A
C
T

491

Measuring profit in business is fairly straightforward. Measuring results in government is not.

—Osborne and Gaebler (1992, 349)

For the corporate executive, the bottom line is profits. Most businesses, whether small, medium, or large, ask: “How much did we earn this quarter? How much did we lose? How do we increase revenues? How do we increase productivity?” An accountant, business manager, planning and management team, or even a research and development division seeks and finds answers to these questions. To explain gains and losses, external auditors annually examine activities of the business and provide reports to shareholders and potential investors.

For the law enforcement executive, the bottom line is less clear. Is the measure of success crime, clearance rates, complaints, calls for service, or use of force incidents? Is it the number of community meetings attended by officers or the quantity and quality of problem-oriented policing projects? A crime analyst, a lieutenant in charge of special projects, an aide to the chief, a patrol officer who has expertise in data management, or a computer guru in the agency may be assigned to compile information. Periodic reports may be written, or an annual report may be produced, but in few instances do these reports fully explain the activities of the agency. To be fair, selecting appropriate measures is a problem shared by administrators in most public agencies that have no single, definable bottom line (Bayley 1994; DiIulio 1993; Wilson 1989, 1993). Furthermore, the administrative problem of identifying and defining those things that are important enough to measure is complicated by the political problem of whether they should be measured and the technical problem of measuring them.

If measurement is a problem within individual police agencies, then developing measures that allow us to compare multiple police agencies is like herding cats. It means developing a reasonable and useful standardized measurement system and encouraging local police agencies to comply with it. Despite the many difficulties inherent in developing comparison measures, comparing police agencies remains a popular idea for a number of audiences, including police administrators and planners, reformers, criminologists and other scholars, journalists, and policy-makers with a variety of agendas.

This essay examines how various features of police organizations have been measured and explained over time and place. The central focus is not on measures developed within individual agencies but on measures that have been, or might be, used to compare police organizations. Although these organizations

have much in common, they also show tremendous variation. Some are large, but many are quite small; some patrol aggressively, arresting offenders for minor public order offenses, others enforce the law with less vigor; some have tall hierarchies and formal command structures, others are less formal, with only a handful of levels; some work closely with communities and spend time formulating customized solutions to local problems, others shun community involvement, instead providing more “traditional” police services. This variation in both what organizations *do* and what they *are* is not unique to police agencies. As Scott (1992) notes, “while organizations may possess common generic characteristics, they exhibit staggering variety—in size, in structure, and in operating processes.” This essay explores efforts to measure and explain variation in American police organizations: variation in what they are and what they do, in form and function, in structure and process, and in policy and practice.

The subject of this essay is police *organizations*. The study of police agencies as organizations is growing, owing its theoretical roots to the sociological and social psychological study of organizations in general.¹ This focus on police as organizations is the common thread linking each section of the essay. We do not examine other frequently studied features of policing, including police culture, police discretionary behaviors (and misbehaviors), individual officer attributes, and other important phenomena occurring at units of analysis that are larger (e.g., states or nations) or smaller (e.g., officers or workgroups) than police organizations.²

There is some ambiguity over what constitutes a police organization (Maguire et al. 1998). As Bayley (1985, 7) notes, “police come in a bewildering variety of forms . . . moreover, many agencies that are not thought of as police nonetheless possess ‘police’ powers.” To reduce the scope of our task, we focus on public police organizations in the United States whose primary purpose is to provide generalized police services, including responding to calls for service, to a distinct residential population.³ This levels the playing field, allowing us to explore variation among organizations with a common purpose.

This essay discusses broad organizational properties rather than particular policies, programs, activities, or structural features. Researchers have produced valuable research on particular features of police organizations such as pursuit policies, D.A.R.E.® programs, the use of one-officer and two-officer patrol cars, and the establishment of special units for various tasks (e.g., narcotics, child abuse, gangs). The line between general and specific organizational properties is admittedly arbitrary. Nevertheless, the goal of the essay is to draw together a diverse body of scholarship on American police organizations. Research on specific (and sometimes esoteric) organizational properties makes it more difficult to consolidate this large body of theory and research. Therefore, we do not

discuss the prevalence of specialized bias-crime units, but we discuss specialization in general; we do not discuss the implementation of new technologies for processing offenders, but we examine innovation; we do not discuss drunk driving enforcement or use of deadly force, but we discuss aggressive patrol strategies and styles of policing.

Even after narrowing the focus in this way, there remain considerable variations among police organizations over time and place.⁴ A substantial body of theory and research has been developed to measure and explain these variations. As one way to organize this body of scholarship on police organizations, we draw an important distinction between what they *do* and what they *are*. These categories sometimes overlap in practice, but there is some precedent in the development of organization theory for treating them separately.

What Police Organizations Do

Processes are not something that we invented in order to write about them. Every company on Earth consists of processes. Processes are what companies do.

—Hammer and Champy (1993, 117)

Like corporate America, police organizations do many things. Most people are unaccustomed to thinking of *organizations* as doing things. After all, organizations are composed of people, and it is the people within them who think, plan, act, decide, debate, respond, cooperate, and all of the other activities and behaviors in which *people* engage. Yet, as Maguire (forthcoming) recently argued:

Organizations are greater than the sum of their parts. They expand and contract, rise and fall, and generally take on lives of their own. Organizations, like individuals and social groups, do not only act, but are acted upon as well. They are influenced, shaped and constrained by a complex interaction of political, social, economic, cultural, and institutional forces. Organizations exhibit patterned regularities, and they can (and indeed should) be studied apart from the people within them (Blau, Heyderbrand, and Stauffer 1966; Blau and Schoenherr 1971).

Recent work by King, Travis, and Langworthy (1997) takes this argument one step further, using a biological, or life course, perspective to study the birth, death, and aging processes of police agencies. Thinking about organizations as separate from the people within them—as “corporate persons”—is essential to understanding what they do (Coleman 1974).

Police organizations do many things. They make arrests, process offenders, find lost children, quell disturbances, respond to emergencies, solve problems, form relationships with the community, and perform many other activities too numerous to summarize briefly. These activities, in the aggregate, constitute the output of police organizations. Most attempts to measure and explain this output have relied on arrest, clearance, and crime statistics.

Systematic collection of data from large samples of police agencies has shown considerable variation in the quantity and quality of this output over time and place. These data are used in many ways. Arrest and clearance statistics, for instance, are frequently used as measures of a police organization's productivity.

As we will show later, the use of these kinds of performance indicators is beginning to fall out of fashion as police executives, scholars, and reformers focus on alternative measures. These data are also used as indicators of a police organization's "style."⁵

Some agencies may emphasize aggressive enforcement of panhandling ordinances, for instance, but others may tend to ignore such minor offenses. Although the concept of organizational style is intangible and difficult to measure, researchers have attempted to draw inferences about policing styles by examining arrest patterns for discretionary offenses such as drunkenness or disorderly conduct (Wilson 1968b). Although police organizations do many different things, data are systematically collected on only a few of these activities. This limits the scope of measures that can be constructed from these data.

One focus of this chapter is to examine variations in police activities, processes, performance, and style over time, across agencies, and across fully functioning subunits within agencies (which often resemble mini organizations). We will trace efforts to measure and explain what police organizations do, from the traditional focus on arrests and clearances to more recent efforts to use problem solving and community partnership strategies. Our discussion spans the gamut of analytical issues, from data collection and analysis strategies to appraisals of reliability, validity, theory, and causation.

Police organizations do many things.

They make arrests, process offenders, find lost children, quell disturbances, respond to emergencies, solve problems, form relationships with the community, and perform many other activities too numerous to summarize briefly.

What Police Organizations Are

Lord! We know what we are, but know not what we may be.

—William Shakespeare, *Hamlet* (1601) act 4, scene 5.1

What a police organization *does* is external, typically taking place outside of the organization: in the community, on the streets, in people's homes. The features that define what a police organization *is* tend to be internal: administrative arrangements, processing routines, structures, communication patterns, and overall "corporate" personalities.⁶ In short, what police organizations *do* takes place within the framework (or context) of what they *are*. The social scientific study of what police organizations are has a much shorter history than the study of what they do. This history parallels a similar split in the study of organizations in general. Although output and performance have always been a primary focus of organizational research, it was not until the late 1950s that "researchers began to conceive of organizations as more than just rationally derived mechanisms for the production of goods and services, but as entities worthy of understanding for what they *are* in addition to what they *produce*" (Maguire forthcoming, 12; emphasis in original).

Researchers began to measure variation in the internal features of police organizations using systematic surveys in the late 1920s. Attempts to explain this variation came later, with theoretical explanations appearing in the 1960s and empirical studies beginning in the mid-1970s. Much of this research examines why we have the police organizations we have and seeks to isolate local characteristics (such as regional, historical, demographic, governmental, or cultural factors) that produce organizational variations from one jurisdiction to another.

The essay is organized into three parts: Measurement, Explanation, and Future Prospects. In part 1, we trace the evolution of measurement in the comparative study of American police organizations. This discussion spans the past 150 years, from crude early attempts to count police output to recent methodologically informed efforts to "measure what matters" (Langworthy 1999). In part 2, we trace the evolution of attempts to explain the variation in police organizations over time and space. This body of work was influenced heavily by research and theory in the sociology of organizations. In part 3, we review some general themes and discuss concrete ways to better measure and explain police organizations.

Measurement

But the main thing is, does it hold good measure?

—Robert Browning, English poet, 1850

Measurement and explanation are the twin pillars of social science. Measurement in the social sciences has the same purpose as in the physical sciences: to quantify, count, or assign meaningful scores to variations in some phenomenon, using valid and reliable methods. Clocks, thermometers, barometers, odometers, speedometers, and bathroom scales surround us, becoming so ingrained in our daily lives that we never stop to think about what they share in common: They are all measuring instruments.

Frequently, the traits measured in the social sciences are more ambiguous than properties like time, humidity, speed, or weight. For instance, psychologists have developed scores of “instruments” for measuring aggression, intelligence, mental illness, and humor. In the business world, researchers measure customer satisfaction, commitment, value systems, and other phenomena that are difficult to quantify. Testing and refining these measuring instruments is a labor intensive process that takes years, as evidenced by the debate over the development of standardized tests to measure individual intelligence and aptitude.

The difficulties in measuring ambiguous social properties apply equally to the police. Researchers, government officials, and police executives have debated the best ways to measure police performance for more than 100 years. This debate remains largely unresolved, though if the various parties agree about anything, it is that existing measurement systems are insufficient and alternatives are difficult to come by. According to Carl Klockars (1999, 198):

[M]easuring in the social sciences is a very sad affair. It is an activity so fraught with mind- and soul-wrenching difficulties that only grossly ignorant beginning students and the least capable or least virtuous of social scientists engage in it with good humor. A warning is in order to any police practitioner who is approached by a quantitative criminologist with a smile on his or her face: Listen very, very carefully, keeping one hand on your wallet and the other on your gun.

The technologies used to record data in police organizations have evolved substantially, from archaic leather log books in police stations to high-speed personal computers that store and process vast amounts of data. The measurements constructed from these raw data have evolved as well, ranging from simple tallies of police activities and outputs to composite measures calculated using advanced statistical methods. In fact, recent research on police organizations

has begun to borrow some of the sophisticated statistical measurement techniques used by educational researchers and psychologists. In this section, we begin by examining the evolution in measures of what the police do, from just after the birth of police organizations to recent findings from a national forum of leading police experts. We then trace various efforts to measure what police organizations are.

Measuring what police organizations do

One of the first areas of measurement for police organizations focused on what they do: their outputs, their performance, and their effectiveness. Since the inception of municipal police departments in the early to mid-1800s, agencies have gauged their performance through arrest rates and reported crimes. In some instances, they recorded the number of lost children returned home, drunks taken home, and lodgers taken into the station house (Monkkonen 1981). As police became more “professionalized,” the crimefighting image began to predominate, and agencies focused on collecting and publishing crime data. With the “discovery” of discretion in the 1960s, researchers and police administrators realized that order maintenance and other noncrime activities were also important to measure. The community policing movement of the 1980s and 1990s has led to a shift in attitudes about what *should* be measured, yet there is little agreement among researchers or police about how to do so. This section examines the types of data that have been collected to measure what police organizations do, including some of the problems those measures present and the ongoing debate about what should be measured.

Most urban police departments were created in the mid- to late 19th century as a result of growing concern over riots and disorder. Yet police neither measured the number of riots they broke up nor kept track of disorder. Instead, to measure the performance of uniformed police, city councils, State assemblies, and police administrators urged the collection of crime statistics (primarily arrests for various crimes) (Monkkonen 1981). As early as 1851, the New York City Police Department collected data on arrests and reported these in their semi-annual reports (Miller 1977). Other large agencies, including Washington, D.C., Boston, Philadelphia, and Chicago, followed suit (see Count-van Manen 1977; Ferdinand 1967, 1972; Lane 1980; Surette 1984; Walker 1977). Police historians have only been able to trace the growth and development of police organizations by looking at annual reports that included arrests, number of officers, and other activities enumerated by the police. These data provide a glimpse of what police organizations did when they were first established (Count-van Manen 1977; Ferdinand 1967; Miller 1977; Monkkonen 1981).

According to Monkkonen (1992), police agencies in the 19th century also engaged in a number of social service functions. They took in “tramps, returned lost children by the thousands, shot stray dogs, enforced sanitation laws, inspected boilers, took annual censuses, and performed other small tasks” (p. 554). Toward the end of the century, they began to drop these functions and focus on crime control; however, some agencies kept careful records of these activities through the early 1900s.

Also near the end of the 19th century, problems in policing, particularly with corruption and inefficiency, began to surface. To resolve these problems, police reformers stressed managerial efficiency and professionalism, catchwords that, in 1894, heralded the development of the forerunner of the International Association of Chiefs of Police (IACP), and independent municipal groups began to collect law enforcement and crime statistics in the early 1900s. For instance, the Bureau of Municipal Research in New York City, organized in 1906, gathered statistics on police administration and operations as part of its investigation of corruption. The bureau not only collected data but also made recommendations to the departments regarding administrative practices. Although reforms were not readily made within the departments as a result of the bureau’s work, these reports established the model for the police survey that, by the 1920s and 1930s, eventually became standard in police administration.

By the late 1920s, IACP and the Federal Bureau of Investigation (FBI) sought to professionalize the police by stressing crimefighting rather than noncrime services. This marked a profound shift in how police characterized their work and led to the collection of national data through the Uniform Crime Reports (UCR) (Uchida 1997). As police systems began to move toward professionalism, August Vollmer (chief of the Berkeley, California, Police Department from 1905 to 1932) and IACP urged departments to maintain records related to crime and operations. The IACP Committee on Uniform Crime Records was formed in 1927 to create a system for collecting uniform police data. Standardized definitions were formulated to overcome regional differences in the definitions of criminal offenses. Over a 2-year period, the committee examined variations among State codes and evaluated the recordkeeping practices of police agencies. By 1929, the committee finalized a plan for crime reporting that became the basis for UCR. The committee chose to obtain data on offenses that come to the attention of law enforcement agencies because they were more readily available than other reportable crime data. Seven offenses, because of their seriousness, frequency of occurrence, and likelihood of being reported to law enforcement, were initially selected to serve as an index for evaluating fluctuations in the volume of crime. These crimes, known as the Crime Index offenses, were murder and non-negligent manslaughter, forcible rape, robbery,

aggravated assault, burglary, larceny/theft, and motor vehicle theft. By congressional mandate, arson was added as the eighth Index offense in 1978.

In 1930, Congress enacted legislation authorizing the U.S. Attorney General to gather crime data. The Attorney General designated the FBI to serve as the national clearinghouse for data collected by the program. As Wellford (1982) indicates, there were two purposes for uniform crime reporting. First, police leaders saw the value of better data for the improvement of management and operation strategies. He cites Vollmer:

How can we ever know the extent or the nature and distribution of crime or discuss the problem intelligently at our conferences or conduct a successful campaign to prevent crime until police statistics are accurately, adequately, and uniformly compiled? (Vollmer 1927).

Second was the desire for data that could combat the perception of “crime waves” created by the media. Police managers were concerned that such crime waves were seen as a reflection of their departments’ inadequacies. They believed that uniform crime data would dispel the notion of crime waves and, therefore, demonstrate the value of police efforts to control crime. The tension between developing a system that would represent a true measure of crimes known to police and developing one that would prove crime was not increasing has been present in UCR since the beginning.

Since their inception in 1930, the Uniform Crime Reports have evolved from a relatively small data collection effort into a large-scale national effort. In its first year of operation, the program collected data from 400 cities in 43 States, representing about 20 million people. In 1998, UCR collected data from more than 17,000 cities in all 50 States, representing about 260 million people. UCR remains a voluntary reporting program, with city, county, and State law enforcement agencies reporting monthly to the FBI the number of part 1 offenses and part 1 and part 2 arrests that have occurred within their jurisdictions. In addition to monthly tallies of offenses and arrests, additional data are captured on particular offenses,⁷ and data on age, sex, race, and ethnicity are collected for arrests. In 44 States, an agency reviews, edits, and compiles the data for statewide UCR reporting and then forwards the data to the FBI for inclusion in the national report.

Although crime data have been systematically collected for nearly 70 years, there has been controversy over whether and how these data actually measure crime and/or police performance. The increase in crime in the early 1960s raised questions about the effectiveness of police. In 1999, with the decrease in crime spanning 7 successive years,⁸ police are praised by the President, the

Attorney General, and local politicians for their effectiveness. Yet many critics question the use of crime data as measures of police performance (Beattie 1960; Kitsuse and Cicourel 1963; Wolfgang 1963; Robison 1966; Seidman and Couzens 1974; Decker 1977; Maltz 1977; Walker 1992). Specifically, critics note that the ratio of recorded to unrecorded crimes may fluctuate by offense, jurisdiction, and year, regardless of the true incidence of criminal events. They also cite the impact of political and organizational pressures on the crime rate, suggesting that fluctuations in the recorded rates may occur independent of true changes. Although others recognize these problems with crime statistics, they indicate that UCR data have some potential (Nanus and Perry 1973; Berk 1974; Wellford 1974). They recommend the exercise of caution when using these data and suggest that crime data be used in combination with other measures. Decker (1981), for example, uses official crime data from UCR, arrest data (also from the FBI, but not widely disseminated), and victimization data for 25 cities. He concludes that "such a combination of measures could lead to a standardization of measures across jurisdictions" and that the three measures "expand the ability of analysts to determine the success of police efforts in dealing with crime."

One partial solution to the problems of UCR is the National Incident-Based Reporting System (NIBRS). NIBRS differs from UCR in a number of ways. The most notable difference is that in NIBRS, individual records related to a distinct crime incident and its associated arrest are recorded. NIBRS includes 52 data elements on 22 primary crime categories. It distinguishes between attempted and completed crimes, collects more detailed information on the victim/offender relationship and the circumstances of an offense, and eliminates the "hierarchy rule."⁹ By collecting detailed data in an incident-based format, "practitioners and researchers will be able to undertake sophisticated, multivariate analyses of crime within a jurisdiction and link demographic and economic data" (Roberts and Jacobs 1994). This system, discussed and debated for almost 20 years, has yet to catch on but "holds great promise for tactical crime analysis, examining the effectiveness of specific law enforcement techniques, determining the allocation of scarce resources, and identifying and comparing trends in criminal justice activities exhibited in different counties or departments" (Coyle, Schaaf, and Coldren 1991).

Discretion came into play not only in decisions to arrest but also in every situation in which the police make choices about doing something.

The discovery of discretion

In the 1960s, the “discovery” of police discretion by legal commentators (Goldstein 1960; LaFave 1965; Davis 1969) and researchers (Wilson 1968b; Reiss 1971; Skolnick 1966) changed how we look at what police do and how we measure their behavior, both organizationally and individually. These writers emphasized the role that discretionary decisionmaking plays in the day-to-day work of police officers. The legal commentators examined discretion in terms of the criminal process, noting that police usurp the power of magistrates by making decisions in the field that are reserved for magistrates. Commentators believed that rulemaking and policy formulation were necessary to control police decisions to invoke the law (Goldstein 1960). Researchers and police administrators took a different approach. They saw that police officers and organizations were involved in complex work and emphasized the seriousness

Overall, the discovery of discretion broadened our understanding of what police organizations do. It meant that the activities of police organizations had to be measured with data that went beyond arrests. This includes many discretionary choices that officers make: to use force; to make motor vehicle stops; to cite, warn, or scold; to be rude or polite; to search; to conduct a field interrogation; and so forth.

and importance of decisionmaking in every part of policing. Herman Goldstein (1963, 1977), for example, emphasized that discretion came into play not only in decisions to arrest but also in every situation in which the police make choices about doing something.

Although these writers demonstrated the use of discretion among *individual* officers, James Q. Wilson (1968b) showed how aggregate patterns of discretion within an agency can reflect an overall *organizational* style. Wilson proposed three major organizational styles of policing: watchman, legalistic, and service. He linked these styles to the policies of the police administrator, local politics, and community structure. But, as we demonstrate later, measuring these styles of policing is not easy. Wilson makes use of arrest data and calls for service for each of his eight agencies, but uses them only to augment his interviews and observations.

Overall, the discovery of discretion broadened our understanding of what police organizations do. It meant that the activities of police organizations had to be measured with data that went beyond arrests. This includes many discretionary choices that officers make: to use force; to make motor vehicle stops; to cite, warn, or scold; to be rude or polite; to search; to conduct a field interrogation; and so forth. Together, these choices shape the image, style, and character of

a police organization. Some agencies make a lot of arrests, use force frequently, and stop motor vehicles regularly; others focus on providing community service and upholding community standards. Although discretion is an individual attribute, it has organizational implications. More importantly, these organizational properties can and should be measured. Yet, for the most part, they are not.

Community policing

Traditional performance measures include reported crime rates, overall arrests, clearance rates, and calls for service and/or response times (Alpert and Moore 1993). Traditional measures, however, emphasize only the crime control aspect of policing and do not adequately address the many other police agency activities. This problem becomes increasingly apparent when we examine the recent interest in, and adoption of, community policing. Several data sources now exist for researchers to begin evaluating how to measure organizational involvement in community policing activities.

Over the past 7 years, there have been nearly a dozen national surveys of community policing in the United States. In 1993, with a grant from the National Institute of Justice, the Police Foundation surveyed more than 1,600 police and sheriffs' departments about their experiences with community policing strategies (Annan 1994; Wycoff 1994). The study was replicated in 1997 (Macro International 1997). National surveys also were conducted by several others, including the National Center for Community Policing and the FBI in 1993 (Trojanowicz 1994); Washington State University in 1993 and 1996 (Maguire, Zhao, and Lovrich 1999); The Urban Institute in 1995 and 1996 (Roth and Johnson 1997); the Bureau of Justice Statistics (BJS) as an adjunct to its 1997 Law Enforcement Management and Administrative Statistics (LEMAS) survey (Reaves and Goldberg 1999); and the University of Nebraska in 1998 (Maguire and Zhao 1999).

Other national data on community policing have been collected by the Justice Department's Office of Community Oriented Policing Services (COPS) since 1994. COPS has collected community policing data from more than 10,000 separate police agencies. For instance, data on 31 community policing activities were collected from a 1994 grant program called COPS FAST (Funding Accelerated for Small Towns) that focused on agencies serving populations of 50,000 or fewer. Nearly 6,000 agencies responded, serving populations ranging from 106 to 49,949. After COPS FAST was completed in 1995, a more comprehensive reporting format was instituted. The Community Policing Information Worksheets collect data on 49 separate community policing activities. Other COPS data collections are based on initial and annual reports completed by grantees. Although these datasets are based on nonrandom samples of agencies

and have other problems, they are still useful for some purposes (see Maguire and Mastroski forthcoming).¹⁰

By the time this chapter appears in print, the newest national survey of community policing will have been completed. Mastroski (1998b) and his colleagues surveyed more than 2,500 police agencies across the United States that employ 10 or more full-time sworn officers. This study improves on previous research in at least five ways. First, the mail survey instrument was designed to elicit the quantity or scope of community policing activities (known by social scientists as the “dosage”), rather than their mere presence or absence. This will result in more precise measurements of the nature and volume of community policing. Second, since community policing likely varies both within and among police agencies, separate surveys were distributed to district commanders in the largest agencies. Third, researchers are examining the measurement structure of the data using methods that are similar to those used in educational and intelligence testing. The purpose of these analyses is twofold: to measure intra-agency as well as interagency variation in community policing, and to ensure that the measures constructed from the data are both reliable and valid (Maguire and Uchida 1998). Fourth, because evidence shows that some respondents to community policing surveys may exaggerate their actual involvement (Maguire and Mastroski forthcoming), respondents were promised full confidentiality. Finally, researchers visited some of the responding agencies, using a variety of semistructured qualitative research methods to record their observations while on site. This served as an additional check for reliability and validity, and produced additional insights about community policing as it is practiced (or not) in American cities and counties.

With these data sources, it appears that measuring community policing at the agency level is possible and plausible. We have explored methods to do so using the Police Foundation data from 1993. Further efforts to measure community policing using newer and better data sources are ongoing (Maguire and Uchida 1998; Parks 1999).

Measuring what police organizations do: Contemporary issues

We have described the kinds of data that have been collected to compare police organizations and spent time discussing how to turn those data into measures. However, there is another whole spectrum of measurement issues that we have not explored: the debate over the kinds of data that should be collected and how they should be used. Policymakers, reformers, scholars, and police administrators now routinely discuss measurement as a method for improving police performance. Some of these ideas involve collecting new kinds of data, others

involve using some of the traditional measures in new ways. We begin with the latter.

One of the most celebrated reforms in policing during the 1990s has been COMPSTAT (computer comparison statistics). First implemented in New York under Commissioner William Bratton, COMPSTAT is now being used by agencies around the United States and in other nations. As implemented in New York, COMPSTAT generates crime statistics for each precinct and uses them to hold precinct commanders accountable for addressing crime in their areas (Bratton 1999; Bratton with Knobler 1998). Blending technology and accountability in this fashion is popular in the private sector, but it is rare in the public sector. According to Bratton, he had “become a staunch advocate of using private-sector business practices and principles for the management of the NYPD, even using the business term ‘reengineered’ rather than the public policy term ‘reinventing’ government” (Bratton with Knobler 1998, 224).

COMPSTAT has been credited with lowering crime rates in several American cities, though some scholars express doubts about these claims (Bratton 1999; Eck and Maguire forthcoming). At a minimum, Bratton mobilized the New York Police Department (NYPD) to raise arrest productivity significantly.¹¹ From an administrative perspective, COMPSTAT represents a significant change. From a measurement perspective, however, it relies on the same crime data that have been criticized as inadequate measures of police performance for the past two decades.

Others also suggest new uses for these traditional measures. In England, for example, there is now a trend toward auditing police agencies (Hough and Tilley 1998; Leigh, Mundy, and Tuffin 1999). This involves having each police agency develop measures of crime and disorder and making these, together with the strategies used to address them, available publicly. As might be expected, concerns are being raised about the extent to which such data are comparable across agencies and the extent to which they can be used as relative measures of police performance. Sherman (1998) adds a new twist to this notion, suggesting the publication of rankings of police agencies, much like *U.S. News and World Report*'s annual ranking of American colleges and universities. Rather than using simple crime rates to rank performance, Sherman suggests using “risk adjusted” crime rates that account for community differences in factors thought to be

The intensive focus on numbers of crimes and arrests may lead police agencies to lose sight of other important goals, such as equity, fairness, or a spectrum of humanistic concerns that Mastrofski (1999) calls “Policing for People.”

associated with crime. This method would pressure agencies with low relative performance to find and adopt more effective strategies.

For some, these suggestions raise the issue discussed in the introduction: Police organizations do not have a readily definable bottom line. Furthermore, the intensive focus on numbers of crimes and arrests may lead police agencies to lose sight of other important goals, such as equity, fairness, or a spectrum of humanistic concerns that Mastrofski (1999) calls "Policing for People" (also see Eck and Maguire forthcoming; Moore and Poethig 1999). As former Commissioner Bratton admits, his strategies for reducing crime in New York came with some consequences:

We defined brutality as unnecessary behavior that caused broken bones, stitches, and internal injuries. But those were not the figures that had gone up significantly. What had risen were reports of police inappropriately pushing, shoving, sometimes only touching citizens. We were taking back the streets . . . we were being more proactive, we were engaging more people, and often they didn't like it. (Bratton with Knobler 1998, 291)

According to Mastrofski, all of the measures discussed thus far ignore a fundamental element of the relationship between police and communities: the nature of police-citizen encounters. He highlights six features of these encounters that should be measured: attentiveness, reliability, responsiveness, competence, manners, and fairness. Like other variables we have discussed, these are characteristics of individual encounters and officers, but collectively, they can be used to characterize and compare police agencies over time and place. Doing so will represent a fundamental challenge for both police administrators and researchers. Measuring these features in ways that are useful for comparing organizations is a challenge for researchers, and administrators will need courage and skill to implement them.

The notion that policing has competing bottom lines is perhaps clearest today. Although the Nation continues to experience dramatic reductions in violent crime, the problems of police-community relations identified three decades ago by the President's Commission on Law Enforcement and Administration of Justice (1967) continue to exist. Concern over use of force by police officers led Congress to order the U.S. Department of Justice (DOJ) to "acquire data about the use of excessive force by law enforcement officers" and to "publish an annual summary of the data acquired." (McEwen 1996, vi). Concern over the practice of "racial profiling" by police officers making traffic stops, a long-time concern of minority communities, has led to calls for measurement on the severity of the problem. Both the U.S. House of Representatives and the Senate are considering the Traffic Stops Statistics Act of 1999, which would

require DOJ to acquire “data from law enforcement agencies regarding the characteristics of persons stopped for alleged traffic violations and the rationale for subsequent searches” (Conyers 1999, E673). Although many observers applaud recent reductions in crime, others continue to wonder whether these reductions are worth the cost in terms of police-community relations, particularly in minority communities. Measurement plays an important role in this debate.

Measuring what police organizations are

Police agencies and researchers are not only concerned about measuring what the police do, they also are interested in other descriptive attributes such as their processes, policies, and structures. Since the 1930s, descriptive information on these internal features of police organizations has been collected, for the most part, on an ad hoc basis. Administrative and management data have been collected intermittently by the International City/County Management Association (ICMA), the Fraternal Order of Police (FOP), the Kansas City Police Department, the Police Executive Research Forum (PERF), the Police Foundation, BJS, and various National Institute of Justice (NIJ) grantees. Through national surveys, these organizations have collected data from police agencies that describe who they are, including salaries, equipment, policies, procedures, personnel, practices, structures, and other information. Only in the past decade, however, has there been a systematic data collection process in place to measure the internal characteristics of police organizations. BJS, through its LEMAS data series, collects input and process data that enable researchers and police executives to better understand police agencies. This section examines the data that are available and briefly discusses research efforts that make use of these data.

The evolution of data collection

The first nationwide collection of operational and administrative data occurred in 1929 under the Bureau of Municipal Research. The bureau surveyed 78 city police agencies and 9 State police forces about police wages and salaries, clothing reimbursement, room and board, sick leave, vacation leave, pension funds, and stability of employment (Beyer and Toerring 1929). In the 1930s, ICMA began the first data collection series, which exists to this day. These data were part of ICMA’s annual *Municipal Yearbook* series, a collection of statistics on cities (and now counties) throughout the country. Initiated in 1934, the *Municipal Yearbook* reported on most facets of city management, from financial planning and form of government to law enforcement. By 1998, the *Yearbook* presented data from more than 7,000 police jurisdictions and included police personnel, salaries, and expenditures.¹²

By 1951, the Fraternal Order of Police and the Kansas City (Missouri) Police Department (KCPD) began collecting annual data from police agencies about their administrative practices. From 1951 to the mid-1980s, FOP collected and published salaries and benefits of chiefs and the rank-and-file, officer education level, and types of legal aid—information important to both administrators and unions. The KCPD collected process data from about 40 large agencies—serving populations of 300,000 to 1 million—for 22 years. Like FOP, the KCPD collected information about salaries and fringe benefits, but for a different purpose: to compare itself with similar agencies. It also asked about computer facilities, uniforms, weapons, and vehicles. By 1973, budgetary considerations forced the department to discontinue the survey, but 4 years later, it joined forces with the Police Foundation to reinstitute the questionnaire. Unlike the previous surveys, however, input data (calls for service) and output data (firearms incidents) were requested. Additional questions were asked about personnel (breakdowns by race and ethnicity), equipment, and special programs. A similar survey was distributed in 1981 with the Police Executive Research Forum.

A number of ad hoc administrative surveys have been fielded by research organizations. In the early 1970s, the Police Foundation, IACP, and the Educational Testing Service analyzed data on personnel practices from 481 respondents (Eisenberg and Kent 1973). In 1983, the National Association of Criminal Justice Planners and BJS collaborated on a survey that was sent to 53 agencies (Cuniff 1983). These data concentrated on four areas: calls for service, agency reports, investigations, and resources.

The most extensive collection of administrative data on a national level occurred through the Workshop in Political Theory and Policy Analysis at Indiana University (Ostrom, Parks, and Whitaker 1978a). Through a grant from the National Science Foundation, Ostrom and her colleagues collected and analyzed data from 1,827 police agencies within 80 Standard Metropolitan Statistical Areas (SMSAs). Five major areas were examined: service conditions, the legal structure, organizational arrangements, manpower levels, and expenditure levels. In phase 1 of the study, researchers collected data through the State capital, from local sheriffs' offices and police agencies in each SMSA, and from interviews with members of more than 600 police agencies. During phase 2, detailed observational data on police-citizen encounters and administrative statistics were collected from three metropolitan areas: Rochester, New York; St. Louis, Missouri; and Tampa-St. Petersburg, Florida.

With the data from phase 1, Ostrom and her colleagues described the services of police departments within the 80 SMSAs. At least 18 technical reports were published at Indiana University, along with their book, *Patterns of Metropolitan Policing* (Ostrom, Parks, and Whitaker 1978a). Few researchers, however, used

these data for further study. Only Robert Langworthy used the data from phase 1 extensively and in combination with the Kansas City General Administrative Survey. His dissertation as well as book and journal articles examine the causes and correlates of organizational structure in large police agencies (see Langworthy 1983, 1985a, 1986). In contrast, many more researchers have used the observational information collected during phase 2 (for a review, see Riksheim and Chermak 1993; Sherman 1980).

Law Enforcement Management and Administrative Statistics

In 1983, BJS noted that the collection of law enforcement data lagged behind data collection for jails, prisons, juvenile institutions, and the courts. To fill this gap, BJS initiated a program for the collection of these data on a national level (Uchida 1986). Through telephone interviews, a national assessment of data needs, and national conferences, Uchida and his colleagues developed the BJS Law Enforcement Management and Administrative Statistics series (Uchida 1986). In 1987, BJS launched the first wave of the new series. The survey was sent to all agencies with 135 or more sworn officers and to a random sample of smaller departments. In 1990, 1993, and 1997, the survey was expanded to include all agencies with more than 100 sworn officers and a random sample of smaller departments. Each of the resulting survey databases contains more than 500 variables on more than 3,000 law enforcement agencies.

The datasets discussed in this section represent only some of those that measure what police organizations are. We have tried to highlight those that either represent significant changes in measurement or that are most widely used. Overall, data describing the internal features of police organizations have evolved and are now of reasonably good quality. BJS publishes the findings from LEMAS surveys in paperback volumes that are useful for both practitioner and research audiences. It also releases the raw data files to data archives that are accessible through the Internet. Scholars now use LEMAS data for a variety of purposes, and their analyses are reported at academic conferences and in journals. BJS has been attentive to requests by scholars and practitioners to modify the survey, while trying to maintain the integrity of the longitudinal data series.

We close this section with a recommendation. When COPS was interested in collecting more data on community policing, it commissioned BJS in 1997 to add a community policing addendum to the existing LEMAS survey. LEMAS could be used to collect different types of data important to public policy on policing. However, it is only conducted every 3 years. BJS could follow the lead of the Census Bureau, which conducts specialized surveys in noncensus years, by conducting surveys on different topics in non-LEMAS years. Although

the topics could change depending on the current policy climate, they could include community policing, use of force, or other issues considered important at the time. These surveys could ride on the success of the existing LEMAS format.

From data collection to measurement

So far in this section, we have tracked the evolution of data systems useful for describing the functions, structures, activities, and performance of American police organizations. These data are available to police professionals, researchers, public policy analysts, and students of the police. Thanks to recent innovations in technology, these datasets are now in electronic archives such as the National Archive of Criminal Justice Data and the Inter-university Consortium for Political and Social Research. Although data sources vary in their quantity and quality, there is still much to be learned by examining them. With all of the archival data available to examine police agencies, relatively few researchers have done so.

Before moving on to our discussion of explanation, we want to carefully emphasize an important difference between data collection and measurement. For many people, data collection and measurement are synonymous. The usual method of doing organizational research on the police is to collect data from a sample of police agencies using telephone or mail surveys, compile the information into percentages and cross-tabulation tables, and publish these in a booklet containing the findings of the study. A typical conclusion from this kind of descriptive research is a statement like the following: "Eighty-two percent of police departments do X, but only 19 percent of departments do Y." These kinds of results are often useful, interesting, and informative, but they constitute the beginning, not the end, of measurement.

In doing scholarly research on the police, researchers and students are frequently interested in examining global or conceptual properties of organizations, such as the style of patrol, the degree of autonomy granted to officers, the decentralization of command, or the effectiveness and efficiency of various policing strategies. The problem for researchers is that these social properties are ambiguous and not directly measurable. To measure these kinds of ambiguous traits, researchers frequently select proxies or indicators that come as close as possible to being direct measures. For instance, most of the empirical research on police styles uses arrest rates as indicators. Yet the variation in arrest rates from community to community depends on many factors that may have little to do with policing styles. The portion of the variation that is not attributable to differences in police styles is the unreliable or error variation, and it has at least two sources: systematic error and random error. One obvious example of systematic error affecting measures of police style constructed from

arrest rates is, quite simply, differences in the number of offenses across communities. If the number of arrests varies in proportion to the number of offenses, then arrest rates may not measure police style at all. It is more likely that they are a partial measure of both the volume of offenses and police style. Sources of random error might include differences in arrest policies and recording practices. Though arrest rates are reasonable (and available) proxies for police style, they are imperfect measures. Despite this, researchers rarely acknowledge the error in their measures or attempt to deal with it.

Recently, social science researchers have begun to use a variety of techniques for transforming individual variables into useful composite measures of ambiguous theoretical concepts (Hayduk 1987; Schumacker and Lomax 1996). Just as a student's responses on a standardized test are combined to form an overall score that measures verbal and mathematical skills, items on a questionnaire might be used to form a similar type of composite score. Policing scholars are now beginning to adopt some of these advanced measurement techniques from other disciplines, especially psychology and education, as a means for measuring police organizations and their attributes. Through the use of structural equation modeling and confirmatory factor analysis, researchers can draw inferences about the reliability of their measures. Recent research applying these techniques to the study of police organizations detected measurement error in some constructs used in previous research (Maguire forthcoming). To illustrate how measurement error might apply in policing research, we present a simple and intuitive example.

When Wilson (1968a, 272) constructed his measure of police styles using arrest rates, he was one of the few researchers to acknowledge the problem of measurement error:

As a substitute measure of law enforcement policy or style, we use the arrest rate for certain offenses where we know police discretion is great and thus where police style is likely to be most evident: larceny, simple assaults, drunkenness, disorderly conduct, and driving while intoxicated. It must be emphasized, however, that police style is not always best measured by arrest rates, even for high discretion offenses.

Because Wilson candidly acknowledged the potential problems of his measure, we examine the extent to which his suspicions were warranted. Using 1994 UCR arrest data on the same five offenses as Wilson used, and with roughly the same sample selection criteria, we estimated a confirmatory factor analysis (measurement) model treating each of the offenses as a measured indicator of a latent police style variable.¹³ Several findings are noteworthy. First, by all conventional standards, and despite our best efforts to the contrary, the model fit

the data poorly.¹⁴ Second, those offenses that were victimless and presumably the most discretionary (driving under the influence, drunkenness, and disorderly conduct), had the lowest loadings on police style. Those that were the least discretionary and likely brought to the attention of police by victims (larceny and simple assault), had the highest loadings. If the common variation across these five variables were truly a measure of police style when responding to discretionary offenses, the opposite pattern would be expected. Third, the indicator with the lowest loading was drunkenness. The latent variable (police style) accounted for only 10 percent of the variation in drunkenness. The rest of the variation in drunkenness was unreliable, which suggests that this indicator contributed mostly error to the model and should be dropped. Fourth, even though the fit of the model improved after dropping drunkenness as an indicator, the model still fit the data poorly ($p=0.001$). The unmistakable conclusion is that this model is full of measurement error and is not a good measure of police style. Any theory suggesting that the shared variance in arrest rates for these five offenses can be used to measure some organizational property needs to be reexamined.

This brief exercise was not meant to denigrate Wilson's (1968b) classic work. Wilson himself referred to his measure of police style as crude; therefore, we suspect he would probably agree with our findings. The point of this exercise was to demonstrate the severity of measurement error that is likely present in many relatively simple constructs used by social scientists to measure organizational phenomena. Yet, researchers have consistently ignored the issue of measurement error.

Once researchers have measured variation in police organizations, the next natural step is to explain why such differences exist. Like measurement, explanation is one of the principal goals of social science research.

There are many sources of error that affect the quality of measures. Some of these are random because they cannot reasonably be controlled or inferred by the researcher. Others are more systematic and can be dealt with through careful research designs. Still others constitute deliberate efforts by those responsible for keeping records to falsify data, as reported in several U.S. cities during the past few years. Addressing the full range of potential error in social measures is beyond the scope of this paper. Our point here is to highlight the need for researchers and policymakers to consider these issues carefully. Detecting and correcting the sources of measurement error in these constructs is an important and worthwhile undertaking, especially when trying to develop explanatory models.

Explanation

Once researchers have measured variation in police organizations, the next natural step is to explain why such differences exist. Like measurement, explanation is one of the principal goals of social science research. Social scientists usually arrive at explanations for social phenomena in one of two ways: induction and deduction. Using the inductive method, researchers collect data and then analyze or search for patterns in that data. Based on their observations and analyses, researchers develop theories. Using the deductive method, they begin by specifying a theory and then collect and analyze data to test the theory. In reality, these two processes tend to overlap. For example, social scientists often begin by stating an explicit theory and collecting data to test the theory (deductive method). Upon finding that the data only partially support the theory, researchers often modify the theory accordingly (inductive method). In the next two sections, we show how social scientists have used both methods to develop, test, and modify explanations about both what police organizations are and what they do.

Social scientists use the term “explanation” to refer to explanations for why a trait varies across time and/or place. For instance, some police organizations are steeped heavily in paramilitary culture; others appear more democratic and less rigid. When social scientists try to explain paramilitarism in police organizations, they are trying to explain why some organizations are more paramilitary than others. Explanations in social science nearly always have the goal of *explaining variation* among units of analysis.

For example, if we believe police organizations in turbulent political climates are less productive than others in terms of clearance rates, then to properly test the theory, we must collect data from a sample of police organizations in different political climates. If we were to study only police organizations in hostile political climates, we could not test the theory because we have nothing with which to compare them. In social science terms, this test would be flawed because the independent variable does not vary. A similar flaw, in which the dependent variable does not vary, is present in much of the current popular management literature. Many books in this genre study successful companies and identify common attributes. The flaw, of course, is that these same attributes might be present in unsuccessful companies, but we cannot know for sure because they were not studied (King 1999).¹⁴ Peters and Waterman’s *In Search of Excellence*, probably the most influential book in this genre, has been criticized for this and other reasons (Aupperle, Acar, and Booth 1986).

The key to developing, testing, modifying, and understanding social science explanations is *comparison*. The comparative method has come to be associated

with multinational research, but comparative research can focus on many types of organized collectivities, from police departments and schools to nations and societies (Blau, Heyderbrand, and Stauffer 1966; Ostrom 1973; Ragin 1987). It is a cornerstone of sociological research on organizations (Langworthy 1986; Scott 1992). The selection of a unit of analysis for comparison depends on the research question. If we want to know why some police organizations are more effective than others, our unit of analysis is police organizations. If, on the other hand, our interest is in how a single organization changes over time, our unit of analysis is the organization at specific points in time (e.g., the month or year). Sometimes the unit of analysis is more complex, combining cross-sections (organizations) and times (years). For instance, if we want to determine if changing the number of officers in municipal police agencies affects clearance rates, we need to collect and analyze data from multiple organizations at different times. Whether comparing multiple organizations, the same organization at different times, or both simultaneously, comparison is central to understanding social science explanations.

This section examines how social scientists have sought to develop explanations for various features of police organizations. Throughout this section, the concepts we have just discussed—explaining variation, units of analysis, and comparison—will continue to appear as important themes. The most common unit of analysis in our discussion is the police organization, and the studies we discuss usually allow for comparisons by including observations from a sample of such organizations. Nearly all of these studies try to explain why some police organizations are different than others, isolating the factors thought to be responsible for producing these variations. One thing that should become clear throughout this essay is how measurement—the focus of the previous section—is inextricably linked with explanation.

Explaining variation in police organizations

The comparative study of police organizations was born in the early 1960s. Following a trend in the administrative sciences, policing scholars began to explore the role of the environment in determining the nature of a police organization. Organizational scholars of that era were profoundly influenced by a series of studies stressing the importance of the environment on organizations. Classic works by Burns and Stalker (1961), Eisenstadt (1959), Emery and Trist (1965), and Lawrence and Lorsch (1967) introduced a new way of thinking (and talking) about organizations and their problems. Based on their influence, scholars, managers, and others interested in organizational life talked about the “fit” between an organization and its environment. The environment consists of everything external to an organization that is important for its functioning and

survival. “Funding agencies, raw materials, clients, potential employees, the media, politicians, rumors, legislation and employees’ unions all reside in an organization’s environment” (Maguire forthcoming).

Initial discussions of the linkage between police organizations and their environments were both subtle and implicit. For instance, Stinchcombe (1963) argued that the distribution of public and private spaces within a community has important effects on administrative practices and aggregate patterns of police behavior. Of particular importance here is his notion that different concentrations of public places within communities might account for differences between urban and rural policing. At around the same time, Wilson (1963) developed a theory linking the professionalism of police agencies to local government structure and political ethos. Though both of these early works seem to have disappeared from the landscape of modern police scholarship, they helped plant the seeds for a growing wave of police research and theory.

Presumably influenced by these earlier works, Reiss and Bordua (1967) highlighted some of the effects that the environment might have on police organizations.¹⁶ They argued that the environmental perspective was especially important for police organizations, since “the police have as their fundamental task the creation and maintenance of, and their participation in, external relationships” (Reiss and Bordua 1967, 25–26). Reiss and Bordua described the “internal consequences” of three broad environmental features: the nature of the legal system, the nature of illegal activity, and the structure of civic accountability. They also noted several other environmental features that might be important in shaping police organizations. It is perhaps one indicator of halting progress in the study of police organizations over the past three decades that important theoretical propositions outlined by Reiss and Bordua have still not been empirically tested.¹⁷

These early works focused attention on some of the factors responsible for variation in police organizations—both what they are and what they do—across time and place. Yet, the appearance in 1968 of James Q. Wilson’s classic book, *Varieties of Police Behavior*, signified the first attempt to formulate a theory of police departments *as organizations* and test the theory using a variety of qualitative and quantitative methods (Langworthy 1986; Maguire forthcoming). Wilson’s book continues to influence police scholarship, though empirical research has still not effectively tested all of Wilson’s propositions (Slovak 1986).¹⁸ Nevertheless, these early works set the stage for three decades of research on interagency variation in police organizations. With this brief historical backdrop in mind, we now discuss the evolution of this body of research, starting with what police organizations do.

Empirical research on what police organizations do

This section reviews efforts to explain some of the external features of American police organizations, including their output, styles, and performance. Wilson's *Varieties of Police Behavior* (1968b) was the first and most influential attempt to explain the output and behavior of police agencies, which include arrest rates and styles of policing. Wilson's theory posited that local contingencies such as characteristics of the population, the form of government, and political culture shape agency behavior and therefore output. Wilson's work was the first to research the causes and correlates of police organizational output, which is most frequently operationalized as aggregate arrest rates for various offenses (Crank 1990; Langworthy 1985b; Monkkonen 1981; Slovak 1986; Swanson 1978).¹⁹ More recent research extends these traditional output measures to include community policing activities, attempting to generate theoretical and empirical explanations for interagency variation in these activities (Maguire et al. 1997; Zhao 1996). Overall, this body of research seeks to determine whether the environmental, historical, and other contextual circumstances (known in organization theory as contingencies)²⁰ of police organizations play roles in shaping organization output and performance. This literature includes a range of theoretical explanations that has not been tested empirically (e.g., Crank 1994; Crank and Langworthy 1992; Duffee 1990). In addition, there is a large body of empirical research in this area that ranges from being nearly atheoretical to almost wholly guided by theory.

Exhibit 1 lists 21 studies that seek to explain variation in what police organizations do. All of the studies meet several criteria: (1) the dependent variable is an organizational property, (2) there is at least one explanatory variable, (3) the study is based on quantitative data, (4) it reports the results of a statistical analysis (loosely defined) of the data, and (5) the total number of observations in the analysis is at least 20 (to allow for adequate comparison). Because our focus is on what police organizations do, we do not include studies in which the dependent variable is a measure of crime. Although police organizations may have an effect on crime rates, crime is not an organizational property; in the parlance of performance measurement, it is an *outcome* rather than an *output*. The remainder of this section will draw on the information presented in exhibit 1. Not all of the studies listed in exhibit 1 will be discussed. Instead, we will select those studies that have paved the way for subsequent research, that have had a profound effect on the field, or that have contributed uniquely to the study of police organizations.

Wilson's *Varieties of Police Behavior* (1968b) was the first empirical study to use quantitative data from a sample of police agencies to explain what police organizations do. This analysis was separate from the bulk of the book, which

Exhibit 1. Empirical studies seeking to explain variation in what police organizations do

Study	Cross-section	Time	Organizational dimension	Measures of organizational dimension	Major findings
Wilson 1968b	146 police departments	1960	Style	Arrest rates for larceny, DWI, drunkenness, disorderly conduct, and simple assault.	The political culture of a community constrains the style of policing.
Ostrom, Parks, and Whitaker 1978b	1,159 police and sheriffs' departments	1974-75	Effectiveness	Nine separate indicators of police effectiveness (too numerous to summarize).	Smaller departments are generally as or more effective than larger agencies.
Swanson 1978	40 police departments	Unknown	Arrest patterns	Arrest rates for 11 offenses.	Organizational factors are less important than environmental factors in shaping arrest patterns in police agencies.
Wilson and Boland 1979	35 police departments	1975	(a) Arrest productivity (b) Patrol aggressiveness	Arrest rates for robbery. Moving violation citations per patrol unit.	Aggressive patrol strategies increase robbery arrests, and workload (crimes per patrol unit) decreases arrest productivity. Political culture and opportunity (automobiles per capita) increase the number of citations issued.

continued

Exhibit 1 (continued)

Study	Cross-section	Time	Organizational dimension	Measures of organizational dimension	Major findings
Monkkonen 1981	(a) 23 police departments (b) 37 police departments 72 police departments 127 police departments	1860–1920 1880 1890 1903	Arrest patterns Arrest patterns	Three categories of arrest rates: total arrests, “initiative” arrests, and crime arrests. Total arrests.	Overall, this historical analysis shows that from 1860 to 1920, the police shifted their attention from controlling the “dangerous class” to controlling crime.
Liska and Chamlin 1984	76 police departments	1972	Arrest patterns	Arrest rates for personal and property offenses.	Supports conflict theory. Variables related to social structure and the racial/economic composition of the community have a substantial effect on arrest rates.
Surette 1984	Chicago Police Department	1873–1973	Arrest patterns	Two categories of arrest rates: the sum of all felony and misdemeanor arrests, and vagrancy arrests.	Police organizational activities are related to economic conditions for only one of the three dependent variables studied: vagrancy arrests. Police strength and serious arrests are unrelated to economic factors.
Langworthy 1985b	152 police departments	1975	Style	Arrest rates for larceny, DWI, drunkenness, disorderly conduct, and simple assault.	Generally supports Wilson’s finding that the political culture of a community is related to styles of policing. By focusing on means and variances of arrest rates, he finds that Wilson’s theory should be regarded as a theory of central tendency rather than a theory of constraints.

Exhibit 1 (continued)

Study	Cross-section	Time	Organizational dimension	Measures of organizational dimension	Major findings
Meagher 1985	249 police departments	Unknown	Style	Frequency with which officers performed 647 job tasks.	Although there is variation in what police officers do across agencies there is a great deal of commonality. Police officers from all agencies perform a common set of core tasks.
Murphy 1986	Unknown	Unknown	Style	Arrest rates.	"Bureaucratic R.C.M.P. detachments produce an aggressive policing style resulting in high arrest rates. Less bureaucratized police departments produce a more informal policing style with low rates of arrest" (abstract).
Slovak 1986	42 police departments	1976-79	Style	Arrest rates for Part 1 violent crimes and property crimes.	Environmental and organizational factors both contribute significantly to the variance in police legalism.
Cordner 1989	84 Maryland police and sheriffs' agencies	1985	Investigative effectiveness	Clearance rates, which are calculated by dividing the number of crimes cleared by either arrest or "exception" by the total number of reported crimes.	Region of the State (essentially a measure of urbanization) is the strongest predictor of investigative effectiveness. The proportion of crimes that include property offenses has weak and inconsistent effects, and workload (crimes per officer) and organization size have no effects.

continued

Exhibit 1 (continued)

Study	Cross-section	Time	Organizational dimension	Measures of organizational dimension	Major findings
Crank 1990	284 Illinois police departments	1986	Style	Arrest rates are computed for four offense categories: trespassing, disorderly conduct, motor vehicle offenses, and various crimes related to the use or sale of marijuana.	Both environmental and organizational variables have an effect on arrest rates. However, the magnitude and direction of these effects are different for urban and rural police departments.
Crank 1992	392 Illinois police departments	1986	Style	Arrest rates for aggravated battery, aggravated assault, burglary, theft, and motor vehicle theft.	Organizational factors are more important than environmental factors in shaping police styles with regard to legally serious crime.
Davenport 1996	78 Texas police departments	1994	Performance	Clearance rates for Part 1 Index offenses.	Most of the organizational and environmental measures used did not have a significant effect on performance. Environmental complexity had the most significant influence.
Zhao 1996	228 police departments	1993	External community policing reforms	Additive index composed of 12 indicators of "externally focused change."	Variation in externally focused change across agencies can be explained by features of the focal organization and by levels of social disorganization in the community.
Maguire et al. 1997	5,726 law enforcement agencies	1994	Community policing	Additive index composed of 31 community policing activities.	Interagency variation in community policing can be partially explained by region of the country and size of the agency.

Exhibit 1 (continued)

Study	Cross-section	Time	Organizational dimension	Measures of organizational dimension	Major findings
Roth and Johnson 1997	1,471 law enforcement agencies	1995 and 1996	External community policing reforms	Additive indices measuring community partnerships (11 items); problem solving (11); and primary (5), secondary (5), and tertiary (4) crime-prevention activities.	Agencies receiving Federal funding were more likely to participate in community partnership and problem-solving activities and, to a lesser extent, in crime prevention activities.
Gianakis and Davis 1998	90 Florida law enforcement agencies	Unknown	Community policing	A nine-category Guttman-type scale measuring the depth of community policing implementation (ranging from patrol philosophy only to restructuring the department and decentralized substations).	Depth of community policing implementation was unrelated to agency type, agency size, accreditation status, or date of community policing adoption.
Maguire and Katz 2000	1,604 law enforcement agencies	1993	Organizational claims regarding community policing	Categorical variable measuring agencies' general claims about participating in community policing.	Police agencies' general community policing claims are loosely coupled with their actual activities. These general claims are shaped in part by the size and region of the agency.
Maguire et al. 2000	1,604 law enforcement agencies	1993	External community policing reforms	Additive indices composed of problem-solving and community partnership activities.	The proposed measurement model of community policing fits the data. Size and region are significant predictors of both internal and external community policing activities.

provided the better known details of his taxonomy of police styles (legalistic, watchman, and service). Wilson's theory was that local political culture constrains (but does not dictate) the style of policing within a community. We have already discussed the methods used by researchers to measure police styles. Wilson argued that measuring both style and political culture would be "exceptionally difficult if not impossible" (p. 271). Nevertheless, considering it to be worthwhile, he constructed a "substitute" measure of political culture focusing on the form of government, the partisanship of elections, and the professionalism of city managers (based on their education and experience). Acknowledging the presence of measurement error in his constructs, Wilson concludes: "[T]he theory that the political culture of a community constrains law enforcement styles survives the crude and inadequate statistical tests that available data permit" (p. 276).

Since the appearance of Wilson's book, a number of empirical studies of police organizational style have appeared. All of them measure police style using arrest rates for some mix of offense types, usually less serious offenses thought to be subject to greater discretion. Most of these studies find that organizational and environmental characteristics play a significant role in shaping police style, though there is little consensus or uniformity about what kinds of explanatory variables are important. Several other studies use arrest rates as a dependent variable but do not treat them as measures of police style. They are usually referred to more generally as indicators of organizational activity, behavior, or productivity.

Other empirical studies listed in exhibit 1 focus on effectiveness or performance, which are usually measured using objective criteria such as clearance rates or subjective criteria such as citizen evaluations of local police performance (Alpert and Moore 1993; Bayley 1994; Parks 1984).²¹ One issue these studies address, in part, is whether bigger police departments are better, as some critics of American policing have claimed (e.g., Murphy and Plate 1977). Subjective studies of police performance conducted by Elinor Ostrom and her colleagues suggest that bigger is not necessarily better.²² Cordner's (1989) examination of investigative effectiveness in Maryland found that the region of the State, which appears to be a proxy for urbanization, was an important predictor, but that crime, workload, and department size were generally insignificant. Davenport (1996) is the only scholar to test a model in which the environment has both a direct effect on department performance and an indirect effect on performance through organization structure. His findings are too numerous to summarize, but the most important predictor of department performance was the complexity of the environment. Overall, this line of research has generated mixed results. Therefore, it is not useful for generating axioms about the

performance and effectiveness of police organizations. Probably the most consistent finding is that larger police organizations are not necessarily more effective, and in many cases they are less effective than smaller agencies.

In the past few years, in response to the need for better measures of what police organizations do, researchers have begun to measure other facets of police behavior. We have already discussed the emergence of a new trend toward measuring the implementation of community policing in American police agencies. Using these measures, researchers have also attempted to explain variation in community policing across agencies. Using data from a national survey of police organizations, Zhao (1996) was the first researcher to test an empirical model explaining community policing. Recognizing that community policing is a catchphrase for a potpourri of different reforms, Zhao divided it into external and internal components, measuring and estimating models for each one separately. Exhibit 1 provides a summary of Zhao's findings regarding externally focused change because this index measures community policing activities occurring outside the police organization and in the community. His results for internally focused change are presented in the next section because these consist primarily of administrative reforms. For the five other community policing studies in exhibit 1, only the external components are listed if they can be easily separated, with the internal components appearing in exhibit 2. These studies use various methods to construct measures of community policing and then try to explain interagency variations in these measures. Probably the most consistent finding in these studies is the important role of region and department size in shaping community policing.²³ Emerging research continues to address the causes and consequences of adopting community policing.

Empirical studies seeking to explain what police organizations are

The topic of this section—explaining what the police are—was the last to emerge of the topics examined in this chapter. The reason, as in organizational studies in general, is that people likely are more interested in how organizations behave and what they produce than in mundane administrative details such as how they are structured. This is especially true in policing, where the bottom line is typically considered to be crime, a subject of fascination to many Americans. Although reams of paper have been expended by reformers trying to convince police administrators to change the structures and internal operating processes of police organizations, scholarly progress in producing theory and research on these organizational features has been slow. In this section, we trace the development of research on internal variation in police organizations, including structure, policy, and other administrative attributes.

Once again, we return to James Q. Wilson's *Varieties of Police Behavior* (1968b). Wilson's analysis did not explicitly consider internal organizational attributes as an object of study, but he refers throughout the book to the structural correlates of police style. Langworthy (1986) considered Wilson's work "the only empirically derived theory of police organization to date." Langworthy (p. 32) summarized Wilson's implicit linkage between style and structure as follows:

Watchman police departments were said to emphasize order maintenance, to be hierarchically flat, unspecialized, and decentralized. Legalistic departments were characterized as oriented toward vigorous law enforcement, hierarchically tall, specialized in law enforcement function, and centralized. Service-style departments were described as responsive to requests for aid or action, highly specialized across a broad range of functions, decentralized in operations, and centralized administratively.

Thus, although Wilson's work is best remembered as a theory of police style, it also contains an implicit theory of police organizational structure.

The first empirical studies in this genre did not appear until the mid-1970s, emerging, like Wilson's work, from political science and urban studies. In 1975, Henderson published an empirical study in the *American Journal of Political Science* on the correlates of professionalism in sheriffs' agencies. The study falls within the class of theory and research that Langworthy (1986) classifies as normative because defining and measuring police professionalism requires the researcher to make personal judgments about what it means to be professional. It was the first (and perhaps only) study to treat professionalism as an organizational, rather than an individual, attribute. In 1976, Morgan and Swanson published an article in *Urban Affairs Quarterly* examining a number of organizational attributes. With little regard for theory, the researchers used exploratory factor analysis to construct both their independent and dependent variables. According to the Social Science Citation Index, neither study has been cited often (seven for Henderson and one for Morgan and Swanson), suggesting that the birth of empirical research on the causes and correlates of police organization was anonymous.

During the 1970s, Elinor Ostrom and her colleagues collected enormous amounts of data on American police organizations, even by today's standards. They examined policing as an "industry," focusing on patterns in the production and consumption of police services. In dozens of publications, most notably their book *Patterns of Metropolitan Policing* (Ostrom, Parks, and Whitaker 1978a), they described how police organizations in metropolitan areas rely on one another for mutual support and to provide specialized services.

Their work defied critics who argued that American policing was a loosely connected patchwork of small and untrained police agencies, often consisting of only a handful of officers (Murphy and Plate 1977; Skoler and Hetler 1969). Although the work of Ostrom and her colleagues made significant contributions to the study of policing, the focus of nearly all of their publications was the metropolitan area and its patterns of service production and consumption, not police organizations.²⁴ For that reason, most of their work falls outside the scope of this essay. Their focus on the internal consequences of police organizational size, however, was one of the earliest studies seeking to explain variations in police organizational structure (Ostrom, Parks, and Whitaker 1978b). We will return to their findings on organizational size.

Probably the most influential work in this area is Robert Langworthy's 1986 book, *The Structure of Police Organizations*. Langworthy argued convincingly that with the exception of James Q. Wilson's work, scholarly attention to police *organizations* had been restricted to normative theories and prescriptions about how they should be structured and what they should be doing. This tendency to prescribe rather than describe and explain, to say what police *should* be doing rather than what they *are* doing and why they are doing it, left a gap in our understanding of police organizations.²⁵ To begin filling this gap, Langworthy borrowed a series of propositions from organization theory (and from Wilson's work), constructing his own theory to explain variation in the structure of police organizations. Using data from two national surveys, including data from Ostrom and her colleagues and the Kansas City General Administrative Survey, Langworthy tested his theory empirically. His analysis was the first comprehensive empirical study to treat the structure of police organizations as a dependent variable. Langworthy (1986, 136) concluded that the causal forces in his study did not appear to exert a significant constraint on organization structure:

This tendency to prescribe rather than describe and explain, to say what police should be doing rather than what they are doing and why they are doing it, left a gap in our understanding of police organizations.

It seems plain that the explanations, size, technology, population mobility, population complexity, and type of local government, although theoretically significant determinants or correlates of agency structure, explain very little of the variance in agency structure. The constraints, when they are suggested by the data, do not appear insurmountable.

These findings suggest that American police executives are, by and large, free to design police organizations as they see fit.

Research on the causes and correlates of police organizational structure continues to emerge. Crank and Wells (1991) found that size exerts a nonlinear effect on structure. King's (forthcoming) recent study on the effects of organizational age found that older police organizations employ fewer civilians than younger ones. Davenport (1996) found that violent crime, resource capacity, and environmental turbulence have mixed effects on measures of structure.²⁶ Maguire's (forthcoming) replication and extension of Langworthy's study found a series of mixed effects of age, size, technology, and environment on structure. Maguire divided the structure of police organizations into two domains: (1) structural complexity, and (2) structural coordination and control mechanisms. He found that the context of police organizations constrains structural complexity (vertical, functional, and spatial differentiation) but not structural control and coordination (formalization, centralization, and administrative intensity). Maguire and Langworthy are currently working with Jihong Zhao to replicate the findings from both of their studies, examining the causes and correlates of structural change during the community policing era. Overall, the study of police organizational structure has entered a stage of incremental development.

Other studies in this genre have examined the environmental and organizational correlates of police innovation and various internal (administratively oriented) community policing reforms. Based on the voluminous literature on innovation diffusion, Weiss (1997) examines two interesting questions: Do police organizations rely on informal communications with other agencies (peer emulation), and, if so, do these contacts result in the diffusion of innovation across agencies? Using innovative methods that we will discuss later, he found that agencies engage in informal information sharing and that peer emulation and cosmopolitanism both shape the adoption of innovations. King (1998) also examines the sources of innovation in police agencies, but his research is rooted more in traditional organizational theory than the diffusion literature. Using factor analysis, King examined the dimensionality of police innovation. He found that innovation is a multidimensional concept consisting of at least five separate dimensions: radical, administrative, technical, line-technical, and programmatic. Furthermore, he found additional evidence that at least some of these dimensions can be further reduced into subdimensions. The findings are too numerous to summarize here, but overall, organizational factors played a stronger role in shaping innovation than did environmental or "ascriptive" factors.

Several studies have examined just one category of innovation: the kinds of administrative changes occurring under the banner of community policing. Zhao (1996) was the first researcher to examine the causes of internally focused changes occurring under community policing. He constructed an additive index of internal change and then sought to explain variation in the measure using a

number of organizational and environmental predictors. His models were able to explain more of the variance in externally focused than internally focused change. In their evaluation of the Justice Department's COPS Office, Roth and Johnson (1997) found that although Federal funding may have affected external elements of community policing, agencies that received the funding were no more likely than nonfunders to have made internal organizational changes. Finally, in a study focusing on measurement rather than explanation, Maguire and colleagues (1999) developed reliable measures of internal change, which they termed "adaptation." Although region and department size were only included in the model for statistical reasons, once again, both were found to have a significant effect on adaptation.

Explaining what the police are—their policies, structures, programs, and other elements—probably represents the next frontier of research on police organizations. The research in this area is relatively undeveloped, and there is an untapped pool of theories to test. For instance, promising theories that were developed in the 1960s have still not been fully tested. These include the work of Reiss and Bordua (1967) and a number of propositions about police agency structure implicit in Wilson's (1968b) theory of police behavior (Langworthy 1986). In addition, there have been a number of recent theoretical contributions in the areas of contingency theory (Maguire forthcoming), institutional theory (Crank 1994; Crank and Langworthy 1992; Katz 1997; Mastrofski and Uchida 1993), resource dependency theory (Katz, Maguire, and Roncek 2000), and various combinations of these theories (Maguire, Zhao, and Lovrich 1999; Mastrofski and Ritti forthcoming). In part 3, we describe these theories and their promise for helping us understand police organizations.

What factors shape police organizations?

Many of the same variables are used to explain interagency variation in both what police organizations do and what they are. Undoubtedly, one reason for this is the availability of these measures in common sources such as Census Bureau publications and data or the *Municipal Yearbook*. Another reason is that many of the same theories are used to explain differences across police agencies. This section briefly reviews some of the factors thought to shape police organizations.

Isolating the factors that shape police organizations with any degree of certainty would require a book-length discussion. The studies listed in exhibits 1 and 2 contain at least 85 separate independent variables, even after combining those that are similar but not exactly the same (such as two different measures of political culture). The following list contains the 14 measures that had at least 1 significant effect in at least 3 separate studies. They are sorted in descending order by the number of studies in which they demonstrated a significant effect:

Exhibit 2. Empirical studies seeking to explain variation in what police organizations are

Study	Cross-section	Time	Organizational dimension	Measures	Major findings
Henderson 1975	65 Florida sheriffs' agencies	1969	Professionalism	Four-item factor score reflecting training and examination policies.	Environmental variables are more significant predictors of professionalism in sheriffs' agencies than the personal characteristics of the sheriff.
Morgan and Swanson 1976	40 police departments	1970	Organizational commitment to community relations	Factor score based on the percentage of employees with CR training and the percentage of CR employees.	Political culture is the only explanatory variable (12 total, 10 of which are factor composites) with a statistically significant effect on organizational commitment to community relations.
Walzer 1970	31 Illinois police departments	1958 and 1960	Efficiency	Municipal police expenditures divided by an index of services that measures the volume of police activity.	As the volume of police activity increases, the ratio of expenditures to activity decreases. This finding suggests an economy of scale in police organizations. Accordingly, the author suggests that police consolidation may be one possible solution to lowering costs.
Ostrom, Parks, and Whitaker 1978b	1,159 police and sheriffs' departments	1974-75	Structure	Administrative intensity and patrol density.	Smaller agencies assign a greater proportion of officers to patrol and a lower proportion to administration.
Monkkonen 1981	23 police departments	1860-1920	Structure	The ratio of nonpatrol to patrol officers is treated as an indicator of overall bureaucratization.	Police organizations became more bureaucratized from 1860-1920. Changes in the structure of policing "came about through external demands upon the police... and, to a lesser extent, through the 'natural' growth of the police bureaucracy" (p. 147).

Study	Cross-section	Time	Organizational dimension	Measures	Major findings
Heininger and Urbanek 1983	100 police departments	1970-80	Civilianization (structure)	Percentage of employees who are not sworn officers.	Growth in civilianization is not associated with any of the five explanatory variables examined.
Langworthy 1983	69 police departments	1977	Structure	Administrative intensity and patrol density.	Refutes previous findings by Ostrom, Parks, and Whitaker (1978), based on measurement and sampling issues. Finds that there is no relationship between size and administrative intensity and only a very weak positive relationship between size and patrol density.
Steinman 1984	949 police departments	1975-76	"Rationalization"	An additive index composed of 10 binary variables relating to various structural and procedural features.	Police departments with higher expenditures and those employing chiefs with managerial training and/or educated chiefs are more rationalized. Rationalization is unrelated to numerous other explanatory variables.
Langworthy 1986	(a) 106 police departments (b) 69 police departments	1974-75 1977	Structure	Five measures of organizational structure were used: administrative overhead; and spatial, hierarchical occupational, and functional differentiation.	The effects of population and agency size on structure could not be disentangled. Agency size is strongly related to spatial differentiation but weakly related to other structural variables. Technology is strongly related to functional differentiation, less so to occupational differentiation, and weakly associated with or independent of other structural elements. Population complexity had a significant effect on only occupational differentiation. Good government cities have fewer ranks and employ more civilians than cities with mayor-council governments or those holding partisan elections.

continued

Exhibit 2 (continued)

Study	Cross-section	Time	Organizational dimension	Measures	Major findings
Slovak 1986*	42 police departments	1976-79	Structure	Three elements of structure were used: administrative intensity (an inverse of span of control), division of labor (percentage of officers in patrol), and civilianization.	Organizational size has no effect on division of labor. "Administrative intensity is a negative function of organizational size (recall that span of control, as measured here, is an inverse indicator of intensity) but is not directly related to differentiation; the innovation of civilianization is promoted by administrative intensity (again, recall the caveat about measurement) but is not directly affected by either size or differentiation" (p. 23).
Crank 1989	505 Illinois police departments	1973 and 1986	Civilianization (structure)	Percentage of employees who are not sworn officers.	Increasing civilianization was associated with community budgetary growth, police organizational downsizing, ruralization, and, to a lesser extent, reductions in crime.
Crank and Wells 1991	399 Illinois police departments	1986	Structure	Measures of four structural variables were used: height (number of ranks), civilianization, concentration, and supervisory ratio.	Variations in size were associated with three of the four dimensions of organizational structure examined: height, concentration, and supervisory ratio. When controlling for organizational size, urbanism did not have an independent effect on organizational structure.

* Slovak's model contains five "structural" variables. Maguire's forthcoming book on police organizational structure argues that two of these variables are not structural. Based on that argument, we include only three of Slovak's five variables here.

Exhibit 2 (continued)

Study	Cross-section	Time	Organizational dimension	Measures	Major findings
Davenport 1996	78 Texas police departments	1994	Structure	Two measures of structure were formed using exploratory factor analysis: organizational complexity and occupational differentiation.	Violent crime is the only one of four environmental variables with a significant (positive) effect on organizational complexity. Resource capacity and environmental turbulence have a significant positive effect on occupational differentiation. Overall, environmental variables explain only 16 percent of the variance in complexity and 17 percent in occupational differentiation.
Zhao 1996	228 police departments	1993	Internal community policing reforms	Additive index composed of six indicators of "internally focused change."	Variation in internally focused change across agencies cannot be explained as well as externally focused change (using organizational and environmental variables).
Maguire 1997	236 police departments	1987-93	Structure	Includes measures of five structural variables: vertical, functional, and occupational complexity; formalization; and administrative density.	Although police organizational structures are changing over time, these changes are unrelated to community policing.
Roth and Johnson 1997	1,471 law enforcement agencies	1995 and 1996	Internal community policing reforms	Additive index measuring organizational changes associated with community policing.	Agencies receiving Federal funding were not more likely than nonfunders to have made internal organizational changes associated with community policing.
Weiss 1997	134 police departments	Unknown	(a) Peer emulation (b) Innovation	Both concepts were measured using measurement models with multiple indicators and multiple "key" informants.	Risk mediation and cosmopolitanism each have a significant effect on peer emulation. Cosmopolitanism has a significant effect on innovation.

continued

Exhibit 2 (continued)

Study	Cross-section	Time	Organizational dimension	Measures	Major findings
King 1998	432 police departments	1993	Innovation	Ten dimensions of innovation were derived from exploratory factor analysis of 27 innovative police programs, activities, and functions.	Innovation is a multidimensional concept. Studies that measure innovativeness using only one or a handful of dimensions are trampling on complexity. Organizational variables play a stronger role than environmental or contextual variables in explaining innovativeness.
King forthcoming	334 police departments	1990	Structure	Five structural variables are included: civilianization, hierarchical and functional differentiation, formalization, and administrative overhead.	Older police departments employ fewer civilians than younger departments. There is no relationship between organizational age and structure for the remaining four dependent variables.
Maguire forthcoming	432 police departments	1993	Structure	Three measures of structural differentiation were used: vertical, functional, and spatial. Also includes three measures of structural coordination and control: formalization, centralization, and administrative intensity.	Environmental and contextual variables play a significant role in shaping structural differentiation but are unable to explain substantial portions of the variance in structural coordination and control mechanisms. The size and age of a police organization, the instability and dispersion of its environment, and the routinization of its social technologies all play a role in shaping its structure.
Maguire et al. 2000	1,604 law enforcement agencies	1993	Internal community policing reforms	Additive index of internal administrative changes related to community policing.	The proposed measurement model of community policing fits the data. Size and region are significant predictors of both internal and external community policing activities.

1. Organizational size (18).
2. City governance (5).
3. Region (5).
4. Concentration (4).
5. Crime patterns (4).
6. Organizational age (4).
7. Political culture (4).
8. Population size (4).
9. Population heterogeneity (4).
10. Poverty/income (3).
11. Urbanization or ruralization (4).
12. Span of control or supervisory ratio (3).
13. Time (3).
14. Vertical differentiation (3).

We are careful not to make too much of these findings. This list is simply intended to illustrate the kinds of variables that researchers have used to explain differences in police organizations and those that have been found important. These findings pertain to several different dependent variables, and neither the direction of effects nor the quality of the studies is considered. Nevertheless, this list illustrates some of the factors commonly thought and found to influence police organizations.

The most frequent and consistent finding in organizational research on police is the importance of organization size. The effects of size are not universal, as Ostrom and her colleagues have demonstrated; the research suggests that size has an important effect on style, structure, and processes, but not necessarily on effectiveness and efficiency. Region also continues to have a significant effect on the administration of public organizations. Yet, to date, researchers have not done a good job of isolating the theoretical reasons for these effects, though many possibilities have been suggested (Maguire et al. 1997). The structure of city governance, together with local political culture, also continues to have a significant effect on police organizations, suggesting that any comprehensive theory of police organizations needs to account for political effects. Another particularly noteworthy finding is the presence of two variables that suggest a historical effect on police organizations: the department's age and the passage of time. Despite assertions to the contrary, police organizations *do* change. The appearance of time and age in this list suggests that they change in ways that are sometimes predictable. Thus, any comprehensive theory of police organizations needs to account for historical effects. The remaining variables are all elements of the organization or its environment, and most are represented in traditional organizational theories.

Future Prospects

The empires of the future are the empires of the mind.

—Sir Winston Churchill (1944, 238)

The comparative study of police organizations has not evolved in a progressive, orderly fashion. Much of the research contains methodological and theoretical shortcomings, and for that reason it has been of limited use for understanding police organizations and the forces that shape them. A byproduct of this limitation is that this research has been of little practical use for police executives and policymakers. More than two decades ago, Dorothy Guyot (1977) bemoaned the lack of empirical research on police organizations, citing Wilson's *Varieties of Police Behavior* (1986b) as the lone exception. Nearly a decade later, Robert Langworthy (1986, 32) echoed Guyot's complaints, arguing that Wilson's work "remains the only empirically derived theory of police organization to date." Through the mid-1980s, police organizational scholarship did not substantially evolve beyond Wilson's seminal work.

Langworthy's 1986 book, *The Structure of Police Organizations*, was an important turning point in police organizational scholarship, although it met with mixed reviews. C. Ronald Huff (1987, 508) called it "an exemplar of theory construction and theory testing," and Peter Manning (1988, 323) described it as "an awkwardly written and jargon-filled monograph." A recent replication of Langworthy's study claimed "there is no doubt that it forged a new road in the study of the police" (Maguire forthcoming, 5). Langworthy's book (and other related work) takes its place among only a handful of other important studies that have blended theory and research in an effort to further our understanding about the structure and function of American police organizations. Perhaps even more importantly, it inspired a new generation of police organizational scholarship (Crank and Wells 1991; King, Travis, and Langworthy 1997; King 1999; Maguire 1997, forthcoming).

Thus, as the new millennium arrives, we cannot complain as forcefully as our predecessors about the status of police organizational scholarship. The past decade has seen a number of improvements in theory, data, and method, though much remains to be done. This section has two simultaneous goals: to diagnose some of the weaknesses in this line of research and to suggest ways that researchers might continue to breathe new life into it. We consider three primary areas: theory, research, and policy.

Theory

Throughout this chapter, we have referenced theories used by scholars to explain interagency variation in police organizations. Some of these theories have received empirical support, others have not, and others remain untested. This section briefly reviews the state of theoretical explanation in the study of American police organizations. Many of these theories make strange bedfellows, often emerging from a variety of disciplines or competing paradigms. Some are difficult to test using the quantitative and comparative methods discussed in this chapter. Others are theoretical statements without clear roots in an established body of theory. Nevertheless, all of them seem to be promising frames of reference. Some readers may find these theories to be abstract, but we urge you to think about their implications for shaping police organizations.

We begin by restating contingency theory because it is an inclusive theory of structure, process, and performance. Briefly stated, contingency theory holds that organizations will be effective only if they remain dynamic, adapting to changes in technology and environment. Technology is used in the broadest sense, referring to tools and strategies used by the organization to process raw materials. In addition to the material technologies that are having such a profound influence in policing (Ericson and Haggerty 1997; Manning 1992), it includes the social technologies used by the police to process and change people and communities (Maguire forthcoming; Mastrofski and Ritti forthcoming). The environments of police organizations are complex, but contingency theory focuses predominantly on the “task environment.” This includes those elements of the environment with direct relevance for the work of the organization. In policing, the task environment includes citizens, courts and other parts of the criminal justice system, patterns of crime and criminality, sources for recruiting and training officers, physical and social attributes of the community, and other external forces that shape the structure and function of police agencies.

Contingency theory is the implicit foundation of nearly every study of police organizations. It is the implicit source of most of the explanatory variables used in models explaining organizational features, such as size, technology, and the various elements of the environment. It is based on the assumption that effective organizations are rational entities seeking to maximize their levels of effectiveness and efficiency. It is somewhat Darwinian in its assumption that organizations that do not adapt to changes in technology and environment will be ineffective and therefore probably fail and be replaced by others (Langworthy

Most researchers who study police organizations would probably not describe them as rational, dynamic, or adaptive.

1992).²⁷ This inherent rationality is why many organizational scholars have abandoned contingency theory (Donaldson 1995). Most researchers who study police organizations would probably not describe them as rational, dynamic, or adaptive. The failure of contingency theory to effectively explain the structure and function of organizations has led to the development of other theories. We now discuss three alternative perspectives on the role of organizational environments: as sources of legitimacy, resources, and information.

Institutional theory has its roots in the early study of organizations by such influential theorists as Talcott Parsons and Philip Selznick. Selznick, for example, described institutionalization as the process by which organizations develop an “organic character” (Perrow 1986) and become “infused with value beyond the technical requirements of the task at hand” (Selznick 1957, 17). Selznick was fascinated by the paradox that organizations are created for rational action but that they never quite succeed in conquering irrationality. Institutional theory has experienced a revival over the past two decades, a trend that many attribute to an influential article by Meyer and Rowan in 1977. Meyer and Rowan argued that the environment is more than a source of raw materials, clients, technologies, and other technical elements essential to the function of an organization. Environments are also the sources of such intangible elements as standards, norms, rumors, myths, symbols, knowledge, traditions, and standards. These elements constitute the institutional environment, and though they are often less rational than elements in the technical environment, they are essential sources of organizational legitimacy. Because organizations require legitimacy to survive and prosper, they are often more responsive to institutional concerns than to technical concerns. Institutional theory has begun to occupy an increasingly important role in the study of police organizations (Mastrofski, Ritti, and Hoffmaster 1987; Crank and Langworthy 1992, 1996; Crank 1994; Mastrofski 1998a; Mastrofski and Ritti 1996, forthcoming; Maguire forthcoming; Maguire, Zhao, and Lovrich 1999; and Maguire and Mastrofski forthcoming). For instance, Katz (1997) recently used institutional theory to explain why a police organization created a gang unit although the community did not have a serious gang problem; it was to achieve legitimacy in the view of powerful local stakeholders. Due to its nature, institutional theory is difficult to measure and test in a definitive manner. Enough has been written about institutional theory in policing that finding ways to test it is an important next step.

Although institutional theory is based on the role of the environment as a source of legitimacy for the organization, resource dependency theory focuses on the environment as a source of valuable resources. The principal statement of resource dependency theory is Pfeffer and Salancik’s (1978) book, *The External Control of Organizations*. Resource dependency theory is essentially

a theory of power and politics, and the methods used by organization actors to secure the flow of resources. Because organizations frequently depend on resources from the environment, they are to a certain extent “externally controlled.” Resource dependency theory has not yet been applied to policing in a comprehensive way, though two recent papers have described its relevance to police organizations (Katz, Maguire, and Roncek 2000; Maguire and Mastrofski forthcoming).

The next theory views the environment as a source of information. Weick (1969) and Duncan (1972) have both demonstrated how various sectors of the environment contain “pools” of information that are critical to the organization. Organizations process this information in a way that decreases “information uncertainty.” As the pace of computerization in police agencies continues to grow, the role of information may become more relevant. Two recent discussions have focused on the centrality of information to police organizations. Peter Manning (1992) outlines the links between organizations, environments, and information-processing technologies such as computer-aided dispatch systems, centralized call collection (911) mechanisms, “expert” systems, management information systems, and other tools designed to increase the organization’s capacity to intake and process information. Manning concludes by suggesting that information technologies have “an indeterminate effect on the organizational structure of policing; technology is used to produce and reproduce traditional practices, yet is slowly modifying them” (1992, 391). Ericson and Haggerty (1997) explore similar though much broader themes in their recent book, *Policing the Risk Society*. They view police organizations as part of a larger network of institutions responsible for identifying, managing, and communicating risks. They argue that policing (at multiple levels) is shaped by external institutions and their need for information about risks. Theories about the environment as a source of information are not yet well developed. In addition, they contain a host of ambiguities about the proper unit of analysis.²⁸ Nonetheless, given the emergence in policing of sophisticated technologies for collecting and processing information, this perspective deserves further attention.

Although all of these theories offer substantial promise for understanding police organizations, there is also a wealth of potential explanations immediately available. Stinchcombe (1963) made a series of early propositions in which the distribution of public and private spaces within communities serves as an important source of variation in police practice administration. His work foreshadowed the emergence of large private spaces policed by private entities, such as malls, amusement parks, and gated communities. Other classic theoretical statements appearing in the 1960s (Bordua and Reiss 1966; Reiss and Bordua 1967; Wilson 1963, 1968a) have still not been adequately tested. These classics need to be dusted off and revived.

Many of the studies reviewed earlier in this chapter have not been adequately rooted in theory. Yet, it seems clear that the reason is not a lack of existing or promising theories. In diagnosing the current state of police organizational scholarship in the United States, we find little reason for concern about the nature or number of theories on which to base solid empirical research. One area for improvement is exploring how a good theory of police organization might differ from a theory of organizations in general or of public service organizations in particular. There is already some evidence that theories designed to explain private organizations, especially those in manufacturing industries, are inadequate for explaining police organizations (Maguire forthcoming). The answer may exist in either the artful blending of existing theories or the emergence of new and better ideas.

Research

In this section, we consider each of the three elements of the research process: data collection, measurement, and explanation.

Data collection

Data collection in policing is currently in an exciting and rapid state of development. Much of this can probably be attributed to the emergence of new technologies for recording, collecting, processing, and distributing data. Police organizations are now experimenting with technology at a record pace, implementing or updating their management information systems, computer-aided dispatch centers, geomapping and other modern forms of crime analysis, mobile data terminals in patrol cars, and many other advances emerging in the past decade. One consequence of the proliferation of information-processing technologies is that police agencies now contain vast archives of data. Although much of this data is not useful for national comparative research, it is changing the face of policing in important ways.

National data collection on police organizations is not in a state of crisis. Police agencies are more open than ever. Careful surveys conducted by researchers, government agencies, and survey firms routinely obtain response rates of 70 to 90 percent. There are numerous sources of data, and although most could be improved, they are generally of decent quality. In a recent article, we criticized some of the data inventories used by government agencies for counting the number of police agencies and officers in the United States over the past several decades (Maguire et al. 1998). Many of the problems cited in that article have been rectified, though some remain. Consequently, current efforts to enumerate the American police are more accurate than ever. Finally, several agencies within the Justice Department now routinely include in their police agency

databases a common FBI code, enabling researchers and policymakers to link separate databases and test interesting new hypotheses. Although there is always room for improvement in the kinds of data that are collected, the methods used to collect data from police organizations tend, on average, to be fairly good.

Our optimism here is not meant to suggest that there remain no challenges. For example, in response to the Violent Crime Control and Law Enforcement Act of 1994, BJS and NIJ have undertaken efforts to measure the use of force by police agencies throughout the Nation. BJS has added supplemental questions on police use of force to its national household survey. Although this strategy is useful for some purposes, it undercounts at least three classes of people who may be more likely to have force used against them by the police: the homeless, the incarcerated and institutionalized, and those without telephones.²⁹ A second strategy, undertaken by IACP with funding from BJS and NIJ, will be to develop a national police use-of-force database based on confidential reporting by police departments of use-of-force incidents. This method, too, contains a number of problems. Most importantly, it relies on official records that may reflect as much about the organization's willingness to record use-of-force incidents as the actual number of incidents that occur. Other entities, including the Police Complaint Center and the American Civil Liberties Union, collect data on excessive force and patterns of discrimination from citizens alleging to be victims of these offenses. Although these may serve a useful social purpose, neither group attempts nor claims to carefully enumerate use-of-force incidents nationally.

If the 1999 Traffic Stops Statistics Bill is enacted, police agencies will face a new challenge with regard to data: collecting detailed information on the characteristics of those who are stopped and the reasons for conducting searches. This enterprise is fraught with the potential for error (and possibly subversion) and will be difficult to implement nationally. The demand for these kinds of alternative measures reflects a point we first raised in the introduction: Policing does not have one bottom line—it has many. The demand for these new data collection efforts reflects a concern for something other than the war on crime and drugs. It reflects a growing concern for equity and fairness on the part of the police. Once again, data collection will play a central, if challenging, role in this issue.

The demand for these kinds of alternative measures reflects [that] policing does not have one bottom line—it has many.

Measurement

We have discussed at length the difference between data collection and measurement. In addition, we demonstrated how relatively simple constructs might

The demand for these new data collection efforts reflects a concern for something other than the war on crime and drugs. It reflects a growing concern for equity and fairness on the part of the police.

not measure what they appear to measure. Measurement error is probably a significant problem in police organizational research but we cannot know for sure because researchers have, by and large, ignored it. Few police researchers have systematically accounted for measurement error in their data. Maguire (forthcoming) and Weiss (1997) both used structural equation modeling techniques to produce reliable measures of organizational variables. Weiss went a step further, using a "key informant" approach to the administration of his survey. He distributed his survey to multiple respondents within each agency as a means of reducing the error associated with any single individual's response. This kind of comprehensive and sophisticated measurement strategy sets a benchmark to which organizational researchers should aspire.

Explanation

The methodologies used in the comparative study of police organizations have improved over the past three decades. Yet, many of the studies we examined are flawed in both theory and method. If we had to identify the single most serious problem in the entire line of research, our choice would undoubtedly be the failure to consistently root empirical studies in theory. Some of the studies with the worst methodological flaws contained flawless reviews of the relevant literature and theory. The indiscriminate use of exploratory factor analysis with little attention to theory is common in much of the research. Judging from this literature alone, it appears that a crucial point in the research process that many people either ignore or find difficult is the translation of a theoretical model into an empirical one. Our goal here is not to denigrate past researchers, but to point out some of the flaws in the research in the hope of deterring future researchers from making the same mistakes. Nearly every study contributes at least one new insight to the literature.

A modest vision for future research

Sometimes it seems that empirical research on police organizations is a lot like making minestrone soup: in the absence of a good recipe (theory), find whatever vegetables that happen to be convenient (the data), toss them into the pot (the model), cook it (execute the statistical program), and see if it tastes good (check the R^2). Continue to make adjustments (capitalizing on statistical chance) to the soup until you like the way it tastes.

Our vision for the future of police organizational research is simple. Begin by explicating a reasonable theory, translate the theory into an empirical model, collect reasonably good data that are useful for testing the theory, turn those raw data points into theoretically meaningful and reliable measures, and then test a model that posits a causal order among the measures. Do not capitalize on statistical chance by endlessly tinkering with the model if it does not fit. If this is the case, return to step one and modify the theory. Recent advances in statistical modeling techniques and software packages that implement these techniques make it easy for most social scientists to become skilled and careful theory testers. This is our “recipe” for achieving incremental progress in the study of police organizations.

Policy

Police executives and policymakers are concerned with the day-to-day realities of their worlds. They want answers about what works and why. They want measures that assist them in making decisions and policy. They want explanations for why things happened. In the academic world of theories, data, and publishing, researchers want precision and statistically significant findings. They want analysis driven by theory. Coming to grips with both of these worlds is difficult but not insurmountable. The policy implications that derive from theory and analysis need to be made explicit by researchers. From our experience, we have found that police executives and policymakers want good measures and explanations, and they want them in ways that are more understandable. They want direct answers to questions such as “How does my department compare with others in terms of community policing or officer performance? Are we on the right track? What should we do that works?” As researchers, at least one of our jobs is to assist policymakers in answering these types of questions. Balancing all of these competing interests—by using adequate theory, collecting good data, formulating accurate measures, developing sound explanations, and describing clearly the implications for policy—is no small task. Pulling all of these separate pieces together is a worthwhile challenge.

Conclusion

This chapter is meant to serve as a resource for scholars, students, policymakers, and others interested in the evolution of research on American police organizations. We have tried to escort readers on a journey from the elementary beginnings of data collection to the birth of scholarly theories; from the development of simplistic measures based on quantitative data to the sophisticated research that is now taking place. Along the way there have been many pitfalls: insufficient attention to conceptualization and theory, unrealistic measures, inadequate

statistical methods, and an overall lack of appreciation for previous research. Looking back on the classics in the field provides a sense of foundation, perhaps with a touch of nostalgia, but countless avenues for refinement and rediscovery remain. Chief among these are two responsibilities that may seem at first like strange bedfellows: doing quality research that (1) is based firmly in new or existing theories, and (2) contributes to the understanding or practice of policing. By tracing the evolution of research on police organizations from past to present, bumps and all, we hope this chapter provides a clear road map of what is to come. As we approach the 21st century, much remains to be learned.

We are grateful to Kimberly Hassell, William King, Robert Langworthy, Steve Mastrofski, Hank Robinson, and the members of the editorial board for their comments on earlier drafts.

Notes

1. This line of research also grew out of the political science tradition of measuring and explaining variations in local government policies and structures (Meyer and Baker 1979; Wilson 1968a, 1968b).
2. Several levels of analysis are commonly used within organization studies. The level of analysis in this essay is called the "organization set." The defining characteristic of the organization set "is that it views the environment from the perspective or standpoint of a specific (focal) organization" (Scott 1992, 126). This limits the scope of the essay to a particular (though broad) analytical framework. Many studies of police organizations are implicitly based on a different level of analysis, including Ostrom, Parks, and Whitaker's (1978a) study of metropolitan areas and Bayley's (1985, 1992) study of nations.
3. This definition excludes agencies that are specialized by function (e.g., fish and wildlife police) or territory (e.g., park or airport police), including most Federal law enforcement agencies, many county sheriffs' departments and State highway patrol agencies, and private security firms. Although using such a restrictive definition reduces the level of variation among the organizations under study, it defines a set of core tasks and functions that all the organizations presumably share.
4. There are also numerous similarities among police organizations. As Wadman (1998) points out, all of the largest municipal police agencies have hierarchical rank structures (though some may be flatter than others); they all have divisions for patrol, investigations, and administration; and they all devote a disproportionate share of their resources to motorized patrol.
5. Although there is some overlap, the style of a police organization is conceptually different from the style of an individual officer (Talarico and Swanson 1979; Wilson 1968b).

6. We are careful to distinguish between what an organization does and what it is by the location of the activity, behavior, or program, rather than its degree of visibility to the public. Much of what the police do *externally* occurs in low visibility settings (Goldstein 1960). On the other hand, the *internal* features of organizations (such as their structures) are often designed to serve as signals to external constituents that the organization is doing the right things (Meyer 1979).
7. For example, the Supplementary Homicide Report collects data on the age, sex, race, and ethnicity of the victim and the offender; the weapon used; and the circumstances of the offense.
8. A number of national and local media, including the Associated Press, *Washington Post*, *New York Times*, *Los Angeles Times*, and *USA Today*, reported the 7-year decline in crime on May 17, 1999, following the release of preliminary data by the FBI.
9. The hierarchy rule limits the reporting of multiple offenses committed within the course of a single criminal incident to the single most serious offense.
10. The COPS datasets contain a number of biases. They are completed by agencies that apply for or receive community policing grants from COPS, which provides respondents with an incentive to exaggerate their involvement in community policing. All of the datasets are based on convenience samples and contain a number of technical problems that limit their utility. They were not collected by social scientists and were not designed for social scientific analysis. Their purpose is to encourage local agencies to think about their community policing efforts and to communicate their progress to Federal authorities. On the other hand, they contain some of the largest sample sizes in police research (Maguire and Mastrofski forthcoming).
11. Misdemeanor arrests rose from 133,446 in 1993 to 205,277 in 1996, but misdemeanor complaints during the same period rose only slightly (Harcourt 1998, 340).
12. Bruce Smith (1949, 282) reported that the Uniform Crime Reports (UCR) occasionally contained "special treatments of material relating to the operation of police forces, the number of police employees, their distribution among the various phases of police work, the types of major police equipment in use, and the method of their employment." These special administrative surveys must have been fielded between the establishment of UCR in 1930 and the publication of Smith's text in 1949.
13. We selected cities between 25,000 and 100,000, just as Wilson did; however, because of the growth of cities over the past three decades, our sample size was much larger ($n=1,633$).
14. Chi-square tests of model fit suggest that the specified model has less than 1 chance in 1,000 of being the correct model (one that is able to closely reproduce the sample data). This finding held through various modifications to the model, including the use of different estimators (maximum likelihood and asymptotic distribution free).

15. A similar issue arises in mortality studies. Studying only the dead to learn about causes and correlates of death is a flawed strategy because we cannot know whether these same conditions might be present in people who lived (Kaufman 1976; King, Travis, and Langworthy 1997).

16. These same themes are also discussed in an earlier article by Bordua and Reiss (1966).

17. In one example, Reiss and Bordua (1967) discuss two environmental variables that are important to the organization: the security of the police chief's tenure and the degree of accountability that the government executive demands from the chief. Cross-classifying these two variables, Reiss and Bordua formed a crude taxonomy of four department types that might reflect variation in political interference into police department affairs. They suggested that these and other environmental variables were important because they "structure the effective range of command and control" (p. 49) in municipal police departments.

18. Slovak (1986, 5) laments that "there is a very real sense in which the promise offered by Wilson's original analysis has gone unfulfilled."

19. Aggregate arrest rates are sometimes used as an indicator of police agency style.

20. In general, structural contingency theory suggests that no single organizational form is ideal for all circumstances (Donaldson 1995; Lawrence and Lorsch 1967). Successful organizations survive by adapting to the contingencies of their specific tasks and environments.

21. We do not include the subjective studies here since they do not really represent organizational variables. They are, nonetheless, important.

22. Ostrom and Parks (1973) found curvilinear relationships between city size and citizen ratings of police performance. For central cities, performance ratings increased as city size approached 100,000 residents, after which ratings decreased; the same relationship was found for suburbs, but the population threshold was only 20,000 residents. The study was criticized for failing to control for "social or economic differences between respondents and their neighborhoods" (Whitaker 1983, 187). Whitaker concludes that the size of a police organization is more important than the size of a political jurisdiction, thus lending support to reform strategies that seek to simulate the feel of small-town policing in large cities through the use of precinct stations, substations, and other strategies. Whitaker's chapter is the most comprehensive (though dated) review of the effects of police agency size.

23. There is a shortage of theory to explain either of these consistent findings. Region size may simply be a proxy for any number of political, historical, economic, or demographic differences between regions. Organizational size seems to affect nearly every aspect of what organizations do. One possible reason that larger police agencies may

report engaging in more community policing activities is simply that they have more employees to assign to such functions.

24. For instance, Clark, Hall, and Hutchinson (1967) treat interorganizational relationships and network properties as contextual rather than organizational variables in their study of police performance.

25. According to Duffee (1990), this problem applies to all sectors of criminal justice. His advice to criminal justice scholars is particularly appropriate—we should describe and explain what criminal justice organizations *do* rather than what they *should be doing*.

26. Unfortunately, Davenport's (1996) study misses one of the most consistent findings in research on structure: the importance of organizational size. Davenport treated size as an organizational rather than a contextual variable. This decision has some precedent in the literature and is not problematic. The problem is that he failed to control for the effects of size on other organizational variables. This is not intended as a blanket indictment of Davenport's study because it contains several features from which students of police organizations can learn.

27. Conventional wisdom in policing suggests that police organizations do not “go out of business” (Travis and Brann 1997). Recent work by William King and his colleagues (King forthcoming; King, Travis, and Langworthy 1997) challenges this assumption. Based on a survey of county sheriffs in Ohio, King documented the death of 104 police agencies (and the birth of an additional 15). He is now replicating this study in several other States.

28. Weick's (1969) discussion is inherently social psychological, but Manning (1992) and Ericson and Haggerty (1997) span levels from the individual to the institution.

29. We are grateful to Paula Kautt at the University of Texas–San Antonio for pointing this out.

References

Alpert, Geoffrey, and Mark H. Moore. 1993. Measuring police performance in the new paradigm of policing. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Annan, Sampson. 1994. Community policing strategies: A comprehensive analysis. Methodology report submitted by the Police Foundation to the U.S. Department of Justice, National Institute of Justice, Washington, D.C.

Aupperle, Kenneth E., William Acar, and David E. Booth. 1986. An empirical critique of *In search of excellence: How excellent are the excellent companies?* *Journal of Management* 12 (4): 499–512.

- Bayley, David H. 1994. *Police for the future*. New York: Oxford University Press.
- . 1992. Comparative organization of the police in English-speaking countries. In *Modern policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.
- . 1985. *Patterns of policing*. New Brunswick, New Jersey: Rutgers University Press.
- Beattie, Ronald. 1960. Criminal statistics in the United States. *Journal of Criminal Law, Criminology, and Police Science* 51:49–51.
- Berk, Richard A. 1974. Performance measures: Half full or half empty? *Social Science Quarterly* 54:762–764.
- Beyer, William C., and Helen C. Toerring. 1929. The policeman's hire. *Annals of the American Academy of Political and Social Sciences* 146 (November): 135–146.
- Blau, Peter M., Wolfe V. Heyderbrand, and Robert E. Stauffer. 1966. The structure of small bureaucracies. *American Sociological Review* 31:179–191.
- Blau, Peter M., and Richard A. Schoenherr. 1971. *The structure of organizations*. New York: Basic Books.
- Bordua, David J., and Albert J. Reiss, Jr. 1966. Command, control, and charisma: Reflections on police bureaucracy. *American Journal of Sociology* 72:68–76.
- Bratton, William J. 1999. Great expectations: How higher expectations for police departments can lead to a decrease in crime. In *Measuring what matters: Proceedings from the Policing Research Institute meetings*, edited by Robert H. Langworthy. Research Report, NCJ 170610. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and the Office of Community Oriented Policing Services.
- Bratton, William J., with Peter Knobler. 1998. *Turnaround: How America's top cop reversed the crime epidemic*. New York: Random House.
- Burns, Tom, and George M. Stalker. 1961. *The management of innovation*. London: Tavistock.
- Churchill, Winston. 1944. Speech at Harvard University, 6 September 1943. In *Onwards to victory*. Boston: Little, Brown and Company.
- Clark, John P., Richard H. Hall, and Bruce Hutchinson. 1967. Interorganizational relationships and network properties as contextual variables in the study of police performance. In *Police and society*, edited by David H. Bayley. Beverly Hills: Sage Publications.

Coleman, James S. 1974. *Power and the structure of society*. New York: W.W. Norton and Company.

Conyers, John, Jr. 1999. Representative Conyers of Michigan introducing the Traffic Stops Statistics Act of 1999, H.R. 1443, 106th Cong., 1st sess. *Congressional Record*, 15 April, vol. 145, no. 52.

Cordner, Gary W. 1989. Police agency size and investigative effectiveness. *Journal of Criminal Justice* 17:145–155.

Count-van Manen, Gloria. 1977. *Crime and suicide in the Nation's capital: Toward macro-historical perspectives*. New York: Praeger.

Coyle, K.R., J.C. Schaaf, and J.R. Coldren, Jr. 1991. *Futures in crime analysis: Exploring applications of incident-based crime data*. NCJ 127201. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Crank, John P. 1994. Watchman and community: Myth and institutionalization in policing. *Law & Society Review* 28 (2): 325–351.

———. 1992. Police style and legally serious crime: A contextual analysis of municipal police departments. *Journal of Criminal Justice* 20:401–412.

———. 1990. The influence of environmental and organizational factors on police style in urban and rural environments. *Journal of Research in Crime and Delinquency* 27 (2): 166–189.

———. 1989. Civilianization in small and medium police departments in Illinois, 1973–1986. *Journal of Criminal Justice* 17 (3): 167–178.

Crank, John P., and Robert H. Langworthy. 1996. Fragmented centralization and the organization of the police. *Policing and Society* 6:213–229.

———. 1992. An institutional perspective of policing. *Journal of Criminal Law and Criminology* 83 (2): 338–363.

Crank, John P., and L. Edward Wells. 1991. The effects of size and urbanism on structure among Illinois police departments. *Justice Quarterly* 8 (2): 170–185.

Cuniff, Mark. 1983. *Beyond crime: Law enforcement operational and cost data*. NCJ 92481. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Davenport, Douglas R. 1996. Public agency performance and structure: Assessing the effects of the organizational environment. Ph.D. diss., Texas Tech University.

Davis, Kenneth C. 1969. *Discretionary justice: A preliminary inquiry*. Baton Rouge: Louisiana State University Press.

- Decker, Scott. 1981. Alternate measures of police output. *American Journal of Police* 1 (1): 23–35.
- . 1977. The evolution of crime statistics as a police problem. *Journal of Police Science and Administration* 6:67–73.
- DiIulio, John J., Jr. 1993. Measuring performance when there is no bottom line. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Donaldson, Lex. 1995. *American anti-management theories of organization: A critique of paradigm proliferation*. Cambridge: Cambridge University Press.
- Duffee, David. 1990. *Explaining criminal justice*. 2d ed. Prospect Heights, Illinois: Waveland Press.
- Duncan, Robert B. 1972. Characteristics of organizational environments and perceived environmental uncertainty. *Administrative Science Quarterly* 17:313–327.
- Eck, John, and Edward R. Maguire. Forthcoming. Have changes in policing reduced violent crime?: An assessment of the evidence. In *The crime drop in America*, edited by Alfred Blumstein and Joel Wallman. New York: Cambridge University Press.
- Eisenberg, Terry, and Deborah Ann Kent. 1973. *Police personnel practices in State and local governments*. Washington, D.C.: International Association of Chiefs of Police and the Police Foundation.
- Eisenstadt, S.N. 1959. Bureaucracy, bureaucratization, and debureaucratization. *Administrative Science Quarterly* 4 (3): 302–320.
- Emery, Fred, and Eric L. Trist. 1965. The causal texture of organizational environments. *Human Relations* (February): 21–32.
- Ericson, Richard B., and Kevin D. Haggerty. 1997. *Policing the risk society*. Toronto: University of Toronto Press.
- Ferdinand, Theodore N. 1976. From a service to a legalistic style police department: A case study. *Journal of Police Science and Administration* 4 (3): 302–319.
- . 1972. Politics, the police, and arresting policies in Salem, Massachusetts since the Civil War. *Social Problems* 19:572–588.
- . 1967. The criminal patterns of Boston since 1869. *American Journal of Sociology* 73:688–698.
- Gianakis, Gerasimos A., and G. John Davis III. 1998. Reinventing or repackaging public services? The case of community-oriented policing. *Public Administration Review* 58 (6).

Goldstein, Herman. 1977. *Policing a free society*. Cambridge, Massachusetts: Ballinger.

———. 1963. Police discretion: The ideal versus the real. *Public Administration Review* 23:140–148.

Goldstein, Joseph. 1960. Police discretion not to invoke the criminal process: Low visibility decisions in the administration of justice. *Yale Law Journal* 69:543–594.

Guyot, Dorothy. 1977. Police departments under social science scrutiny. *Journal of Criminal Justice* 5:105–118.

Hammer, Michael, and James Champy. 1993. *Reengineering the corporation*. New York: HarperCollins Publishers.

Harcourt, Bernard E. 1998. Reflecting on the subject: A critique of the social influence conception of deterrence, the broken-windows theory, and order maintenance policing New York style. *Michigan Law Review* 97:291–389.

Hayduk, Leslie A. 1987. *Structural equation modeling with LISREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.

Heininger, B.L., and J. Urbanek. 1983. Civilianization of the American police: 1970–1980. *Journal of Police Science and Administration* 11 (2): 200–205.

Henderson, Thomas A. 1975. The relative effects of community complexity and of sheriffs upon the professionalism of sheriff departments. *American Journal of Political Science* 19 (1): 107–132.

Hough, Michael, and Nick Tilley. 1998. *Auditing crime and disorder: Guidance for local partnerships*. Briefing Note. London: Home Office Police Research Group.

Huff, C. Ronald. 1987. Organizational structure and innovation in urban police departments. *Public Administration Review* 47 (6): 508–509.

Katz, Charles M. 1997. Police and gangs: A study of a police gang unit. Ph.D. diss., University of Nebraska at Omaha.

Katz, Charles, Edward R. Maguire, and Dennis W. Roncek. 2000. The creation of specialized police gang units: Testing contingency, social threat, and resource dependency explanations. Unpublished manuscript, University of Nebraska at Omaha.

Kaufman, Herbert. 1976. *Are government organizations immortal?* Washington, D.C.: Brookings Institution.

King, William R. Forthcoming. Time, constancy, and change in American municipal police organizations. *Police Quarterly*.

———. 1999. Personal communications with Edward R. Maguire, March–May.

- . 1998. Innovativeness in American municipal police organizations. Ph.D. diss., University of Cincinnati.
- King, William R., Lawrence F. Travis III, and Robert H. Langworthy. 1997. Police organizational death. Paper presented at the 1997 Annual Meeting of the American Society of Criminology, 22 November, San Diego.
- Kitsuse, John, and Aaron Victor Cicourel. 1963. A note on the uses of official statistics. *Social Problems* 11:131–139.
- Klockars, Carl B. 1999. Some really cheap ways of measuring what really matters. In *Measuring what matters: Proceedings from the Policing Research Institute meetings*, edited by Robert H. Langworthy. Research Report, NCJ 170610. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and the Office of Community Oriented Policing Services.
- LaFave, William R. 1965. *Arrest: The decision to take a suspect into custody*. Boston: Little, Brown and Company.
- Lane, Roger. 1980. Urban police and crime in nineteenth-century America. In *Crime and justice: An annual review of research*, edited by Norval Morris and Michael Tonry. Vol. 2. Chicago: University of Chicago Press.
- Langworthy, Robert H. 1992. Organizational structure. In *What works in policing?*, edited by Gary W. Cordner and Donna C. Hale. Cincinnati: Anderson Publishing Company.
- . 1986. *The structure of police organizations*. New York: Praeger.
- . 1985a. Police department size and agency structure. *Journal of Criminal Justice* 13:15–27.
- . 1985b. Wilson's theory of police behavior: A replication of the constraint theory. *Justice Quarterly* 2 (1): 89–98.
- . 1983. The effects of police agency size on the use of police employees: A re-examination of Ostrom, Parks, and Whitaker. *Police Studies* 5:11–19.
- , ed. 1999. *Measuring what matters: Proceedings from the Policing Research Institute meetings*. Research Report, NCJ 170610. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and the Office of Community Oriented Policing Services.
- Lawrence, Paul R., and Jay W. Lorsch. 1967. Differentiation and integration in complex organizations. *Administrative Science Quarterly* 12 (1): 1–47.
- Leigh, Adrian, Gary Mundy, and Rachel Tuffin. 1999. *Best value policing: Making preparations*. Police Research Series paper 116. London: Home Office Police Research Group.

Liska, A.E., and M.B. Chamlin. 1984. Social structure and crime control among macrosocial units. *American Journal of Sociology* 90 (2): 383–395.

Macro International. 1997. *Community policing: 1997 national survey update of police and sheriffs' departments*. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Maguire, Edward R. Forthcoming. *Context, complexity, and control: Organizational structure in American police agencies*. Albany: State University of New York Press.

———. 1997. Structural change in large municipal police organizations during the community policing era. *Justice Quarterly* 14 (3): 701–730.

Maguire, Edward R., and Charles M. Katz. 2000. Loose coupling, sensemaking, and community policing in American police agencies. Unpublished manuscript, University of Nebraska at Omaha.

Maguire, Edward R., Joseph B. Kuhns, Craig D. Uchida, and Stephen M. Cox. 2000. Measuring community policing at the agency level. Unpublished manuscript, University of Nebraska at Omaha.

———. 1997. Patterns of community policing in nonurban America. *Journal of Research in Crime and Delinquency* 34:368–394.

Maguire, Edward R., and Stephen D. Mastrofski. Forthcoming. Patterns of community policing in the United States. *Police Quarterly*.

Maguire, Edward R., Jeffrey B. Snipes, Craig D. Uchida, and Margaret Townsend. 1998. Counting cops: Estimating the number of police departments and police officers in the USA. *Policing: An International Journal of Police Strategies & Management* 21 (1): 97–120.

Maguire, Edward R., and Craig D. Uchida. 1998. Measuring community policing at the precinct and agency levels. Paper presented at the 1998 Annual Meeting of the American Society of Criminology, 14 November, Washington, D.C.

Maguire, Edward R., and Jihong Zhao. 1999. Structural change in large municipal police organizations. Paper presented at the Annual Conference on Criminal Justice Research and Evaluation, 20 July, Washington, D.C.

Maguire, Edward R., Jihong Zhao, and Nicholas Lovrich. 1999. Dimensions of community policing. Unpublished manuscript, University of Nebraska at Omaha.

Maltz, Michael. 1977. Crime statistics: A historical perspective. *Crime & Delinquency* 23:32–40.

- Manning, Peter K. 1992. Information technologies and the police. In *Modern policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.
- . 1988. Review of *The structure of police organizations* by Robert Langworthy. *Administrative Science Quarterly* (June): 323–327.
- Marvell, Thomas B., and Carlisle E. Moody. 1996. Specification problems, police levels, and crime rates. *Criminology* 34 (4): 609–646.
- Mastrofski, Stephen D. 1999. *Policing for people*. Ideas in American Policing Series, March. Washington, D.C.: Police Foundation.
- . 1998a. Community policing and police organization structure. In *How to recognize good policing: Problems and issues*, edited by Jean-Paul Brodeur. Newbury Park, California: Sage Publications.
- . 1998b. *Patterns of community policing: Tracking organizational change undertaken with COPS funding, phase two*. Washington, D.C.: U.S. Department of Justice, Office of Community Oriented Policing Services.
- Mastrofski, Stephen D., and R. Richard Ritti. Forthcoming. Making sense of community policing: A theory-based analysis. *Police Practice and Research: An International Journal*.
- . 1996. Police training and the effects of organization on drunk driving enforcement. *Justice Quarterly* 13 (2): 291–320.
- Mastrofski, Stephen D., R. Richard Ritti, and Debra Hoffmaster. 1987. Organizational determinants of police discretion: The case of drinking-driving. *Journal of Criminal Justice* 15:387–402.
- Mastrofski, Stephen D., and Craig D. Uchida. 1993. Transforming the police. *Journal of Research in Crime and Delinquency* 30 (3): 330–358.
- McEwen, Tom. 1996. *National data collection on police use of force*. NCJ 160113. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics and National Institute of Justice.
- Meagher, M.S. 1985. Police patrol styles: How pervasive is community variation? *Journal of Police Science and Administration* 13 (1): 36–45.
- Meyer, Fred A., Jr., and Ralph Baker, eds. 1979. *Determinants of law-enforcement policies*. Lexington, Massachusetts: Lexington Books.
- Meyer, John, and Brian Rowan. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology* 83 (2): 340–363.

Meyer, Marshal W. 1979. Organizational structure as signaling. *Pacific Sociological Review* 22:481–500.

Miller, Wilbur R. 1977. *Cops and bobbies: Police authority in New York and London, 1830–1870*. Chicago: University of Chicago Press.

Monkkonen, Eric H. 1992. History of urban police. In *Modern policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and justice: A review of research*. Chicago: University of Chicago Press.

———. 1981. *Police in urban America: 1860–1920*. Cambridge: Cambridge University Press.

Moore, Mark H., and Margaret Poethig. 1999. The police as an agency of municipal government: Implications for measuring police effectiveness. In *Measuring what matters: Proceedings from the Policing Research Institute meetings*, edited by Robert H. Langworthy. Research Report, NCJ 170610. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and the Office of Community Oriented Policing Services.

Morgan, David R., and Cheryl Swanson. 1976. Analyzing police policies: The impact of environment, politics, and crime. *Urban Affairs Quarterly* 11 (4): 489–510.

Murphy, Christopher J. 1986. The social and formal organization of small-town policing: A comparative analysis of RCMP and municipal policing. Ph.D. diss., University of Toronto.

Murphy, Patrick V., and Thomas Plate. 1977. *Commissioner: A view from the top of American law enforcement*. New York: Simon & Schuster.

Nanus, B., and L. Perry. 1973. A planning oriented measure of crime and delinquency. *Journal of Criminal Justice* 23 (1): 32–40.

Osborne, David, and Ted Gaebler. 1992. *Reinventing government*. New York: Addison-Wesley.

Ostrom, Elinor. 1973. On the meaning and measurement of output and efficiency in the provision of urban police services. *Journal of Criminal Justice* 1:93–112.

Ostrom, Elinor, and Roger B. Parks. 1973. Suburban police departments: Too many and too small? In *The organization of the suburbs*, edited by Louis H. Masotti and Jeffrey K. Hadden. Urban Affairs Annual Reviews Series, vol. 7. Beverly Hills: Sage Publications.

Ostrom, Elinor, Roger B. Parks, and Gordon P. Whitaker. 1978a. *Patterns of metropolitan policing*. Cambridge, Massachusetts: Ballinger.

———. 1978b. Police agency size: Some evidence on its effects. *Police Studies* (March): 34–46.

- Parks, Roger B. 1999. Measuring community policing. Paper presented at the 1999 Annual Meeting of the American Society of Criminology, 18 November, Toronto.
- . 1984. Linking objective and subjective measures of performance. *Public Administration Review* (March/April): 118–127.
- Perrow, Charles. 1986. *Complex organizations: A critical essay*. 3d ed. Glenview, Illinois: Scott, Foresman and Company.
- Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *The external control of organizations*. New York: Harper & Row.
- President's Commission on Law Enforcement and Administration of Justice. 1967. *Task Force report: The police*. Washington, D.C.: U.S. Government Printing Office.
- Ragin, Charles C. 1987. *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.
- Reaves, Brian A., and Andrew L. Goldberg. 1999. *Law enforcement management and administrative statistics, 1997: Data for individual State and local agencies with 100 or more officers*. NCJ 171681. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Reiss, Albert J., Jr. 1971. *The police and the public*. New Haven: Yale University Press.
- Reiss, Albert J., Jr., and David J. Bordua. 1967. Environment and organization: A perspective on the police. In *The police: Six sociological essays*, edited by David J. Bordua. New York: John Wiley & Sons.
- Riksheim, Eric, and Steven Chermak. 1993. Causes of police behavior revisited. *Journal of Criminal Justice* 21 (4): 353–382.
- Roberts, David J., and Seth F. Jacobs. 1994. *Demonstrating the operational utility of incident-based data for local crime analysis: Reporting systems in Tacoma, Washington, and New Bedford, Massachusetts*. NCJ 145860. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Robison, Susan. 1966. A critical view of the Uniform Crime Reports. *Michigan Law Review* 64:1031–1054.
- Roth, Jeffrey A., and Calvin C. Johnson. 1997. COPS context and community policing. Paper presented at the 1997 Annual Meeting of the American Society of Criminology, November, San Diego.
- Schumacker, Randall E., and Richard G. Lomax. 1996. *A beginner's guide to structural equation modeling*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Scott, W. Richard. 1992. *Organizations: Rational, natural, and open systems*. 3d ed. Englewood Cliffs, New Jersey: Prentice Hall.

Seidman, David, and M. Couzens. 1974. Getting the crime rate down: Political pressure and crime reporting. *Law & Society Review* 8:457–493.

Selznick, Philip. 1957. *Leadership in administration*. New York: Harper & Row.

Sherman, Lawrence W. 1998. *Evidence-based policing*. Ideas in American Policing Series, July. Washington, D.C.: Police Foundation.

———. 1980. Causes of police behavior: The current state of quantitative research. *Journal of Research in Crime and Delinquency* 17:69–100.

Skoler, D.L., and J.M. Hetler. 1969. Criminal administration and the local government crisis—The challenge of consolidation. *Prosecutor* 5 (4).

Skolnick, Jerome. 1966. *Justice without trial: Law enforcement in a democratic society*. New York: John Wiley & Sons.

Slovak, Jeffrey S. 1986. *Styles of urban policing: Organization, environment, and police styles in selected American cities*. New York: New York University Press.

Smith, Bruce, Jr. 1949. *Police systems in the United States*. 2d rev. ed. New York: Harper & Row.

Steinman, M. 1984. Rationalizing police operations: Some explanatory factors. *Journal of Criminal Justice* 12 (3): 221–233.

Stinchcombe, Arthur L. 1963. Institutions of privacy in the determination of police administrative practice. *American Journal of Sociology* 69:150–160.

Surette, Raymond. 1984. A note on the historical relationship between the economy and police organizational activities. *Criminal Justice Review* 9 (2): 47–52.

Swanson, Cheryl. 1978. The influence of organization and environment on arrest practices in major U.S. cities. *Policy Studies Journal* 7:390–398.

Talarico, Suzette M., and Charles R. Swanson. 1979. Styles of policing: An exploration of compatibility and conflict. In *Determinants of law-enforcement policies*, edited by Fred A. Meyer, Jr., and Ralph Baker. Lexington, Massachusetts: Lexington Books.

Travis, Jeremy, and Joseph E. Brann. 1997. Introduction to *Measuring what matters, part two: Developing measures of what the police do*. Research in Action, NCJ 167255. Washington, D.C.: U.S. Department of Justice, National Institute of Justice and the Office of Community Oriented Policing Services.

- Trojanowicz, Robert. 1994. *Community policing: A survey of police departments in the United States*. Washington, D.C.: U.S. Department of Justice, Federal Bureau of Investigation.
- Uchida, Craig D. 1997. The development of the American police: An historical overview. In *Critical issues in policing: Contemporary readings*. 3d ed. Edited by Roger G. Dunham and Geoffrey P. Alpert. Prospect Heights, Illinois: Waveland Press.
- . 1986. Law enforcement management and administrative statistics: Research and development, final report. Manuscript prepared for the Institute of Criminal Justice and Criminology, University of Maryland.
- Vollmer, August. 1927. Presidential Address. In *Proceedings of the annual meeting of the International Association of Chiefs of Police*. Cited in Charles F. Wellford, Redesigning the Uniform Crime Reports, *American Journal of Police* 1 (2) (1982): 76–93.
- Wadman, Robert C. 1998. Organizing for the prevention of crime. Ph.D. diss., Idaho State University.
- Walker, Monica A. 1992. Do we need a clear-up rate? *Policing and Society* 2:293–306.
- Walker, Samuel. 1977. *A critical history of police reform: The emergence of professionalism*. Lexington, Massachusetts: Lexington Books.
- Walzer, C. 1970. Economies of scale and municipal police services. Ph.D. diss., University of Illinois.
- Weick, Karl E. 1976. Educational organizations as loosely coupled systems. *Administrative Science Quarterly* 21:1–19.
- . 1969. *The social psychology of organizing*. Reading, Massachusetts: Addison-Wesley.
- Weiss, Alexander. 1997. The communication of innovation in American policing. *Policing: An International Journal of Police Strategies and Management* 20 (2): 292–310.
- Wellford, Charles F. 1982. Redesigning the Uniform Crime Reports. *American Journal of Police* 1 (2): 76–93.
- . 1974. Crime and the police: A multivariate analysis. *Criminology* 12:195–213.
- Whitaker, Gordon P. 1983. Police department size and the quality and cost of police services. In *The political science of criminal justice*, edited by Stuart Nagel, Erika Fairchild, and Anthony Champagne. Springfield, Illinois: Charles C. Thomas.

Wilson, James Q. 1993. The problem of defining agency success. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

———. 1989. *Bureaucracy: What government agencies do and why they do it*. New York: Basic Books.

———, ed. 1968a. *City politics and public policy*. New York: John Wiley & Sons.

———. 1968b. *Varieties of police behavior: The management of law and order in eight communities*. Cambridge: Harvard University Press.

———. 1963. The police and their problems: A theory. *Public Policy* 12:189–216.

Wilson, J.Q., and B. Boland. 1979. *Effect of the police on crime*. Criminal Justice Perspectives, NCJ 58531. Washington, D.C.: U.S. Department of Justice, Law Enforcement Assistance Administration.

Wolfgang, Marvin. 1963. The Uniform Crime Reports: A critical appraisal. *University of Pennsylvania Law Review* 111:708–738.

Wycoff, Mary Ann. 1994. *Community policing strategies. Final report*. NCJ 159603. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Zhao, Jihong. 1996. *Why police organizations change: A study of community oriented policing*. Washington, D.C.: Police Executive Research Forum.

Standards and Measures of Court Performance

by Ingo Keilitz

The Trial Court Performance Standards (TCPS), a comprehensive organizational court performance system, were 10 years in the making. TCPS significantly advanced the scope of inquiry of performance measurement of the courts and other components of the justice system from one of conceptualization and identification of constructs, variables, and operational definition to one of critical review and evaluation of actual implementation. In contrast to numerous model conceptual approaches to performance measurement and exhortations promoting the idea of court performance measurement, TCPS come with specific directions for how it is to be done. Included are the elements of a complete organizational performance measurement system: (1) the abstract concepts or constructs of desired performance, (2) their concrete representations or variables, and (3) the operational definitions and procedures for measuring the variables.

Despite their widespread use by courts throughout the United States, TCPS are not well known outside the community of court practitioners. This essay reviews the constructs, variables, and operational definitions of TCPS; traces the history of their development; and explores their contribution to our knowledge about the performance measurement of courts and other components of the justice system.

Ingo Keilitz is President of Sherwood Consulting in Williamsburg, Virginia.

A
B
S
T
R
A
C
T

559

The terminology and logic of performance measurement—that of “inputs,” “outputs,” and “outcomes”—are then applied to the enterprise of court performance measurement itself, in particular the strategies for the acceptance, adoption, and use of TCPS in light of general resistance to organizational performance measures among court practitioners.

A
B
S
T
R
A
C
T

Courts, like other organizations funded by tax dollars, increasingly are held accountable for their performance. No longer content to prioritize services based on needs and demands, the public wants assurances of effective services at reasonable costs. The application of court performance standards and measures is a way to assess what the public gets for its money (Epstein 1988). The *Trial Court Performance Standards* (TCPS), almost 10 years in the making, were published in four volumes in July 1997 (Commission on Trial Court Performance Standards 1997a, 1997b, 1997c, 1997d). These standards represent a milestone in the development of concepts, techniques, and strategies to examine the performance of courts and the justice system. They are an example of public policy reform launched as an effort to establish measures for the goals of a branch of government.¹

TCPS provide a conceptual framework and the operational steps for identifying desired results, creating and tracking measurable indicators of progress toward those results, and assessing court system performance. They represent a new understanding—a new paradigm—of the role and responsibilities of courts. The *Trial Court Performance Standards With Commentary*, the first of the two major volumes, describes the standards (i.e., the broad performance goals or major areas of emphasis) toward which courts should strive and the role and responsibilities that these standards impose (Commission on Trial Court Performance Standards 1997d). The second major volume, *Trial Court Performance Standards and Measurement System Implementation Manual*, describes the performance measures and the methods by which courts can gauge their performance according to the standards (Commission on Trial Court Performance Standards 1997b).²

Despite their widespread use by courts throughout the United States and in several foreign countries, TCPS are not well known outside the community of court practitioners.³ The purpose of this essay is to stimulate critical review and experimentation directed not only at the substance of TCPS but also at their application by courts. The essay reviews the constructs, variables, and operational definitions of TCPS; traces the history of their development; and explores their contribution to our knowledge about the performance measurement of courts and other components of the justice system. The terminology and logic of performance measurement—that of “inputs,” “outputs,” and “outcomes”—are then applied to the enterprise of court performance measurement itself, in particular, the strategies for

The Trial Court Performance Standards represent a milestone in the development of concepts, techniques, and strategies to examine the performance of courts and the justice system.

the acceptance, adoption, and use of TCPS and the progress of public policy reform launched by TCPS in light of general resistance to organizational performance measures among court practitioners.

In their influential 1992 book, *Reinventing Government*, David Osborne and Ted Gaebler encouraged public organizations to measure results more than process, effectiveness more than efficiency, and broad policy outcomes more than processes or program outcomes. They cited a preliminary version of TCPS⁴ as a good example of appropriate performance measurement focused on results and outcomes (what ends courts actually accomplish with the means at their disposal) rather than on inputs (e.g., structures and processes). By focusing on results that matter to those served by the courts rather than those who run them, TCPS avoid the most common pitfalls of performance measurement in the public sector—a failure to identify important outcomes and a confusion of inputs (e.g., processes, resources, and structures) with outputs.

In contrast to numerous model conceptual approaches to performance measurement and exhortations promoting the *idea* of court performance measurement (see, for example, U.S. Department of Justice 1993), TCPS come with specific directions for how it is to be done. TCPS comprehensively describe issues that seem obvious in retrospect but that public managers have not, in the words of John J. DiIulio, Jr., addressed “with sufficient regularity or seriousness of purpose” (1993, 149). These issues are (1) what court performance ought to be, (2) what factors determine performance “outcomes,” (3) what results are desired and by whom, (4) how and by what means court performance should be measured on a day-to-day basis by court practitioners, and (5) what the goals of courts as public organizations are (i.e., the constructs that are the foundation of the variables and measures of performance).

The Conventional Wisdom About Court Performance: 1906–76

Beginning with Roscoe Pound’s speech on “the causes of popular dissatisfaction with the administrations of justice” to the American Bar Association in 1906 (Pound 1937) that “kindled the white flame of progress” in judicial administration (Wigmore 1937, 176), the “conventional wisdom” that the problems of courts are best addressed by innovations in their structure and their processes dominated reform for 70 years (Gallas 1976, 35–36). A causal link was simply assumed among structures, resources, and processes (inputs) and their immediate products, such as the number of cases heard and services provided (outputs), and court effectiveness and the well-being of those served by

the courts (outcomes). For example, the belief that court unification—the consolidation and simplification of court structure—has a direct impact on court performance was not seriously questioned until the late 1970s (see Henderson et al. 1984; Rottman and Hewitt 1996). This conventional wisdom, emphasizing structure and process over results and outcomes, is exemplified by the American Bar Association's *Standards Relating to Court Organization* (revised in 1990) and *Standards Relating to Trial Courts* (revised in 1992) (American Bar Association 1990, 1992).

Interest in performance measurement and courts as units of analysis gained momentum from empirical research in the late 1970s and early 1980s, much of it conducted by the National Center for State Courts in the area of the pace of litigation and causes of court delay. This research suggested that there was no one best way to guarantee desired results and that courts with similar formal rules, structures, and procedures can produce dramatically different outputs and outcomes (see Goerdts 1998). Meanwhile, the focus of research in courts expanded beyond relatively narrow topical emphases (e.g., the role of defense counsel, pretrial release, and plea bargaining) to a broader perspective of the courts as public organizations operating in the context of the justice system as a whole.⁵ The idea that courts must be managed like other public and private organizations to achieve desired results had gained a foothold in judicial administration.

In 1978, the National Institute of Law Enforcement and Criminal Justice (now the National Institute of Justice) of the U.S. Department of Justice commissioned a series of studies of performance measurement in the criminal justice system, including courts, police, prosecution, defense, adjudication, corrections, and the system as a whole (Cook et al. 1982; Whitaker et al. 1982; Jacoby 1982; Grizzle et al. 1982). This body of work stands as the immediate precursor to TCPS. One of the four volumes describing the results of these studies sought to develop a conceptual framework and methodology to be used in constructing performance measures for metropolitan adult felony courts (Cook et al. 1982, v). The authors surveyed three bodies of literature: research and analysis focusing on improving performance; research and theory defining performance; and applied research focusing on measuring performance. The literature survey was supplemented by interviews of court personnel and observations of court operations. “[W]e have not uncovered,” the authors concluded, “a well-articulated consensus concerning such key issues as: what performance is, how performance should be measured, what the proper goals of courts are, what factors determine court outputs, or the meaning of such terms as ‘due process,’ ‘efficiency,’ and ‘justice’” (Cook et al. 1982, 6). To a large extent, these issues were not resolved but simply framed as research issues that needed further attention.

Given the complexity of the task, it should not be surprising, at least in retrospect, that these early studies did not produce “a complete, intricately detailed blueprint for the construction of a performance measurement system,” but rather only a “conceptual framework that will guide future development of such a blueprint.” It would take another 5 years to begin, and 15 years to create, this “complete, intricately detailed” blueprint (Cook et al. 1982, 11). One commentator, assessing the status of court performance measurement in 1987 at the eve of the TCPS project, concluded that “[i]f the courts are to improve during the next thirty years everyone associated with them—managers, judges, researchers, and others—must radically increase the attention to evaluation of performance” (Feeny 1993, 477).

Overview of TCPS

TCPS include the elements of a complete organizational performance measurement system: (1) the abstract concepts or constructs of desired performance, (2) their concrete representations or variables, and (3) the operational definitions and procedures for measuring the variables. The elements are grouped in five performance areas that embrace the fundamental purpose or mission of courts:

- Access to Justice.
- Expedition and Timeliness.
- Equality, Fairness, and Integrity.
- Independence and Accountability.
- Public Trust and Confidence.

These areas suggest alternative ways of viewing the fundamental role and responsibilities of courts, such as providing and appearing to provide individual justice in individual cases; resolving disputes; upholding Federal and State constitutions; working independently of, but in cooperation with, other branches of government; promoting the rule of law; protecting individuals from the arbitrary use of government power; making a formal record of legal proceedings; and encouraging behavior that adheres to societal norms as expressed in statutes, ordinances, and regulations. Within each of the five performance areas, three to six standards describe goals (or major areas of emphasis) for court performance. Standards in two of the five performance areas—Expedition and Timeliness and Equality, Fairness, and Integrity—emphasize the courts’ fundamental dispute resolution functions. The standards in the three other performance areas focus on the functions of courts as organizations and their relations with other organizations and the public.

Theoretically, the five performance areas are accorded equal weight and importance. If effective participation and access to justice are denied, for example, it is of little consequence to those affected that a court *potentially* performs well in the other four areas. Justice delayed is justice denied, even if barriers to access are overcome. But expedition and timeliness are not everything. If the train is not heading where you want to go, it matters little how fast it is going. A court system that lacks public trust and confidence has diminished power and legitimacy.

In practice, the performance areas are systematically linked, and actions in one will affect another. If a minority group loses its trust and confidence in the court system and believes that it will not be dealt with fairly and equally, its effective participation and access to justice are limited. Court performance in different areas may conflict, requiring balance and choice by court managers and leaders. A public information campaign promoting a court's domestic violence program may increase access afforded women seeking civil protection orders, but it may also increase the court's caseload to a point that timeliness and expedition of cases are threatened.

For each performance area, succinct statements—"black letter" standards—describe the required performances. Each standard is followed by commentary to explain and clarify it. A total of 22 standards cover all 5 areas. Each standard is linked with a set of specific performance measures (i.e., variables, operational definitions, and procedures for measuring the variables) and data collection methods, techniques, and forms—a total of 68 measures for all 22 standards.⁶

TCPS encourage courts to conduct continuous self-assessment and improvement as part of routine court management, planning, and leadership. Each of the 68 measures is presented in easy-to-understand language—including a description of the measure's purpose and how it aligns with the standard and performance area, planning and preparations for taking the measure, data collection procedures and forms, data analysis and reporting, and references to other resources. This allows court practitioners to arrange the specific performance measurement with little or no professional research assistance. Some measures and their specific methods build on others and should be conducted in a particular sequence. Others stand alone and can be taken independently. Some measures, such as Measure 1.1.3, Audibility of Participants During Open Court Proceedings, are relatively easy to

TCPS encourage courts to conduct continuous self-assessment and improvement as part of routine court management, planning, and leadership.

Performance, as defined by TCPS, is oriented toward outcomes or results instead of level of effort or output. An “outcome” is a meaningful result, a condition of well-being, for the individuals, groups, or communities served by the courts.

conduct; others, such as Measure 3.3.3, Equality and Fairness in Sentencing, are more complex and time consuming.

The measurement system employs numerous data-gathering methods and taps diverse data sources. Included are familiar methods, such as court and case record reviews and tallies of case filings and dispositions, as well as other social science techniques less familiar to courts, such as systematic observation, simulations, surveys of various reference groups, and group techniques. Different “evaluators” and data collectors are prescribed depending on the object or subject of the measure. Trained volunteers, for example, are recommended for conducting structured observations of court proceedings and simulations of public access to court information, and court staff are recommended for taking measures involving record reviews. Some measures—such as measure 5.1.3,

General Public’s Perceptions of Court Performance—suggest the participation of professional consultants or court staff with expertise in data analysis or survey methodology. One of the four TCPS publications, *Planning Guide for Using the Trial Court Performance Standards and Measurement System*, discusses broader issues related to the measurement process framed by three questions: Who should oversee the measurement process? How should the measure or measurement process be adapted for a specific court? How should the results be used? (Commission on Trial Court Performance Standards 1997b.)

Performance, as defined by the TCPS areas of performance, standards, and measures, is oriented toward outcomes or results instead of level of effort or output. An “outcome” is a meaningful result, a condition of well-being, for the individuals, groups, or communities served by the courts, not merely an “output,” or efforts of the courts (e.g., number of cases heard) not tied to the needs of citizens and the overarching purposes of government. To state that TCPS are “oriented” toward important outcomes is to suggest a qualification. That is, although the focus of TCPS is clearly on outcomes, the individual standards and measures vary considerably in the extent and breadth of their identification of outcomes versus mere outputs.

Some of this variation is attributable simply to the difficulty of identifying meaningful outcomes and creating useful variables, measures, and operational definitions of those outcomes, especially in performance areas where no performance measures and indicators exist. What are the desired specific outcomes of

access to justice, of equality and fairness, and of independence and accountability? During the development of the measures, the TCPS project was plagued by what staff dubbed “process creep,” the tendency to identify standards and measures associated with inputs (the resources that a court uses to produce services) and outputs (the number and types of services delivered) rather than with important results and outcomes.

The following five subsections summarize the major elements of TCPS by court performance area, beginning with Access to Justice.

Access to Justice

The five standards defining the performance area of Access to Justice—presented first because they address the initial entry of litigants and other court users into the judicial system—require that courts be accessible to the individuals they should serve. They require a court, in effect, to eliminate all barriers—physical, geographic, procedural, cognitive, psychological, and attitudinal—to court services. Like all 22 standards, the 5 standards in the area of Access to Justice encourage courts to view their performance from that of the citizens who are served by the courts (outcomes) rather than that of those who run the courts (who are likely to focus on inputs and outputs).

- 1.1. **Public Proceedings:** The trial court conducts its proceedings and other court business openly.
- 1.2. **Safety, Accessibility, and Convenience:** Trial court facilities are safe, accessible, and convenient to use.
- 1.3. **Effective Participation:** The trial court gives all who appear before it the opportunity to participate effectively, without undue hardship or inconvenience.
- 1.4. **Courtesy, Responsiveness, and Respect:** Judges and other court personnel are responsive to the public and accord respect to all with whom they come into contact.
- 1.5. **Affordable Cost of Access:** The costs of access to trial court proceedings and records—whether measured in terms of money, time, or the procedures that must be followed—are reasonable, fair, and affordable.

Twenty-one specific performance measures are associated with these five standards. Together, the measures provide both breadth and depth of measurement of a court’s performance in offering public access to justice. Prescribed methods

of measurement include structured observations, interviews, surveys, record searches, and reviews. Three measures call for administering surveys (forms and instructions are provided) to individuals who are “regular users of the courthouse.” The information sought relates to safety and security, the ease of “doing business” with the court, and the courtesy and respect experienced by these court users.

The method prescribed most often for measuring access is observation (sometimes combined with simulation). Observers systematically record what they see and hear. There are 12 measures of this type. Two other measurement methods rely on data collected through interviews and examination of court records and written policy documents. Some of the measures of this type focus on case data. For Measure 1.3.1, focused on effective legal representation of children in abuse and neglect proceedings, for example, court case records are examined and the individuals involved in the cases are surveyed and interviewed to document how the guardian *ad litem* process actually worked for selected cases. Other measures focus on administrative documents. For example, Measure 1.5.1, Inventory of Assistance Alternatives for Financially Disadvantaged, requires examination of forms, brochures, and written policies to evaluate court efforts to facilitate affordable access alternatives for individuals with low incomes. Finally, measures addressing the issues of court security (Measure 1.2.1) and interpreter services (Measure 1.3.2) require evaluation by outside experts in their respective areas.

Expedition and Timeliness

Court reform in the past 20 years has focused on the pace of litigation and reduction of delay in case processing. The three standards in the second performance area expand the requirement of timely case processing to all court activities.

- 2.1. Case Processing: The trial court establishes and complies with recognized timelines for timely case process while keeping current with its incoming caseload.
- 2.2. Compliance with Schedules: The trial court disburses funds promptly, provides reports according to required schedules, and responds to requests for information and other services on an established schedule that assures their effective use.
- 2.3. Prompt Implementation of Law and Procedure: The trial court promptly implements changes in law and procedure.

The 10 measures for this area's three standards assess how promptly the court processes cases, files required reports, disburses funds, and implements new legal and procedural changes. Because of the diversity of performances required by the three standards, a wide range of measurement techniques are employed—record review, observation and simulation, surveys, interviews, and structured group techniques. Several measures associated with this performance area—length of time to case disposition, ratio of case dispositions to case filings, and the age of the pending caseload—are familiar to most court managers and judges. Four other measures draw on State and local sources of information to determine whether a court is performing non-case-related functions (e.g., distributing funds and providing reports, information, and services) in a timely manner. Satisfactory performance requires not only that provision of reports be timely but also that they be completed in a manner useful to the person or agency requesting the information or report. Finally, Measures 2.3.1, Implementation of Changes in Substantive and Procedural Laws, and 2.3.2, Implementation in Changes in Administrative Procedure, relate to the promptness with which a court implements externally mandated changes.

Equality, Fairness, and Integrity

This requirement is articulated by the following six standards:

- 3.1. Fair and Reliable Judicial Process: Trial court procedures faithfully adhere to relevant laws, procedural rules, and established policies.
 - 3.2. Juries: Jury lists are representative of the jurisdiction from which they are drawn.
 - 3.3. Court Decisions and Actions: Trial courts give individual attention to cases, deciding them without undue disparity among like cases and on legally relevant factors.
 - 3.4. Clarity: The trial court renders decisions that unambiguously address the issues presented to it and clearly indicate how compliance can be achieved.
 - 3.5. Responsibility for Enforcement: The trial court takes appropriate responsibility for the enforcement of its orders.
 - 3.6. Production and Preservation of Records: Records of all relevant court decisions and actions are accurate and properly preserved.
- The six standards address the constitutional guarantees of due process and equal protection under the law. The standards emphasize integrity and fidelity to established laws and procedure and require courts not only to be explicit in their

orders but also to ensure their enforcement. The first standard requires that trial courts adhere to laws they are responsible for upholding. The equality and fairness afforded to litigants and disputes are determined not only by judicial officers but also by juries. Standard 3.2 requires that trial courts do their utmost to encourage equality, fairness, and integrity by ensuring that individuals called for jury duty are representative of the population from which the jury was drawn.

Standard 3.3 focuses on what many consider the essence of justice—the requirement that a court’s decisions and actions be based on legally relevant factors consistently applied in all cases. Decisions and actions should be the result of individual attention paid to individual cases. Because clarity and understanding are prerequisites of compliance and enforcement, Standard 3.4 requires courts to render decisions that clearly address the issues and specify how compliance with their decisions can be achieved.

Standard 3.5 requires courts to take responsibility for the enforcement of their orders, acknowledging that the responsibility is more often than not shared with its justice partners and other branches of government. Finally, Standard 3.6 requires the proper and accurate preservation of court records. Records of court decisions and court process constitute, in an important sense, the law. Both accuracy and reliable and prompt access to these records are fundamental to the purposes of courts.

A total of 23 performance measures are associated with the six standards in the area of Equality, Fairness, and Integrity. They are intended to provide systematic performance information on the many facets of this complex and important area. For most of the standards, associated measures use similar data elements, data-gathering procedures, and methods of analysis. For example, for five of the six measures of Standard 3.6, a common database is used to assess the integrity of the court’s record management systems. The measures prescribe use of some portion of the same pool of cases to examine the extent to which court records are adequately stored. The use of a joint database is suggested for other standards. For example, Measures 3.3.3, Equality and Fairness in Sentencing, and 3.3.4, Equality and Fairness in Bail Decisions, rely on the same set of cases and the same methodological approach in determining whether legally irrelevant factors play a role in sentencing and bail decisions. A court that decides to measure a given standard will find that it can apply all of the measures within that standard in an efficient manner.

The most common approach to the measures in this area is the analysis of case-related information. Case files are used as a basic source of data for 17 of the 23 measures. In some instances, the information in files is gathered and analyzed to assess the fairness of court decisions, such as in bail and sentencing.

Case-related information is also used in Standard 3.1 as a means of determining the extent to which the court adheres to laws and procedures. Here the case-related information is used as a way to verify compliance with laws.

The next most common approach is the use of mail questionnaires to assess the views of key participants in the trial court process. Different measures target different samples of respondents. For example, Measure 3.3.3 seeks to determine how court employees and attorneys assess a court's performance in applying the law. Measure 3.3.1 targets the bar's view toward the fairness of court decisions and actions. Measure 3.3.2 provides a parallel survey of court users. Measure 3.6.6 is directed at the views of attorneys toward the adequacy of the court record when cases are appealed.

Finally, the three measures associated with Standard 3.2 call for an examination of court records pertaining to the selection of jurors. The lists of potential jurors are compared with other sources of information, such as census reports, to determine inclusiveness, randomness, and representativeness.

Independence and Accountability

The five standards in the fourth performance area recognize the importance of the independence of the judiciary, the courts' institutional integrity, and separation of powers. At the same time, however, the standards require courts to maintain effective working relationships (comity) with other branches of government and justice system partners. Focusing on courts as public organizations, these standards require them to balance independence with public accountability, practice good stewardship of resources, ensure that their internal personnel practices meet the highest standards, and be a responsive component of government.

- 4.1. Independence and Comity: The trial court maintains its institutional integrity and observes the principle of comity in its governmental relations.
- 4.2. Accountability for Public Resources: The trial court responsibly seeks, uses, and accounts for its public resources.
- 4.3. Personnel Practices and Decisions: The trial court uses fair employment practices.
- 4.4. Public Education: The trial court informs the community about its programs.
- 4.5. Response to Change: The trial court anticipates new conditions and emergent events and adjusts its operations as necessary.

In contrast to the measurement approach in the other four performance areas—which is largely *prescriptive* in its detailing of specific measures and indices—the measurement approach in the area of Independence and Accountability is largely *heuristic*. Rather than defining specific measures of performance, it requires methods by which a court proceeds along empirical lines to identify the people, events, and activities needed to develop valid and feasible performance measures to assess independence and accountability. And rather than initially taking a specific measurement, the court engages in a process that will allow it to make inferences about its performance based on empirical results. Field tests and demonstrations of experimental measurement approaches for the standards and measures (see the section “A Brief History of the TCPS Project”) indicated that performance assessment in this area is highly context-driven. Differences in the sizes of courts, the statutory frameworks governing court funding, and the structural arrangements of essential justice system services make it difficult to prescribe a standard set of measurement approaches.

The measures should only be undertaken following the formation of a steering committee of judges and court managers. This group will be involved in planning data collection, considering the significance of the results, and integrating the findings from all the measures into an overall view of court performance in this area. Structured group techniques, such as the Nominal Group Technique and Ideawriting, led by a skilled facilitator, are recommended to save time during steering committee meetings and to maximize objectivity in the committee’s work. The use of the steering committee, in conjunction with research efforts that may be undertaken by court staff or consultants, combine fact-gathering, value clarification, decisionmaking, and action.

***A judicial system
derives its authority
and legitimacy from
those it serves.***

Public Trust and Confidence

A judicial system derives its authority and legitimacy from those it serves. The central question posed by the three standards in this area is whether court performance—in accordance with standards in the other four performance areas—actually instills public trust and confidence.

- 5.1. Accessibility: The public perceives the trial court and the justice it delivers as accessible.
- 5.2. Expeditious, Fair, and Reliable Court Functions: The public has trust and confidence that basic court functions are conducted expeditiously and fairly and that court decisions have integrity.

5.3. Judicial Independence and Accountability: The public perceives the trial court as independent, not unduly influenced by other components of government, and accountable.

A court must attend to four major constituencies served by the courts. They vary in type and in the extent of contact and experience with the courts. At the most general level is the local community or the “general public”—the vast majority of citizens who seldom experience the courts and may have formed their beliefs by watching TV, reading newspapers, or surfing the Internet. A second constituency is a community’s opinion leaders—the newspaper editor, the reporter assigned to courts, the police chief, the mayor, members of the board of supervisors, business leaders, and members of court watch committees. A third constituency includes citizens who appear in court on a regular basis—referred to as “regular court users.” These include attorneys, litigants, witnesses, jurors, victims, family, friends, and representatives of those who appear before the court. This group obviously has direct experience and knowledge of the court and its routine functions and activities. The last constituency consists of those individuals employed by the court system—judicial officers, managers, and other staff—who may have an insider’s perspective on how well the court is performing.

The three standards in this area are associated with 14 measures, all but three of them drawn from the other performance areas. Because the court’s performance with regard to Public Trust and Confidence is dependent in large part on its performance in other performance areas like Access to Justice and Fairness and Equality, several of the measures rely on *informed* opinions (i.e., opinions of individuals who have had contact with the court). For example, in Measure 1.2.6, Evaluation of Accessibility and Convenience by Court Users, the ease and convenience of conducting business with the court is measured through a survey of regular court users—court employees, attorneys, probation officers, and jurors.

Three measures address public trust and confidence squarely: Measure 5.1.1, Court Employees’ Perceptions of Court Performance; 5.1.2, Justice System Representatives’ Perceptions of Court Performance; and 5.1.3, General Public’s Perceptions of Court Performance. The first measure is conducted through a mail survey of court employees, the second through a focus group discussion with representatives of the various components of the justice system, and the third through a telephone survey of the general public.

A Brief History of the TCPS Project

The TCPS project has drawn much of the blueprint for court performance measurement called for by Cook and his colleagues in 1982 and has resolved many of the issues they identified for more attention (p. 11). The 10-year, multimillion-dollar project followed a traditional course of applied social research—research and development; testing, refinement, and demonstration; implementation and institutionalization. The National Center for State Courts (NCSC) and the Bureau of Justice Assistance (BJA) of the U.S. Department of Justice initiated the project in 1987 to develop measurable performance standards for State trial courts. At the time, State court systems saw themselves stretched beyond their capacities. Court personnel were experiencing fatigue and burnout in the face of overwhelming increases in drug-related cases. Pressure on the courts created a sense of urgency. In contrast to the prior research on performance measurement, which emphasized conceptual framework development, the TCPS project was launched as an ambitious applied research effort with implications for major public policy reform. Even in its early stages, the project set lofty goals that reached beyond the identification of concepts. Project representatives asserted that TCPS would “define a philosophy, and a valid and widely shared conception, of what optimum trial court performance entails. [T]hese standards will prove a valuable resource for self-assessment and self-improvement of trial courts and provide better ways of meeting the needs of those served by the courts” (Commission on Trial Court Performance Standards 1997a, vii).

Research, innovation, and development: 1987–90

The initial phase of the TCPS project was a 3-year research, innovation, and development effort that began in August 1987 and ended in mid-1990. The project team consisted of a 14-member policymaking and advisory body, the Commission on Trial Court Performance Standards, that included appeals and trial court judges, trial court administrators, a court clerk, an academic researcher, and NCSC project staff. The project staff, which numbered from 5 to 12 during this time, included social scientists, research attorneys, and program administrators. The critical question, “What *should* courts be accomplishing?” that prior attempts had largely failed to answer (Cook et al. 1982, 184–185), was addressed at the outset.

The project team quickly discovered that there was little specific guidance in the literature of court administration on how to measure trial court performance and no consensus on the broader goals against which to measure performance. The success of this difficult and important initial stage of goal setting, as others have suggested in different contexts (see Danegger et al. 1999, 5, 20), was in

no small part made easier by the political legitimacy and credibility of the 14 members of the Commission on Trial Court Performance Standards. It is doubtful that the project would have progressed as far and as fast without the political standing that the Commission gave the project.

The following objectives and strategies were established for the project within the first year after several months of deliberations about the general contours of the project:

- Development of a manageable number (20–25) of standards of trial court performance through a process that included the preparation of a series of “briefing papers” by project staff, consideration of the issues raised by the briefing papers by the Commission and project staff, and the crafting of standards and commentary in five performance areas. The identification of the five performance areas was the subject of the first several briefing papers and meetings of the Commission.
- Development of a comprehensive measurement system built around the standards, including performance measures or indicators, data collection methods and techniques by which measures would be taken, requirements for data, and a performance evaluation scheme by which the measurement system could be applied by trial courts throughout the country.
- Field testing and application of the performance standards and measurement system in selected “demonstration” courts.
- Dissemination, promulgation, and acceptance (institutionalization) of TCPS by key judicial organizations and several States.

In 1990, a tentative version of the performance standards, measures, and commentaries, published in the form of a 41-page booklet (Commission on Trial Court Performance Standards 1990), and a companion videotape describing the TCPS project, were widely distributed for review. The initial objective of the project—a manageable number of performance standards grouped in five performance areas and a tentative set of associated measures—was largely accomplished by 1990. Comments and suggestions for improvement of the standards were received from judges; elected and appointed court managers at the State and local levels; judicial administration scholars; representatives of various national, State, and local judicial administration organizations; and other interested individuals and groups.

In general, the tentative standards were well received once they were seen in print. Although many observers of the work of the project expressed early skepticism about any attempts to identify what courts *ought* to be doing in any

meaningful way, most agreed with the sentiment of a judge who commented after reviewing a tentative version of the standards, "These things are like Mom, apple pie, and the American way. Who can quibble with them?"

At the same time, in addition to the standards and commentaries, the Standards Project staff developed a tentative measurement system to accompany the standards (Trial Court Performance Standards Project 1990). The development of many of the measures was based on a performance-measurement process with a long history of use in industrial and organizational psychology (Smith and Kendall 1963) that entails identifying critical factors of effective and ineffective behaviors and scaling them along an effectiveness dimension.

In contrast to the standards expressing the broad goals of courts against which performance would be gauged, the tentative measurement system that was summarized in the 1990 booklet of TCPS drew strong negative reactions from the field. The summary included only a brief narrative and tabular description of 75 specific measures associated with the 22 standards, the object or subject of the measurement (e.g., child support orders), the methods of measurement (e.g., case file reviews), and the person or persons responsible for taking the measure (by whom the measure is applied).

A number of judicial groups and individuals raised strong objections to an advance draft of this tentative version of the performance measures, foreshadowing political and attitudinal obstacles that would confront the application of TCPS in later phases of the project. The ready acceptance of the standards (i.e., the performance goals and the principles and ideals on which they were based) would not be extended to the associated measurement system. It seemed as if the articulation of specific measures—a total of 75—signaled a seriousness of purpose of the project that had not been evident to observers before. One group of California judges, for example, persistently voiced its objections to the president of NCSC, the chief justice of California, and representatives of BJA, the agency of the U.S. Department of Justice funding the TCPS project. Despite repeated assurances by the project team that TCPS were intended for self-assessment and improvement, the California group contended that TCPS was in fact a mechanism that NCSC planned to use to build a national system of "accreditation" of courts, a development the group rejected as inappropriate. The group used language lifted from 3-year-old grant applications, obtained from BJA under the authority of the Freedom of Information Act, to bolster its contention. The group also complained that TCPS would be used inappropriately to gauge the performances of individual judges.

Cautioning researchers and policymakers about any efforts to define measures for the goals of justice agencies, John J. DiIulio, Jr., has observed that "it is

often the most dedicated and caring government workers who are anywhere from suspicious to downright dismissive of any attempt to define and apply such measures” (1993, 155). In an attempt to placate the judicial organizations whose endorsement of TCPS was deemed critical to their acceptance and use, project staff wrote the caveat about court performance measurement that appears, emphasized by italics, on the second page of the introduction to the *Trial Court Performance Standards With Commentary*: “The use of the standards as a basis for cross-court comparisons or as part of a national regional accreditation of State Courts is not intended or recommended. . . . The standards and accompanying measurement system also are not intended, nor are they appropriate, for gauging the performance of individual judges” (Commission on Trial Court Performance Standards 1997a, 2). These cautionary words accurately reflect a negative mental model of court performance measurement that was, and probably continues to be, pervasive in the courts. It seemed to be based on tacit assumptions and fears that performance measurement is something done *to* us, not *for* us, by third parties who do not necessarily have the court’s best interests in mind; that it is predicated on serious misunderstandings about how courts work; that it is accomplished with methods that are insensitive to the courts’ unique working environments and operations; that it results in numbers that bear little relationship to what is purported to be measured; and that it is more likely to be harmful than helpful.

Peter M. Senge, a prominent management theorist, defines mental models as the deeply ingrained assumptions, generalizations, and images that shape how we understand the world and take actions (1990, 174–204; Senge et al. 1994, 235–293). Differences in mental models explain why two people can observe the same phenomenon and describe it quite differently. The classic story, “The Emperor’s New Clothes,” Senge writes, is less about ignorant people than it is about people bound by powerful images of the emperor’s dignity that prevented them from seeing him as naked. The TCPS project team learned early in the project’s life that they would ignore this negative model of performance measurement at the risk of jeopardy to the project. It seemed that acceptance and use of TCPS would depend on continued assurances of autonomy of control and use of TCPS by trial courts themselves.

Testing, refinement, and demonstration: 1989–95

As the measurement system developed, 75 tentative measures were tested and refined. Trial courts in Arizona, Michigan, and Ohio contributed to this process by serving as test sites for the tentative measures. Testing consisted largely of informal efforts by project staff to take the various measures with the assistance

of court staff. Although court staff were questioned about the feasibility, utility, and likelihood of use of the measures, no attempts were made at the time to incorporate the measures into the day-to-day operations of the “test” courts.

Beginning in August 1990, work began on the next phase of the project, which involved broader testing and demonstration of the feasibility and utility of the measures. The objectives of this phase were (1) to move the project from innovation to limited implementation in selected States, where the use of the trial court performance standards and their accompanying measurement system could be tested; (2) to continue refining and adapting the measurement system to meet the requirements of trial courts and the State administrative offices of the courts; and (3) to provide the foundation for the acceptance and institutionalization of the standards and measurement system as a useful tool of judicial administration.

The demonstrations were conducted in 13 trial courts of different sizes in New Jersey, Ohio, Virginia, and Washington.⁸ In consultation with project staff and the administrative office of the courts in the respective States, trial courts agreed to demonstrate a selected number of performance measures. Because the Conference of State Court Administrators and the Conference of Chief Judges had already endorsed the tentative version of TCPS, and the project had gained a certain notoriety, many trial courts viewed participation with the continuing project as “demonstration sites” as a way to enter the national spotlight. Consequently, recruitment of demonstration courts turned out to be relatively easy.

A number of considerations determined what measures would be taken by what demonstration court, including the interests and needs of the courts and the State court administrative offices, the need of the project to demonstrate all measures in as many different operating environments as possible, and the size and resources of the court. Each of the 75 measures was taken as prescribed by the tentative measurement system by at least one of the demonstration courts, with most measures demonstrated by more than one (Trial Court Performance Standards Project 1990).

The demonstrations led to a revision of the tentative measures—including elimination of some, simplification of others, and confirmation of most in revised form. The 75 measures in the tentative 1990 version of TCPS eventually were reduced to 68 measures in the final version.

In retrospect, it is clear that the project failed to take full advantage of the demonstrations that directly engaged approximately 100 trial court personnel, and indirectly perhaps twice that number, in 13 sites for more than a year. Despite efforts by project staff to fit the demonstrations as much as possible

into the day-to-day operating environments of the courts, the demonstrations of the measures remained, for the most part, artificial. Although the participants claimed benefits of the demonstrations for their own management and planning, the measurements seemed to be demonstrated because of a requirement of the TCPS project, not because the measures necessarily were needed or desirable for purposes of management decisionmaking in the demonstration courts. Most of the participants in the 13 demonstration courts viewed their efforts as contributions to national research and development by NCSC, not as an opportunity to demonstrate real benefits of performance measurement in their courts. Even today, when asked to comment on the technical, administrative, and political obstacles that stood in the way of implementing TCPS in their courts, participants speak more of the difficulties of the demonstration *per se* (e.g., interactions with the TCPS project team) than the merits of the TCPS measurement system from their perspectives.

By tightly controlling the measures taken and prescribing the steps by which they were to be taken, the project probably limited the information gleaned about the application of TCPS in actual operating court environments. As a result, although the demonstrations served the purpose of acquiring information for refinement of the technical aspects of the measurement system (e.g., the ability of courts to take certain measures without outside assistance, the difficulties of obtaining samples, and the ease of use of data collection instruments), they contributed little to knowledge of how and under what circumstances TCPS would be used by courts with no obligation to use them. Although project staff consulted with and visited demonstration sites frequently and diligently recorded comments and suggestions for improvement of the measures and the measurement process, their focus was almost exclusively the revision of the individual measures, their operational definitions, and the description of the methods by which the measures were to be taken. Whatever information that may have been obtained that would address the most frequently asked question today, "How are TCPS being used and with what results?" was not retained. Notwithstanding the "artificiality" of the demonstrations, this question could have been addressed at relatively little cost by a systematic effort to inquire about, record, and catalog the input, output, and outcomes produced by the performance measurement efforts of demonstration courts. It seems likely that a similar effort would need to be made in the future with courts using TCPS today (see the section "The Inputs, Outputs, and Outcomes of Performance Measurement").

Institutionalization of TCPS: 1995–2000

In June 1995, the Commission, meeting for the last time in Baltimore, expressed the vision that by 2000, all State courts would be using TCPS in a systematic or organized way. Late in 1994, the first intensive 3-day training course on the use

Between 1990 and 1999, five major judicial and court management organizations, representing most State courts, endorsed TCPS by formal resolutions, proclamations, or publications: the Conference of Chief Justices, the Conference of State Court Administrators, the National Association for Court Management, the College of Probate Judges, and the American Judges Association.

of TCPS, “Organizing Your Court: How to Use the *Trial Court Performance Standards*,” was conducted in Tucson, Arizona, with 27 court managers and judges from throughout the country. Sponsored by NCSC’s Institute for Court Management, the course focused on the use of TCPS as: (1) a common language for the description, classification, and communication of court activities; (2) a means for self-assessment, self-improvement, and accountability; (3) a conceptual framework for understanding and improving court performance; and (4) a guide for court planning, management, and leadership.

Although the Commission’s vision that all State courts would be using TCPS by 2000 is unlikely to be realized, the extent of the institutionalization of TCPS in court management and judicial administration is broad.

Endorsements

Between 1990 and 1999, five major judicial and court management organizations, representing most State courts, endorsed TCPS by formal resolutions, proclamations, or publications: the Conference of Chief Justices, the Conference of State Court Administrators, the National Association for Court Management, the College of Probate Judges,⁹ and the American Judges Association.

State-level efforts

Administrative offices of State courts or supreme courts in at least 20 States adopted TCPS as a basis for their long-term or strategic planning. California, for example, codified TCPS as a California Judicial Council Rule and is using them as the framework of its ambitious strategic plan (Commission on the Future of the California Courts 1993).

General jurisdiction courts

An estimated 1,200 trial courts (approximately 40 percent) are using TCPS in some way. For example, the 19th Circuit Court in Illinois and the 26th Judicial District Court in North Carolina currently are using TCPS to guide their strategic planning efforts.

This and subsequent estimates of adoption or use of TCPS are based on the direct experiences and informed judgments of the TCPS project, NCSC's Institute for Court Management, and their associates actively engaged in the promulgation of TCPS. A precise definition of "use" of TCPS remains to be made. For the purpose of these estimates, however, use was loosely defined as any effort by courts, beyond simple exposure to TCPS by training and education, to apply TCPS to their management, planning, or leadership.

Limited jurisdiction courts

An estimated 7,000 (40 percent) municipal, family, and other limited jurisdiction courts are using TCPS. For example, in 1996, the Los Angeles Municipal Court, with funding from the State Justice Institute, attempted to conduct all 68 performance measures as part of its strategic planning process (Anabis-Straub n.d.). The North County (San Diego) Municipal Court's project using TCPS to improve customer service (Lane 1998) was nominated for the Ralph N. Kleps Award for improvement in the administration of justice in California courts. With funding from the State Justice Institute, the Family Court of Delaware is adapting TCPS for family courts.

Other court efforts

Other State and Federal courts, court organizations, and foreign courts (including those in Austria, Australia, Canada, Hong Kong, and the Philippines) are using TCPS (see, for example, Gryphon Consulting Services 1998).

Publications and dissemination

Since 1990, an estimated 20,000 copies of various publications featuring TCPS have been distributed by NCSC, BJA, and the State Justice Institute.

Education and training

About 1,500 court managers, judges, and other public officials have been exposed to TCPS in national, State, and local courses conducted by the Institute for Court Management of NCSC and its associates since the end of 1994. Numerous presentations of TCPS have been made at meetings of national judicial organizations. With funding from BJA, NCSC recently established a TCPS resource center that maintains an Internet listserv and a database of individuals interested in TCPS. The center is producing an educational video, a CD-ROM version of the *Trial Court Performance Standards and Measurement System*, and other resource materials.

Advancing the Inquiry Beyond Goals

TCPS make it possible to advance the inquiry into performance measurement of the courts—and other components of the justice system¹⁰—from one of conceptualization and identification of constructs, variables, and operational definition to one of critical review and evaluation of a comprehensive measurement system currently in use in the courts. This is a significant and potentially rewarding advance for justice system researchers, policymakers, and practitioners. Questions of what the goals of court performance *ought* to be can give way, at least somewhat, to questions about whether the performance goals, principles, and ideals expressed by TCPS are the appropriate ones. Do TCPS concentrate on significant measures of court performance? Are the measures relevant to other components of the justice system? Do the measures together constitute a “balanced scorecard” of important outputs and outcomes (Kaplan and Norton 1992, 1993)? Does the measurement system produce data that are accurate, consistent over time, and valuable to both practitioners and policymakers? These questions can be addressed empirically by evaluations in courts using TCPS.

In their 1982 synthesis of the extant literature, Cook and his colleagues lamented: “One looks in vain for a well-articulated formal hierarchy of goals for criminal justice agencies against which the performance of these agencies could be measured. . . . [P]rior attempts to identify an agreed upon set of goals for the court system were unsuccessful; apparent consensus broke down on the particulars of what courts *ought* to be doing” (Cook et al. 1982, 184–185). Clearly, the overall value of a performance measurement system is dependent on the quality of the individual measures. However, without minimizing the importance of TCPS’ greatly detailed procedures for measuring performance, the most significant contribution of TCPS to the literature of performance measurement may be their “formal hierarchy of goals”—the theoretical concepts, abstractions, and new categories of thinking about the fundamental responsibilities of courts—on which TCPS are built. Goal formulation is recognized as perhaps the most critical step in the strategic planning process. The effort is as much political as it is rational (see Bryson 1995, 10–13). Researchers and policymakers may have underestimated the difficulty of identifying and formulating the broad goals of courts as they looked to the more technical challenge of creating measures and indicators of the work performed and results achieved by courts.

Organizational performance standards for courts and other public organizations are largely meaningless if they are detached from constructs they are intended to represent. At best, they are grounded in historical understanding, democratic vision, and civic ideals (see DiIulio 1993b). Traditionally, government performance measurements are developed from the perspectives of

government managers, not of citizens, and the two may differ greatly. TCPS clarify what citizens want from their courts and what results are desired. Citizens want ready access to the justice delivered by the courts; they want that access to be safe, relatively convenient, and affordable. Once they have gained access, they want their business with the courts dealt with expeditiously and fairly, according to the facts and according to established rules. They want their disputes to get individual attention and to be dealt with fairly. They want their courts to be independent of other branches of government and other agencies to assure that decisions and actions are based solely on legally relevant factors. Ultimately, they seek trust and confidence in the courts. These civic ideals are identified by TCPS' five performance areas: Access to Justice; Expedition and Timeliness; Equality, Fairness, and Integrity; Independence and Accountability; and Public Trust and Confidence in the courts.

The underlying principles of TCPS support these constructs: (1) a focus on outcomes (results and performance) rather than inputs, (2) courts as organizations (i.e., the organization as the unit of analysis), (3) the public as definers of desired results (i.e., service orientation), and (4) responsiveness to community needs. By focusing on results that matter to those served by the courts, rather than those who run them, TCPS avoid the most common pitfalls of performance measurement in the public sector—a failure to identify important outcomes and a confusion of inputs (e.g., processes, resources, and structures) with outputs and outcomes. TCPS represent a shift from thinking about courts as individual judges making individual decisions (one judge, one court) to thinking about courts as public organizations—as a system of structures, people, methods, and practices brought together to achieve specific ends. Viewing the courts as organizations makes Standard 3.5, Responsibility for Enforcement, and Standard 4.5, Response to Change, for example (see previous section, “Overview of TCPS”), more acceptable to judges who might otherwise consider these standards as threats to the separation of powers and an endorsement of judicial activism. TCPS' emphasis on performance changes the focus from the perspective of insiders (those who run the courts) to those who are served by the courts. TCPS represent a change of thinking about the courts' role in society from isolation and independence to interdependence and community responsibility. The welfare and quality of life of the community and its citizens should matter to courts as organizations.

TCPS represent a shift from thinking about courts as individual judges making individual decisions to thinking about courts as public organizations—as a system of structures, people, methods, and practices brought together to achieve specific ends.

They define not only courts' guiding ideas—their direction, values, and purpose—but also their structures and processes. The emergence of specialized courts, such as drug courts, community courts, teen courts, and domestic violence courts, reflect the importance of community to how we structure and run our courts.¹¹

The Inputs, Outputs, and Outcomes of Performance Measurement

What are the benefits or desired results of the adoption and deployment of court performance measurement? Although there appears to be widespread recognition that organizational performance measurement is only a tool, not an end in itself, the desired results or ends of court performance measurement have yet to be specified with any precision. Court researchers and practitioners engaged in court performance measurement have not examined their own efforts with any regularity or rigor. They need to examine not only what they really *do* but, more important, what they and their performance measurement processes really get *done* in terms of their impact on the functioning and effectiveness of the courts.¹² Meritorious outcomes resulting from performance measurement, like those of management and strategic planning in the courts, are simply assumed (see Keilitz, Davis, and Benedict in press).

The logic and language of performance measurement can be put to use as a framework for reviewing and evaluating court performance measurement as a strategy to achieve desired results. Using this framework, court performance measurement, like anything the courts do, entails resources and strategies (inputs) that act on the operating environment of the justice system and community to produce services (outputs) that, in turn, produce demonstrable changes in the well-being of the public and the community served by the courts (outcomes). What are the *inputs* of performance measurement in the courts—the human, financial, facility, and material resources expended? What are the *outputs*—the activities, procedures, and services produced by performance measurement? Finally, what are the *outcomes* of performance measurement—the results or impacts of the inputs and outputs on the court and the community? The language and logic of performance measurement can be, and arguably should be, productively applied to efforts of performance measurement themselves. TCPS' comprehensive measurement system, a system that has actually been used by courts, makes this inquiry feasible.

The input-output-outcome framework also may be useful for addressing the question of what constitutes “adoption” and “use” of a performance measurement

system, a question that has plagued the TCPS project since its demonstration phase. The adoption of TCPS by a court or a State administrative office of the courts, absent a showing of sufficient inputs of human, financial, and material resources, for example, would raise questions of the legitimacy of the “adoption” of the measurement system. “Use” could be construed at the level of output or outcome. The publication of performance measurement results would constitute a legitimate output but would not be considered an outcome of performance measurement.

The actual inputs of court performance measurement can be framed in accordance with the efforts of planning, preparation, data collection, data analysis, and reporting prescribed for the various performance measures of TCPS: number of court staff and others employed to prepare for the measurement effort and to take the measures; amount of time and money expended; the scope and amount of data taken; the number and size of samples drawn; and equipment, materials, and facilities required. (As noted earlier, although it would have been feasible to catalog the inputs of performance measurement in this fashion during the demonstration phase of the TCPS project, it was not done.) Outputs of performance measurement may include, for example, the type and number of measures taken, amount of information acquired, reports and presentations made, and the form that these reports and presentations take (e.g., a publication or a real-time guidance system that includes several indices joining multiple measures). Finally, the broad categories of outcomes may include decisions made, actions taken, and results achieved as a result of performance measurement in the management, planning, and leadership of the courts in the performance areas identified by TCPS (i.e., Access to Justice, Expedition and Timeliness, and so forth).

Although an input-output-outcome framework of inquiry may facilitate the evaluation of performance measurement using TCPS and improvements in individual measures and the measurement system as a whole, its immediate value may lie in promoting and overcoming resistance to the use of performance measurement. It is common but not sufficient simply to proclaim the advantages of performance measurement—focus, attention, understanding, control, accountability, prediction, influence, and strategy development—and to expect its effective implementation. We must first acknowledge and address the negative mental models that impede its successful use. The “discipline” of mental models is an essential program of study and practice of learning organizations (Senge 1990; Senge et al. 1994). When it applied to court performance measurement, it requires continually clarifying and improving the picture of court performance measurement and seeing how it shapes important decisions and actions.

Being rigorously explicit about the inputs, outputs, and expected outcomes of court performance measurement undoubtedly will help in creating a more accurate—albeit not necessarily a more favorable—picture of court performance measurement as an enterprise. The experience of the TCPS project suggests that proponents of performance measurement may have been at the same time too sanguine and too imprecise about the benefits of performance assessment (exhortations such as “What gets measured gets attention” and “What gets counted, counts” remain largely unsupported) and too reticent to explore the limitations of performance measurement.

The fear that measures of case disposition times or disposition (clearance) rates, for example, would put a court, or even an individual judge, in an unfavorable light is not necessarily unfounded. Indeed, the negative mental model that gives birth to the fear merely acknowledges the law of unintended consequences. By recognizing that cross-court comparisons unfavorable to a particular court and evaluations of individual judges based on organizational performance measures *can* occur but that careful attention to the outputs and outcomes of performance measurement may control unintended consequences, if not eliminate them, proponents of performance measurement can do much to advance their cause. Restrictions imposed on the output of performance measurement (such as restricting access to individual judges’ contribution to aggregate data) and explicit methods limiting the intended outcomes of performance measurement may do much to blunt the negative mental model of performance measurement and advance the value of performance measurement for court administration, planning, and leadership.

Conclusion

TCPS have achieved the “well-articulated consensus” about basic issues of court performance that has eluded prior research and policy development (Cook et al. 1982, 6). TCPS deserve serious consideration by justice system researchers and policymakers, first, because they represent a comprehensive, outcome-oriented organizational performance system for courts—including a hierarchy of goals and principles, variables, and operational definitions of methods—and, second, because they are in use today in about a third of State courts. They provide researchers and policymakers the opportunities to advance the inquiry into justice system performance beyond identification of issues to evaluation and further development of elements of an accepted paradigm in use today.

A logical line of advanced inquiry is research and evaluation of the elements of TCPS as conceived and implemented in the State courts. How are TCPS being used and with what results? Do TCPS concentrate on significant measures of

court performance? Are the measures relevant to other components of the justice system? Do the measures together constitute a “balanced scorecard” of important outputs and outcomes? Does the measurement system produce data that are accurate, consistent over time, and valuable to both practitioners and policymakers? The logic and language of performance measurement may be put to use as a framework for reviewing and evaluating court performance measurement as a strategy to achieve desired results. What are the inputs of performance measurement in the courts—the human, financial, facility, and material resources expended? What are the outputs—the activities, procedures, and services produced by performance measurement? What are the outcomes of performance measurement—the results or impacts of the inputs and outputs on the court and the community? TCPS’ comprehensive measurement system, a system that has been used by courts, makes this line of inquiry feasible.

In addition to this obvious line of inquiry, several other avenues of research and development are opened up by TCPS. TCPS provide a framework that can facilitate the development of new and improved measures of performance of courts and other components of the justice system. By framing meaningful performance outcomes, such as access to justice, TCPS serve to highlight new avenues of research and development. The performance area of Access to Justice is fertile ground for the development of new meaningful outcome measures. This is particularly true for Standard 1.3, Effective Participation. Access and accommodation for *pro se* litigants—individuals appearing in court without representation by lawyers—has become a major problem for State courts in the past 10 years. In some jurisdictions, at least one of the parties is not represented in the majority of domestic relations cases, for example. This issue is not highlighted by Standard 3.1; indeed, it had not been an identified problem at the time of the development of TCPS. Effective participation for litigants without lawyers and access to justice, however, may well be, at least at the conceptual level, a more powerful outcome measure than those currently associated with Standard 1.3. In what proportion of cases, and what case types, do litigants appear without lawyers? How do the proportion and types of cases with *pro se* litigants compare with other courts in jurisdictions of similar socioeconomic profiles? Are courts effectively closing their doors and denying effective participation to *pro se* litigants if the percentage of those served by the courts slips below a certain standard? These are questions relevant to research and development of new performance outcomes and measures in the area of Access to Justice.

Another fruitful area of research and development opened up by TCPS is the creation and development of various indices that combine several measures in a discrete area of performance. “The key to having a successful set of metrics is paring down your database to the vital few key metrics that are linked to your success,” advises Mark Graham Brown in *Keeping Score: Using the Right*

Metrics to Drive World-Class Performance (1996, 4). Multiple measures in a “family of metrics” can be assigned weights according to their importance and combined in an aggregate statistical index. An example of such an index might join the four measures associated with Standard 2.1, Case Processing: time to case disposition (cycle time), disposition or clearance ratio, case backlog, and certainty of trial date. These four measures of case processing, expressed as proportions, would be reduced to one number. The resulting index would require calculating the measures as prescribed in TCPS with some deviations to accommodate the aggregation of the measures into an index.

Justice policymakers, researchers, and practitioners can learn something about how to fashion and implement performance measurement, not only for courts but for other components of the justice system, from the history of the development of TCPS and the experiences of courts actually using TCPS. The State courts and court organizations—general and limited jurisdiction courts and State administrative offices of the courts—that have adopted and used TCPS can serve as laboratories for innovation and development of justice system performance measurement.

Notes

1. Another example is welfare reform (see Nathan 1988).
2. The other two volumes of the four-volume set, the *Planning Guide* and the *Program Brief*, provide direction for using TCPS as a planning and evaluation tool and address questions and issues that policymakers and court officials are likely to encounter in implementing TCPS.
3. Professor George F. Cole has published perhaps the only academic article describing TCPS in any detail (see Cole 1993). In contrast, over the past several years, most issues of *The Court Manager*, the publication of the National Association for Court Management, have mentioned TCPS.
4. Approximately 10,000 copies of this tentative version of TCPS were distributed between 1989 and 1996 (Commission on Trial Court Performance Standards 1990).
5. The interest in alternative dispute resolution, which inspired a burst of innovation and research peaking in the mid-1980s, also drew attention to issues of measurement of the quality of justice. See Tyler 1989 (this article is one of several in a special symposium issue focused on the quality of dispute resolution) and Hensler 1988.
6. These elements are described in detail in two of the four volumes of TCPS (see Commission on Trial Court Performance Standards 1997b, 1997d).

7. At a conceptual level, outcomes are matters of common sense. They are what is important to those who are served by the courts and not necessarily to who run the courts—those who in their day-to-day work might attend primarily to outputs produced by the court without reference to what difference those outputs make to individuals, groups, and communities. For an interesting discussion of outcomes and results-based decisionmaking, see Danegger et al. 1999.

8. In New Jersey, the five demonstration courts were the Superior Courts of Atlantic County, Burlington County, Morris County, Ocean County, and Somerset County. In Ohio, the Common Pleas Courts of Meigs County, Stark County, and Wayne County participated. The Fairfax County Circuit Court was the single demonstration court of TCPS in Virginia. In Washington, the three demonstration courts were the Superior Courts of Spokane County, Thurston County, and Whatcom County (see Saari 1995).

9. All 22 standards with commentary appear with no substantive changes in “Probate Court Performance,” *National Probate Court Standards* (Commission on National Probate Court Standards 1993, sect. 1, 11–26).

10. Justice agencies other than courts have examined TCPS. In May 1997, for example, the Allen County (Indiana) Juvenile Probation Department used TCPS as the basis of a program entitled “Strategic Thinking, Planning, and Strengthening Your Executive Team.” Many of the participants in the “core” course of NCSC’s Court Executive Development Program, “Trial Court Performance Standards,” offered twice a year since 1994, represent justice agencies other than courts.

11. Rottman, Efke, and Casey suggest a new role for courts: that of becoming more responsive to the needs of the community. This role follows a similar one taken by other components of the justice system growing out of a strategy of policing—community policing—focused on establishing a problem-solving partnership with communities (see Rottman and Efke 1998; Rottman and Casey 1999).

12. Of course, court researchers and practitioners have been exposed to the purported benefits of performance measurement—focus, attention, understanding, better decision-making, control, enhanced accountability, prediction, influence, and strategy development—but the description of these benefits to court systems has remained at a broad conceptual level (see also Alpert and Moore 1993).

References

Alpert, Geoffrey P., and Mark H. Moore. 1993. Measuring police performance in the new paradigm of policing. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

American Bar Association, Judicial Administration Division. 1992. *Standards relating to trial courts*. Vol. 2 of *Standards of judicial administration*. Chicago: American Bar Association.

———. 1990. *Standards relating to court organization*. Vol. 1 of *Standards of judicial administration*. Chicago: American Bar Association.

Anabis-Straub. n.d. Implementation & institutionalization of the Trial Court Performance Standards—Los Angeles Municipal Court. Report developed under State Justice Institute grant no. SJI-96-08-W-T-A-034, awarded in 1996.

Brown, Mark Graham. 1996. *Keeping score: Using the right metrics to drive world-class performance*. New York: Quality Resources.

Bryson, John M. 1995. *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement*. San Francisco: Jossey-Bass.

Cole, George F. 1993. Performance measures for the trial courts, prosecution, and public defense. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Commission on the Future of the California Courts. 1993. *Justice in the balance—2020*. San Francisco: California Administrative Office of the Courts.

Commission on National Probate Court Standards. 1993. Probate court performance. Sec. 1 of *National probate court standards*. Williamsburg, Virginia: National Center for State Courts.

Commission on Trial Court Performance Standards. 1997a. *Planning guide for using the Trial Court Performance Standards and measurement system*. NCJ 161568. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Assistance.

———. 1997b. *Trial Court Performance Standards and measurement system implementation manual*. NCJ 161567. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Assistance.

———. 1997c. *Trial Court Performance Standards and measurement system—Program brief*. NCJ 161569. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Assistance.

———. 1997d. *Trial Court Performance Standards with commentary*. NCJ 161570. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Assistance.

———. 1990. *Trial Court Performance Standards and commentary*. Williamsburg, Virginia: National Center for State Courts.

Cook, Thomas J., Ronald W. Johnson, with Ellen Fried, John Gross, Mary Wagner, and James Eisenstein. 1982. *Basic issues in courts performance*. NCJ 84430. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Danegger, Anna E., Carol E. Cohen, Cheryl D. Hayes, and Gwen A. Holden. 1999. *Juvenile Accountability Incentive Block Grants: Strategic planning guide*. Summary, NCJ 172846. Washington, D.C.: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

DiIulio, John J., Jr. 1993a. Measuring performance when there is no bottom line. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

———. 1993b. Rethinking the criminal justice system: Toward a new paradigm. In *Performance measures for the criminal justice system*. Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. NCJ 143505. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Epstein, Paul D. 1988. *Using performance measurement in local government*. New York: National Civic League Press.

Feeney, Floyd. 1993. Evaluating trial court performance. In *Handbook of court administration and management*, edited by Steven W. Hays and Cole B. Graham, Jr. New York: Marcel Dekker.

Gallas, Geoffrey. 1976. The conventional wisdom of State court administration: A critical assessment and an alternative approach. *Justice System Journal* 2 (Spring): 35–55.

Goerdts, John. 1998. *Examining court delay: The pace of litigation in 26 urban courts, 1987*. Williamsburg, Virginia: National Center for State Courts.

Grizzle, Gloria, Jeffrey S. Bass, J. Thomas McEwen, Deborah M. Galvin, Ann G. Jones, Harriet D. Mowitt, and Ann D. Witte. 1982. *Basic issues in corrections performance*. NCJ 84405. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Gryphon Consulting Services. 1998. *Hong Kong judiciary, JISS Phase III feasibility study*. Eureka, California: Gryphon Consulting Services.

Henderson, T.A., C.M. Kerwin, N. Miller, R. Guynes, H. Saizow, R.C. Grieser, and C. Baar. 1984. *The significance of judicial structure: The effect of unification on trial court operations*. NCJ 89501. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Hensler, Deborah R. 1988. Researching civil justice: Problems and pitfalls. *Law and Contemporary Problems* 51 (3) (Summer): 55–65.

Jacoby, Joan E. 1982. *Basic issues in prosecution and public defender performance*. NCJ 84431. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

- Kaplan, Robert S., and David P. Norton. 1993. Putting the balanced scorecard to work. *Harvard Business Review* 71 (5) (September/October): 134–142.
- . 1992. The balanced scorecard—Measures that drive performance. *Harvard Business Review* 70 (1) (January/February): 71–79.
- Keilitz, Ingo, Denise Davis, and Jerry Benedict. In press. Strategic planning for success. *Court Manager* 14 (2): 29–33.
- Lane, Ron. 1998. Beyond the vision: Using the Trial Court Performance Standards to improve customer service in the trial courts. *Court Manager* 13 (2/3): 14–18.
- Nathan, Richard P. 1988. *Social science in government: Uses and misuses*. New York: Basic Books.
- Osborne, David, and Ted Gaebler. 1992. *Reinventing government*. Reading, Massachusetts: Addison-Wesley.
- Pound, Roscoe. 1937. The causes of popular dissatisfaction with the administration of justice. Reprinted in *Journal of the American Judicature Society* 20 (1): 178–187.
- Rottman, David, and Pamela Casey. 1999. Therapeutic jurisprudence and the emergence of problem-solving courts. *National Institute of Justice Journal* 240 (July): 13–17.
- Rottman, David, and Hillery S. Efke. 1998. *A guide to court and community collaboration*. Williamsburg, Virginia: National Center for State Courts.
- Rottman, David B., and William E. Hewitt. 1996. *Trial court structure and performance: A contemporary appraisal*. Williamsburg, Virginia: National Center for State Courts.
- Saari, David J. 1995. Testing the measures of Trial Court Performance Standards. *Court Management & Administration Report* 6 (4): 1–16.
- Senge, Peter M. 1990. *The fifth discipline: The art & practice of the learning organization*. New York: Doubleday.
- Senge, Peter M., Art Klein, Charlotte Roberts, Richard B. Ross, and Bryan J. Smith. 1994. *The fifth discipline fieldbook: Strategies and tools for building a learning organization*. New York: Doubleday.
- Smith, P.C., and L.M. Kendall. 1963. Retranslating of expectations: An approach to the construction of unambiguous anchors of rating scales. *Journal of Applied Psychology* 47:149–155.
- Trial Court Performance Standards Project. 1990. *Measurement of trial court performance: 1990 supplement to the tentative Trial Court Performance Standards with commentary*. Williamsburg, Virginia: National Center for State Courts.

Tyler, Tom R. 1989. The quality of dispute resolution procedures and outcomes: Measurement problems and possibilities. *Denver University Law Review* 66 (3): 419–436.

U.S. Department of Justice, Bureau of Justice Statistics—Princeton University Study Group on Criminal Justice Performance Measures. 1993. *Performance measures for the criminal justice system*. NCJ 143505. Washington, D.C.

Whitaker, Gordon P., Stephen Mastrofski, Elinor Ostrom, Roger Parks, and Stephen L. Percy. 1982. *Basic issues in police performance*. NCJ 84429. Washington, D.C.: U.S. Department of Justice, National Institute of Justice.

Wigmore, John H. 1937. Roscoe Pound's St. Paul address of 1906. *Journal of the American Judicature Society* 20 (1): 176–178.

Appendix: *Criminal Justice 2000* Volumes and Chapters

Volume 1. *The Nature of Crime: Continuity and Change*

Introduction to Volume 1

The Changing Nature of Crime in America

*by Gary LaFree, Robert J. Bursik, Sr., James Short, and
Ralph B. Taylor*

Theoretical Developments in Criminology

by Charles R. Tittle

The Politics of Crime and Punishment

by William Lyons and Stuart Scheingold

Dynamics of the Drug-Crime Relationship

by Helene Raskin White and D.M. Gorman

Criminal Justice Discovers Information Technology

by Maureen Brown

Explaining Regional and Urban Variation in Crime: A Review of Research

by Graham C. Ousey

Change and Continuity in Crime in Rural America

by Ralph A. Weisheit and Joseph F. Donnermeyer

A Century of Juvenile Justice

by Philip W. Harris, Wayne N. Welsh, and Frank Butler

Changes in the Gender Gap in Crime and Women's Economic Marginalization

by Karen Heimer

On Immigration and Crime

by Ramiro Martinez, Jr., and Matthew T. Lee

Volume 2. Boundary Changes in Criminal Justice Organizations

Introduction to Volume 2

A Century of Changing Boundaries

by Charles M. Friel

The Privatization and Civilianization of Policing

by Brian Forst

The Changing Boundaries Between Federal and Local Law Enforcement

by Daniel C. Richman

The Governance of Corrections: Implications of the Changing Interface of Courts and Corrections

by Christopher E. Smith

Brick by Brick: Dismantling the Border Between Juvenile and Adult Justice

by Jeffrey A. Butts and Ojmarrh Mitchell

The Changing Boundaries of the Criminal Justice System: Redefining the Problem and the Response in Domestic Violence

by Alissa Pollitz Worden

The Internationalization of Criminal Justice

by Dr. Richard H. Ward

Community Justice: A Conceptual Framework

by David R. Karp and Todd R. Clear

Volume 3. Policies, Processes, and Decisions of the Criminal Justice System

Introduction to Volume 3

Policies, Processes, and Decisions of the Criminal Justice System

*by Julie Horney, Ruth Peterson, Doris MacKenzie, John Martin,
and Dennis Rosenbaum*

Prison Use and Social Control

by James P. Lynch and William J. Sabol

Changing the Contours of the Criminal Justice System To Meet the Needs of Persons With Serious Mental Illness

by Arthur J. Lurigio and James A. Swartz

Assessing Correctional Rehabilitation: Policy, Practice, and Prospects

by Francis T. Cullen and Paul Gendreau

The Evolution of Decisionmaking Among Prison Executives, 1975–2000

by Kevin N. Wright

Community Justice and a Vision of Collective Efficacy: The Case of Restorative Conferencing

by Gordon Bazemore

Community Policing in America: Changing the Nature, Structure, and Function of the Police

by Jack R. Greene

Criminal Justice and the IT Revolution

by Terence Dunworth

Thirty Years of Sentencing Reform: The Quest for a Racially Neutral Sentencing Process

by Cassia C. Spohn

The Convergence of Race, Ethnicity, Gender, and Class on Court Decisionmaking: Looking Toward the 21st Century

by Marjorie S. Zatz

National Criminal Justice Reference Service (NCJRS)

Box 6000

Rockville, MD 20849-6000

Volume 4. Measurement and Analysis of Crime and Justice

Introduction to Volume 4

Measurement and Analysis of Crime and Justice:

An Introductory Essay

*by David Duffee, David McDowall, Lorraine Green Mazerolle, and
Stephen D. Mastrofski*

The Self-Report Method for Measuring Delinquency and Crime

by Terence P. Thornberry and Marvin D. Krohn

**Self-Report Surveys as Measures of Crime and Criminal
Victimization**

by David Cantor and James P. Lynch

Theory, Method, and Data in Comparative Criminology

by Gregory J. Howard, Graeme Newman, and William Alex Pridemore

Spatial Analyses of Crime

by Luc Anselin, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita

Measuring the Costs and Benefits of Crime and Justice

by Mark A. Cohen

**Measuring the Sexual Victimization of Women: Evolution, Current
Controversies, and Future Research**

by Bonnie S. Fisher and Francis T. Cullen

**Measurement and Analysis of Drug Problems and Drug
Control Efforts**

by Jonathan P. Caulkins

Fear of Crime in the United States: Avenues for Research and Policy

by Mark Warr

**Measurement and Explanation in the Comparative Study of
American Police Organizations**

by Edward R. Maguire and Craig D. Uchida

Standards and Measures of Court Performance

by Ingo Keilitz

About the National Institute of Justice

The National Institute of Justice (NIJ), a component of the Office of Justice Programs, is the research agency of the U.S. Department of Justice. Created by the Omnibus Crime Control and Safe Streets Act of 1968, as amended, NIJ is authorized to support research, evaluation, and demonstration programs; development of technology; and both national and international information dissemination. Specific mandates of the Act direct NIJ to:

- Sponsor special projects and research and development programs that will improve and strengthen the criminal justice system and reduce or prevent crime.
- Conduct national demonstration projects that employ innovative or promising approaches for improving criminal justice.
- Develop new technologies to fight crime and improve criminal justice.
- Evaluate the effectiveness of criminal justice programs and identify programs that promise to be successful if continued or repeated.
- Recommend actions that can be taken by Federal, State, and local governments as well as by private organizations to improve criminal justice.
- Carry out research on criminal behavior.
- Develop new methods of crime prevention and reduction of crime and delinquency.

In recent years, NIJ has greatly expanded its initiatives, the result of the Violent Crime Control and Law Enforcement Act of 1994 (the Crime Act), partnerships with other Federal agencies and private foundations, advances in technology, and a new international focus. Examples of these new initiatives include:

- Exploring key issues in community policing, violence against women, violence within the family, sentencing reforms, and specialized courts such as drug courts.
- Developing dual-use technologies to support national defense and local law enforcement needs.
- Establishing four regional National Law Enforcement and Corrections Technology Centers and a Border Research and Technology Center.
- Strengthening NIJ's links with the international community through participation in the United Nations network of criminological institutes, the U.N. Criminal Justice Information Network, and the NIJ International Center.
- Improving the online capability of NIJ's criminal justice information clearinghouse.
- Establishing the ADAM (Arrestee Drug Abuse Monitoring) program—formerly the Drug Use Forecasting (DUF) program—to increase the number of drug-testing sites and study drug-related crime.

The Institute Director establishes the Institute's objectives, guided by the priorities of the Office of Justice Programs, the Department of Justice, and the needs of the criminal justice field. The Institute actively solicits the views of criminal justice professionals and researchers in the continuing search for answers that inform public policymaking in crime and justice.

To find out more about the National Institute of Justice,
please contact:

National Criminal Justice Reference Service,
P.O. Box 6000
Rockville, MD 20849-6000
800-851-3420
e-mail: askncjrs@ncjrs.org

To obtain an electronic version of this document, access the NIJ Web site
(<http://www.ojp.usdoj.gov/nij>).

If you have questions, call or e-mail NCJRS.

