

CONTENTS

VOLUME 12 • ISSUE 4 • NOVEMBER 2013

Special Issue

EDITOR'S PREFACE

- Addressing Causal Uncertainty in the Application of Criminological Research to
Public Policy569
Thomas G. Blomberg

CRIMINOLOGY, CAUSALITY, AND PUBLIC POLICY

EDITORIAL INTRODUCTION

- Seeking Causality in a World of Contingency: Criminology, Research,
and Public Policy571
Thomas G. Blomberg, Julie Mestre, and Karen Mann

EXECUTIVE SUMMARY

- Overview of: "Translating Causal Claims: Principles and Strategies for
Policy-Relevant Criminology"585
Robert J. Sampson, Christopher Winship, and Carly Knight

RESEARCH ARTICLE

- Translating Causal Claims: Principles and Strategies for Policy-Relevant
Criminology587
Robert J. Sampson, Christopher Winship, and Carly Knight

POLICY APPLICATION ESSAYS

- Family-Focused Interventions to Prevent Juvenile Delinquency: A Case Where
Science and Policy Can Find Common Ground617
Abigail A. Fagan

- Evidence and Public Policy: The Example of Evaluation Research in Policing651
Daniel S. Nagin and David Weisburd

- Supermax Prisons: The Policy and the Evidence681
Daniel P. Mears

AFTERWORD

- Linking Evidence and Criminal Justice Policy721
Alfred Blumstein

Guide to Preparing Manuscripts

Editorial Policy—*Criminology & Public Policy* (CPP) is a peer-reviewed journal devoted to the study of criminal justice policy and practice. The central objective of the journal is to strengthen the role of research findings in the formulation of crime and justice policy by publishing empirically based, policy-focused articles. Authors are encouraged to submit papers that contribute to a more informed dialogue about policies and their empirical bases. Papers suitable for CPP not only present their findings, but also explore the policy-relevant implications of those findings. Specifically, appropriate papers for CPP do one or more of the following:

- Strengthen the role of research in the development of criminal justice policy and practice
- Empirically assess criminal justice policy or practice, and provide evidence-based support for new, modified, or alternative policies and practices
- Provide more informed dialogue about criminal justice policies and practices and the empirical evidence related to these policies and practices
- Advance the relationship between criminological research and criminal justice policy and practice

The policy focus of the journal requires articles with a slightly different emphasis than is found in most peer-reviewed academic journals. Most academic journals look for papers that have comprehensive literature reviews, provide detailed descriptions of methodology, and draw implications for future research. In contrast, CPP seeks papers that offer literature reviews more targeted to the problem at hand, provide efficient data descriptions, and include a more lengthy discussion of the implications for policy and practice. The preferred paper describes the policy or practice at issue, the significance of the problem being investigated, and the associated policy implications. This introduction is followed by a description and critique of pertinent previous research specific to the question at hand. The methodology is described briefly, referring the reader to other sources if available. The presentation of the results includes only those tables and graphs necessary to make central points (additional descriptive statistics and equations are provided in appendices). The paper concludes with a full discussion of how the study either provides or fails to provide empirical support for current, modified, or new policies or practices. The journal is interdisciplinary, devoted to the study of crime, deviant behavior, and related phenomena, as found in the social and behavioral sciences and in the fields of law, criminal justice, and history. The major emphases are theory; research; historical issues; policy evaluation; and current controversies concerning crime, law, and justice.

Manuscript Submissions—Manuscripts are to be submitted electronically to cpp@fsu.edu. The manuscript should be submitted in one Word (.doc) file with tables and figures in the same document as the manuscript text. Additional documents, including cover letters or memos to the editor, may also be e-mailed as supplemental files. Although we strongly encourage on-line submission, those who prefer not to submit on-line may send a CD to Julie Mestre Brancale, Managing Editor, Florida State University, Center for Criminology & Public Policy Research, 325 John Knox Road Building L-102, Tallahassee, FL 32303.

An executive summary of approximately 150 words and a brief biographical paragraph describing each author's current affiliation, research interests, and recent publications should accompany the manuscript.

Papers accepted for publication should comply with American Psychological Association guidelines concerning nonsexist language. We accept three formats for digital artwork submission: Encapsulated PostScript (EPS), Portable Document Format (PDF), and Tagged Image Format (TIFF). We suggest that line art be saved as EPS files. Alternately, these may be saved as PDF files at 600 dots per inch (dpi) or better at final size. Tone art, or photographic images, should be saved as TIFF files with a resolution of 300 dpi at final size. For combination figures, or artwork that contains both photographs and labeling, we recommend saving figures as EPS files, or as PDF files with a resolution of 600 dpi or better at final size. More detailed information on the submission of electronic artwork can be found at:

<http://authorservices.wiley.com/bauthor/illustration.asp>

The American Society of Criminology regards submission of a manuscript to this journal as a commitment to publish herein; simultaneous submission to another journal is unacceptable. Every effort will be made to notify authors of editorial decisions within 3 months of manuscript and processing fee receipt. Note that CPP publication decisions arrive by postal mail, so please indicate whether you prefer e-mail correspondence to postal mail upon submission of your manuscript.

Please consult and conform to the CPP Style Sheet, which is available at cpp.fsu.edu, prior to submission.

Criminology & Public Policy (Print ISSN #1538-6473/On-line ISSN #1745-9133) is published quarterly on behalf of the American Society of Criminology, 1314 Kinnear Road, Suite 212, Columbus, OH 43212 by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774.

Information for subscribers. *Criminology & Public Policy* is published in four issues per year. Institutional print and on-line subscription prices for 2013 are: US\$328 (U.S.), US\$403 (rest of world), €258 (Europe), £206 (U.K.). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2009, where available. For other pricing options, including access information and terms and conditions, please visit www.wileyonlinelibrary.com/access. A subscription to *Criminology & Public Policy* also includes 4 issues of *Criminology*.

Delivery Terms and Legal Title. Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability.

Journal customer service. For ordering information, claims, and any inquiry concerning your journal subscription, please visit www.wileycustomerhelp.com or contact your nearest office. *Americas* e-mail: cs-journals@wiley.com; telephone: +1 781 388 8598 or +1 800 835 6770 (toll free inside the U.S. and Canada). *Europe, Middle East, and Africa* e-mail: cs-journals@wiley.com; telephone: +44 (0) 1865 778315. *Asia Pacific* e-mail: cs-journals@wiley.com; telephone: +65 6511 8000. *Japan* For Japanese speaking support, email: cs-japan@wiley.com; telephone: +65 6511 8010 or telephone (toll free): 005 316 50 480. **Visit our Online Customer Get-Help** available in 6 languages at www.wileycustomerhelp.com.

Membership. American Society of Criminology membership is available to all interested people. Individuals interested in joining should complete and return the application forms on the American Society of Criminology Web site (www.asc41.com). Membership in the American Society of Criminology includes a subscription to *Criminology* and *Criminology & Public Policy*.

Back issues. Information about the availability and prices of back issues can be obtained from the American Society of Criminology by e-mail (asc@asc41.com) or telephone (614-292-9207).

Mailing and on-line access. Journal Customer Services, John Wiley & Sons, Inc., 350 Main St., Malden, MA 02148-5020. Postmaster: Send all address changes to *Criminology & Public Policy*. View this journal online at wileyonlinelibrary.com. For submission instructions, subscription and all other information, visit [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1745-9133](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1745-9133).

Access to this journal is available free online within institutions in the developing world through the AGORA initiative with the FAO, the HINARI initiative with the WHO and the OARE initiative with UNEP. For information, visit www.aginternetwork.org, www.healthinternetwork.org, and www.oarescience.org. *Criminology & Public Policy* is published by Wiley Periodicals, Inc., Commerce Place, 350 Main Street, Malden, MA 02148; Tel: (781) 388-8200; Fax: (781) 388-8210.

ISSN 1538-6473 (Print)

ISSN 1745-9133 (Online)

Wiley contacts. Ang Wen Ting, Production Editor (e-mail: CAPP@wiley.com); Kristin McCarthy, Advertising (e-mail: kmccarthy@wiley.com).

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. Since launching the initiative, we have focused on sharing our content with those in need, enhancing community philanthropy, reducing our carbon impact, creating global guidelines and best practices for paper use, establishing a vendor code of ethics, and engaging our colleagues and other stakeholders in our efforts. Follow our progress at www.wiley.com/go/citizenship.

Copyright © of The American Society of Criminology. All rights reserved. No portion of the contents of *Criminology & Public Policy* may be reproduced without the permission of the American Society of Criminology. Please forward all rights and permission requests to the American Society of Criminology, 1314 Kinnear Road, Suite 212, Columbus, OH 43212.

Disclaimer. The Publisher, the American Society of Criminology and Editor cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, the American Society of Criminology and Editor.

Printed in the USA by The Sheridan Group.

Addressing Causal Uncertainty in the Application of Criminological Research to Public Policy

Thomas G. Blomberg

Florida State University

With the publication of this Issue 4, Volume 12, I complete 6 years as Editor of *Criminology and Public Policy* (CPP). Throughout this period, a persistent challenge in CPP's efforts to advance the role of research in public policy has been how to deal with causal uncertainty and/or largely contingent research findings. Specifically, in the peer review of manuscripts submitted for publication consideration to CPP, the primary deliberation has been upon evaluations of the degree to which manuscripts provide sufficiently compelling causal arguments to justify new or innovative policy recommendations. The challenge is magnified, in part, because the standards used by scholars for determining compelling causal arguments are not uniform or clear. This special issue addresses this challenge and associated ambiguity. The issue includes: (1) an introduction to criminology's ongoing quest to establish causality; (2) an assessment of principles and strategies for informing policy in a causally uncertain and contingent world; (3) individual applications of a proposed approach to the areas of crime prevention, law enforcement, and imprisonment; and (4) a concluding assessment of the capacity of this proposed approach for advancing the use of research in public policy. The specific intent of this special issue is to contribute to the ongoing dialogue and quest to develop greater clarity and consistency in how criminologists can more effectively apply their research to public policy despite ongoing causal uncertainty.

Since its inception 12 years ago, CPP has emerged as one of criminology's premier journals. In 2010, CPP applied for inclusion into ISI and early this summer CPP was accepted into ISI's Social Science Index. Additionally, in 2011, CPP began developing timely special policy issues that are featured at annual Congressional luncheons in Washington, D.C. To date, these special policy issues and Congressional luncheons have addressed: (1) Imprisonment and Crime (2011); (2) A Signaling Perspective on Employment-Based

Direct correspondence to Thomas G. Blomberg, Florida State University, College of Criminology and Criminal Justice, Eppes Hall, 145 Convocation Way, Tallahassee, FL 32306 (e-mail: tblomberg@fsu.edu).

Reentry (2012); and (3) this issue of *Criminology, Causality and Public Policy* (2013). Moreover, in 2011, the American Society of Criminology's Best Article Award was given to William Spellman's CPP article, "Crime, Cash and Limited Options: Explaining the Prison Boom," (Vol. 8(1), 2009). Clearly, while challenges remain in the continuing effort to advance research-informed public policy and practice, CPP is making important progress.

I would like to express my sincere appreciation to the editorial board and the numerous reviewers that I have had the pleasure of working with over the past 6 years. My thanks to Chris Eskridge for his excellent assistance and support with the multiple logistical arrangements for CPP's annual Congressional luncheons. Additionally, my thanks and gratitude to Ms. Julie Mestre, CPP's Managing Editor, for her outstanding editorial skills, enthusiasm, and hard work. I believe we can all look forward to CPP's continued "onward and upward" trajectory under the capable leadership of CPP's new Co-Editors Bill Bales and Dan Nagin.

EDITORIAL INTRODUCTION

CRIMINOLOGY, CAUSALITY, AND PUBLIC POLICY

Seeking Causality in a World of Contingency Criminology, Research, and Public Policy

Thomas G. Blomberg

Julie Mestre

Karen Mann

Florida State University

In recent years there has been growing recognition of the increasing importance of applying criminological research to public policy. This has been reflected in several initiatives taken by the American Society of Criminology (ASC) including its establishment of *Criminology and Public Policy* (CPP) in 2000 to strengthen the role of research in policy and practice. Additionally, the ASC is now sponsoring annual Congressional luncheons where current research on timely crime and justice policy and practice topics are presented to Congressional members and staffers. Moreover, in 2009, the ASC formed a consortium with the Academy of Criminal Justice Sciences (ACJS) and the Association of Doctoral Programs in Criminology and Criminal Justice, which represents approximately 4,000 criminologists and related behavioral scientists and nearly 40 doctoral programs, in an effort to establish a stronger research involvement in national crime and justice policy. A particularly noteworthy initiative by the consortium was the successful advocacy for increased independence in the grants-making and publication processes of the National Institute of Justice (NIJ), the Bureau of Justice Statistics (BJS), and the Office of Juvenile Justice and Delinquency Prevention (OJJDP).

Nonetheless and despite these unprecedented efforts to elevate the role of criminological research in public policy, an important question remains. Namely, will criminologists be able to successfully integrate their traditional research and science role with an emerging public policy role? A number of criminologists argue that because criminology's research and associated knowledge base is not causally certain, they should not and cannot responsibly inform public policy. In fact, some even caution that criminologists could end up doing

Direct correspondence to Thomas G. Blomberg, Florida State University, College of Criminology and Criminal Justice, Eppes Hall, 145 Convocation Way, Tallahassee, FL 32306 (e-mail: tblomberg@fsu.edu).

more harm than good by promoting the application of incomplete research knowledge to public policy. However, other criminologists who are actively engaged in research and public policy share the belief that the use of the best available research knowledge in public policy is a logical and necessary alternative in the absence of causal certainty upon which to inform criminal justice policy (Latessa, Cullen, Gendreau, 2002).

The purpose of this special issue is to address the criminological question of how best to advance criminal justice policy and practice in the absence of causal certainty while simultaneously employing rigorous standards of scientific methodology.

Criminology's Development as a Scientific Discipline

In its disciplinary beginnings, at the turn of the 20th century, criminology was focused upon informing policy initiatives. For example, the Chicago School expressly sought the development of descriptive and explanatory insights to inform policies that might ameliorate the perceived criminogenic features of Chicago's slum communities and thereby hopefully reduce crime (Burgess and Bogue, 1967). The underlying criminological belief of this period was that by determining the causes of crime, associated prevention and rehabilitation policies would follow (Allen, 1964; Blomberg and Lucken, 2010). The determination of specific antecedent causes to criminal behavior that were to guide subsequent prevention, treatment, and rehabilitation policies and practices was an ideal that has yet to be accomplished.

During much of the subsequent 20th century, the discipline became more focused upon establishing itself as a recognized scientific discipline through the discovery and validation of causal explanations of crime. For example, following the Chicago School, a series of theories on the causes of crime were provided by Merton (1938), Sutherland (1947), Cohen (1955), Cloward and Ohlin (1960), among others. Throughout this period, it was generally assumed that the criminal justice system operated according to its formally prescribed goals and practices with "disinterested professionalism," and was therefore, not subject to sustained study of reform policy initiatives (Skolnick, 1965).

Despite the prominence of the 20th Century Rehabilitative Ideal, beginning in the late 1960s, the United States witnessed the patterned inability of the criminal justice system to legally and effectively control crime as well as the period's unprecedented civil disobedience. This realization led to a series of "critical law in action" studies (Lemert, 1970; Messinger, 1969; Skolnick, 1965, 1966; Sudnow, 1965; among others) that documented patterned "goal versus practice" disparities of our criminal justice agencies.

In 1968, the U.S. President's Commission on Law Enforcement and Administration of Justice published *The Challenge of Crime in a Free Society* (1967). The prevailing thinking was that crime, while a problem with local origins and impacts, was also a national problem that required a national strategy or a "War on Crime" (Feeley and Sarat, 1980). Subsequently, the Law Enforcement Assistance Administration (LEAA) was established following federal passage of the Omnibus Crime Control and Safe Streets Act

of 1968 to assist in implementing the policy recommendations made by the President's Commission.

During the next decade, LEAA operated with annual budgets that averaged approximately \$800 million to administer block grants to states and local jurisdictions that were intended to assist in implementing the specific policy recommendations of the President's Commission. While much has been written about LEAA's failure to successfully win the war against crime, little has been written about the agency's role in accelerating the growth of academic criminology. Through LEAA's Law Enforcement Educational Program (LEEP), college and university students received substantial loans to fund their undergraduate and graduate educations. LEEP and LEAA's funding contributed to a significant expansion in the number of college and university criminology programs.

Given that criminology was *not* generally recognized as an established academic or scientific discipline and its subject matter had applied and policy implications, the precise mission of university and college criminology programs has and continues to generate debate and divisions among individual program faculty and students. The early goal of some university programs was centered upon applied concerns while other programs were more focused upon research and theory related to crime's causes. Today, there are 38 university PhD programs in criminology (Association of Doctoral Programs in Criminology & Criminal Justice, 2013), and as many senior criminological researchers can attest, their graduate training was likely centered upon theory, research methods, and causality rather than ways to communicate and/or apply criminological research knowledge to policy initiatives. This graduate training largely continues today.

Criminology's focus upon scientific development and recognition is reflected by the strong professional interest in the *US News and World Report* rankings of Criminology PhD programs and studies of faculty article counts, citations, and external funding levels, etc. (Kleck and Barnes, 2011; Mustaine and Tewksbury, 2009). Yet, there has been little serious recognition of criminology's public policy accomplishments (Cullen, 2005; Petersilia, 1991). However, this now decades-old focus upon increasing criminology's "scientific stature" has become a bit blurred in more recent years with growing interest in public policy. This is evidenced by the Stockholm Award in Criminology given for outstanding achievement in criminological research and the application of research results in reduction of crime policies, the growing prominence of ASC's journal, *CPP*, and the previously mentioned policy related efforts by ASC, ACJS, and the Association of Doctoral Programs in Criminology and Criminal Justice.

Arguably, the discipline of criminology is becoming more secure in its scientific standing and increasingly recognizes that, for example, the quest for the "root cause(s)" of crime is unending (Rein and Winship, 2000). It appears that a growing number of criminologists have concluded that the discipline's scientific development and associated quest for causality, while an ongoing process, can, nonetheless, responsibly influence policy with what is now known versus what might be known in the future.

Some Consequences of Causality as the Empirical Standard in Criminology

An illustration of the mixed sentiment among criminologists concerning the causality and policy question occurred at the 2009 Annual Meeting of the ASC. The meeting's theme was titled "Criminology and Criminal Justice Policy," and featured a series of plenary sessions on various perspectives, challenges, and possible remedies for increasing criminology's policy role. A panel paper by Wellford (2009) was provocatively titled, "Criminologists Should Stop Whining about Their Impact on Policy and Practice." His argument was that criminological research has not been sufficiently established as a science to inform policy and practice responsibly.

This general sentiment is shared by a number of other criminologists and reflects some of the reasoning developed earlier by Tittle (2004). Tittle (2004) argued that criminologists simply do not have sufficient knowledge to responsibly inform public policy. He suggested that existing criminological knowledge is shaky at best and poses more dangers for policy formulation than beneficial prospects. To illustrate his point, Tittle (2004: 1641) claimed that there is not a single criminological "issue about which even a modestly demanding critic could be convinced by the available research evidence." Tittle cited such examples as: What causes crime? Does arresting domestic violence abusers deter future domestic violence? Does gun control prevent violence? Does the death penalty curb capital crime? Tittle concluded that criminologists are as likely to be wrong as they are to be right in the process of applying their research to policy initiatives, therefore not only doing no good but instead doing potential harm in the process because they do not possess research that provides the sufficient cause-effect evidence essential for responsibly guiding crime related policy.

In a related discussion, Burawoy's American Sociological Association's (ASA) Presidential Address, "For Public Sociology" in 2004, generated similar debate within sociology. According to Burawoy, engaging various "publics" is critical to sharing information, broadening understandings, and increasing what we know about the world. However, he also acknowledged the "primitive stage" that sociologists are in regarding how to effectively engage "publics." Burawoy (2005) discussed three possible justifications for sociologists to engage in politics or initiate a public policy stance: professional self-defense, policy intervention, and public engagement. The concept discussed within this article is parallel to what he refers to as "public engagement."

Tittle (2004) argued, in response to Burawoy's call for public engagement, that Burawoy was making an incorrect assumption, namely that sociologists actually have good knowledge that can be successfully applied to public policy. Wellford (2009) echoed Tittle's argument and further speculated that criminology may have had too much policy influence, rather than too little, given its limited stage of scientific development and lack of causal knowledge that enables more certain explanations and predictions to inform effective policy. Wellford urges criminologists to focus on criminology as a science and to refrain from any meaningful policy role until such time as criminology becomes recognized as a

valid scientific discipline, one that possesses more specific and empirically validated causal arguments.

The debate, then, within criminology is not over the importance of criminology's science being applied to policy development. Rather, the debate is centered upon when criminology's research will be sufficiently established as a science before it is used toward that end. For some criminologists the fundamental standard for establishing scientific criminology and a legitimate policy role relates to the capacity to determine cause-effect relationships. For other criminologists, the necessary threshold for a crime-related policy function is not the requirement to identify cause-effect relationships. Rather, it is the application of current and best available research knowledge to policy-relevant questions as they arise. This disciplinary debate can be characterized as revolving around the question: "is the pursuit of perfect the enemy of the pursuit of good?"

Since its positivist beginnings, much of criminology has been preoccupied with scientifically identifying the causes of crime. Underlying this quest for causality is the assumption that the occurrence of events, like crime, is determined by a cause and effect relationship (i.e., X causes Y, and if X causes Y, how large is the effect of X on Y? And is this effect larger than other causes?). For criminology, as with other social and many natural sciences, it has not yet been possible to isolate a single cause of crime. Rather, criminologists have found that in the case of delinquency, there are many contingent factors that increase the likelihood that delinquency will occur (e.g., poverty, living in slums, negative peer group associations, dropping out of school, broken homes) rather than a single cause that will lead to delinquency. However, such multiple correlations do not provide a predictive function (i.e., not all youth living in slums, with negative peer associations, lacking a high school diploma, and from broken homes are delinquent and/or some delinquents are not from the slums, or school drop-outs, etc.). The presence of multiple factors does not necessarily mean that distinct casual effects cannot be identified. A common policy struggle, then, is the identification of the most appropriate and alterable risk factor that a particular policy or practice should address. To further complicate this process, the numerous factors that influence an individual's decision to commit a crime do not exist in a vacuum—they co-exist and co-occur with other variables and create a research environment that is impossible to fully control or account for all nonconstant factors.

As previously stated, the discipline's lack of causal certainty in its research findings keeps many criminologists away from attempting to have a serious public policy role. Regarding the limits of causal analysis, Rein and Winship (2000) have concluded that social science produces weak causal theories with modest effect sizes that can explain only a small portion of the variability in the dependent variable. The authors reference Jon Elster's (1991) argument that complete explanations may never be forthcoming in the social sciences (2000: 36). Rein and Winship (2000) also highlight the dilemma of identifying the most appropriate factor or correlate that policy should address when the causal modeling involves multiple independent variables. Moreover, policies that target correlates of crime or delinquency risk

factors can be limited because some of these correlates and risk factors are not alterable (i.e., age, gender, race). Given the limitations of trying to link policies with validated causality, Rein and Winship (2000) argue for research-driven policies that, as mentioned earlier, target the problem itself rather than its causes. Research-driven policies that target the problem would have two objectives: (1) to directly intervene and (2) to more effectively ameliorate the alterable conditions of the problem in question rather than eliminating its uncertain causes. While challenges remain, it seems that criminology is poised to expand its disciplinary focus of causal analysis to include linking its existing and best available research knowledge to policy aimed at ameliorating the malleable conditions of the problem rather than the single pursuit of identifying and eliminating the uncertain causes of the problem. The question is how can criminologists simultaneously pursue causal and public policy functions?

The argument guiding this special issue is that criminologists can and should continue to pursue their scientific interests in causality while simultaneously pursuing ways to inform crime-related policy with the best available research knowledge. Regarding their policy role, the argument suggests that criminologists “target the problem” which may not directly address causes *per se* but rather the proximate and malleable manifestations of the problem in question. An example concerns the role of educational achievement for incarcerated delinquent youth upon post release recidivism. While no direct cause and effect relationships have been established in this area of research, the best available research knowledge documents that delinquent youth who enter incarceration with numerous educational deficiencies and experience substantial educational achievement through the use of highly qualified and certified teachers providing individualized instruction are more likely to return to school upon release. Further, the research demonstrates that if youth remain in school following release, their likelihood of recidivism is substantially reduced. However, current national policy and practice in juvenile justice education can be characterized as grossly uneven and inferior in quality (i.e., low employment of certified teachers, lack of individualized instruction, etc.) particularly compared to that in public schools. Yet, increasing educational achievement through quality and individualized instruction while delinquents are incarcerated increases their likelihood of return to school upon release thereby reducing their recidivism. Employing this best available research knowledge to current juvenile justice education policy and practice seems clearly preferable to waiting until more causal certainty is established among incarceration, educational achievement, and recidivism (Blomberg, Bales, Mann, Piquero, and Berk, 2011; Blomberg, Bales, and Piquero, 2012).

Challenges for Criminology, Causality, and Public Policy

Combining scientific research roles with public policy roles creates potential tensions or complications for criminologists. It requires them to be accessible and to confront immediate issues in current headlines, which come in the form of journalists' questions, and urgent e-mail inquiries—while simultaneously being comprehensive, reflexive, and

objective. Moreover, it is difficult to appear before state legislative or congressional bodies and deliver compelling research and policy arguments in the often-restricted time limits of, for example, five minutes before the U.S. Congress, or to provide concise answers to pressing policy questions from research that is far from definitive or causally conclusive.

Because so much of the evidence produced by scholarly work involves contingent and nuanced relationships, this often renders it less understandable to larger public, legislative, or policymaking bodies. Very importantly, the linkage of scholarship and public policy involves a delicate balance: it is not necessarily a matter of identifying and advocating for a particular policy or practice; rather, it often consists of identifying and explaining the choices and likely consequences involved in various policy options from the best available and relevant research knowledge. Certainly, there are areas of criminological research that enable informed and helpful presentations of the likely consequences associated with various policy options even though this research does not provide specific cause-effect conclusions. Some of these research areas include: mass incarceration; supermaximum prisons; punitive drunk driving laws; prison drug treatment, aftercare, and recidivism; and electronic surveillance, among others.

How, then, can criminology's research and knowledge responsibly inform public policy? Certainly, our theories of the causes of crime indicate that no single cause or associated policy such as punitiveness or "getting tough" will prove successful. Further, our studies of the criminal justice system have demonstrated that while it is difficult to appropriately implement reforms because of common impediments—professional resistance, politics, bureaucratic obstacles, ideological conflicts, etc.—it is possible to overcome these impediments and implement reforms with fidelity, evaluate the results, and identify the likely outcomes of these reforms. Our knowledge of crime's correlates and the operations and tendencies of the criminal justice system can provide helpful information in crime and justice policy decisions thereby simultaneously advancing criminology and public policy. By analogy, while the medical profession knows much about cancer, it does not have a scientifically proven cause and associated cure for this disease. Rather, the medical profession conducts research on government monitored drugs and surgical procedures that have themselves been developed from prior research and have been shown to make a positive and even life-and-death difference for afflicted patients with cancer despite the lack of validated causal findings. In sum, the medical profession does, in fact, target the problem with what is currently known in the effort to ameliorate the problem's conditions as recommended for the social sciences by Rein and Winship (2000).

We contend that criminologists should follow a similar "target the problem" ethic in their efforts to apply the best available research knowledge to public policy. Criminologists should be involved in sharing what they do and do not know in the effort to advance crime and justice policies. As medical knowledge has limitations on the causes of cancer and numerous other diseases, criminology and public policy is not an all or nothing proposition. Rather, it is possible to improve crime and justice policy decisions with what we now know

in ever-advancing and incremental steps as more scientific knowledge and causal inferences become available.

To be effective in a proactive role that informs public policy with the best available research knowledge, criminologists will need to shift from their traditional role on campus, the classroom, and their comfortable relationships with like-thinking professional colleagues and invoke more informed strategies. Researchers will need to increase their tacit knowledge about all aspects of the utility of their research. Awareness of the mechanisms by which policy makers and practitioners *acquire* and *interpret* research and the ways in which policy makers and practitioners *use* research evidence will facilitate efforts in such a pursuit.

Policy makers and practitioners acquire research through a number of channels. To inform policy with research, it is important to be familiar with the intermediary organizations that broker and disseminate research evidence (e.g., advocates, lobbyists, nonpartisan private organizations, foundations). It is also important to anticipate the most likely political, economic, and social context in which the process occurs. Further, researchers should anticipate that within the political context, a research finding will not remain static with one particular meaning or interpretation. Policy makers and practitioners are continuously interpreting new information and integrating it into their respective knowledge/ideological frameworks. Policy makers and staff routinely acquire new research articles or excerpts of findings and attempt to use that evidence to justify positions, understand problems, and inform their decisions. In some instances, research findings will be misinterpreted, partially extracted, or otherwise misused in the process. As policy formulations are influenced by research evidence and diffused across social networks, individuals who are part of the networks are interpreting (and reinterpreting) the ideas and research evidence. Similarly, as research evidence is used in organizational decision making, its meaning is being interpreted and reinterpreted in relation to local needs, contexts, and constraints. Knowing the various channels, strategies, and processes explicitly or implicitly employed to make policy makers aware of research evidence will assist researchers to frame their findings so that they can be more clearly understood and more responsibly incorporated into policy.

Once research evidence is known to policy makers and practitioners, criminologists should be familiar with the ways in which these nonacademics *use* research evidence and the role that research evidence plays in policy and practice. There are, no doubt, various ways in which research can be used by policy makers and practitioners. However, there are at least four examples of the use of research that are relevant to this discussion—tactical, imposed, instrumental, and conceptual (Nutley, Walter, Davies, 2007; Weiss, 1979). *Tactical use*, where evidence may play more of a passive role, is when research evidence is invoked to justify an existing policy, position, practice, reform initiative, legislation, or other stance. This is more of a symbolic or strategic use of evidence rather than a proactive, preemptive, or explicit use. A different recent trend with federal program funding awards is to mandate that entities use such funding to implement “evidence-based practices” or “proven practices” (e.g., “what works research” and the “what works clearinghouse”). This type of research use is

an *imposed use*. Likewise, when empirical evidence is directly applied to the programmatic or policy decision-making process, it can be considered an *instrumental use*. If research evidence informs or influences policy makers' and practitioners' interpretations and thought processes of problems, emerging challenges, or situations, it is considered to be a *conceptual use*.

A singularly frustrating occurrence that will be inevitable in some policy settings, is when policy makers ignore empirical evidence and direct policy in the opposite direction because it is more politically viable (e.g., refusal to reverse their own position; hesitation to vote against the will of the majority or their constituents' sentiments; fear of appearing soft on crime; or because they are reacting to a highly visible media account of an isolated incident that, in their opinion, demands a certain response). This type of resistance will occur even in the face of implementing a more costly policy—corrections and criminal justice are areas in which, sometimes, costs take a back seat to the message being sent to the public. This will, in all likelihood, frustrate and discourage researchers; however, it should not be interpreted as a sign that practitioners and policy makers are not open to evidence.

Knowing how policy makers and practitioners interpret, process, or make sense of empirical evidence is also important because it provides guidance regarding the most policy relevant way to report research findings. The risk with this task is that interpretations will vary according to individual characteristics and values, and with broader or macro-level factors such as political climate, context, fiscal constraints, and visibility of the issue. Anticipating the way in which these factors affect interpretations of the relevance, validity, meaning, or implications of research evidence and how policy makers appraise the quality and nature of the evidence will also serve the researcher.

Drawing on experience with the Florida legislature, the entire process of making policy makers aware of evidence, guiding their interpretation of evidence, and assisting policy makers in using it is certainly time consuming and requires a substantial commitment on the part of the researcher. Given the existing pressure for scholarly research, stellar teaching, and service, academic researchers may not be able to commit the necessary time in order to effectively impact policy initiatives. However, effort can be consolidated and shared within departments and organizations. It is hard to ignore the fact that the single most critical element of being effective in this role is establishing a relationship with policy makers, practitioners, and legislative staff. A relationship will open doors and facilitate the remainder of the process. Another important component of this process is to summarize empirical research findings into very specific actionable policy recommendations that include, if possible, proposed policies and/or suggestions for amending existing statutes. Researchers must marshal as many relevant resources as possible in order to effectively penetrate the barriers and impediments surrounding the policy-making processes.

The executive and legislative branches of government (federal and state legislatures and governors) are the primary “drivers” of criminal justice policy. Criminologists

generally have limited access to these “drivers” while criminal justice agencies are required to provide information, data, and testimony to governors and legislative committees on a regular basis. For example, when proposed criminal justice legislation is being considered/debated in the legislature, legislators routinely call upon administrators and staff from applicable criminal justice agencies to give testimony and provide supporting information related to the substantive and fiscal impact of the proposed legislation. To the extent that criminologists collaborate with criminal justice agencies to apply research knowledge on the policy issues under consideration, the greater the likelihood that these policy decisions will be based upon the best available research knowledge.

Accomplishing such a task is a challenge during state legislative sessions due to time constraints and the pace with which legislation is drafted, considered, amended, and adopted. Whether at the local, state, or national levels, policy making is often hurried, rash, reactive, and shortsighted. For example, the Florida legislative session allows 60 days for debate, amendment, voting, and re-amendment of hundreds of pieces of legislation—often, the rushed pace handicaps the vetting process for critical issues and there is limited consideration of research evidence. Clearly, the presentation of relevant research findings, with full disclosure of the methodological limitations, is preferable to hurried, reactive and/or politicized policy decisions that are far too often based upon perceptions rather than the best available research knowledge.

Argument for Criminology, Causality, and Public Policy

Considering the costly fiscal impact of many criminal justice and corrections policies and practices coupled with the current limited and still shrinking fiscal capacity, the situation demands that policy initiatives be evidence-based. In time and with effort, citizens may become even more inclined to demand that state revenue—their tax dollars—be used to support only those programs where positive outcomes have been established. For example, rather than blindly and precipitously mandating the privatization of prisons on a massive scale, which recently occurred in Florida, as a result of the allure of cost savings that have yet to be demonstrated elsewhere, it would be more prudent and fiscally responsible to first conduct an empirical evaluation comparing state-operated prisons versus privately operated prisons. Putting aside the fact that criminal justice policies strive to ensure the safety of citizens and that they tend to infringe upon individual rights and freedoms (of the accused) as the most compelling reasons to enact only evidence-based policies and practices, the scarcity of resources and current downward economic spiral should compel researchers, policy makers, and administrators in this direction.

Despite the absence of causal certainty, empirical evidence continues to make valuable contributions to criminology’s body of knowledge which, in turn, can responsibly inform public policy. However, caution must be exercised to ensure accurate accounts of the weight of the evidence or the best available research knowledge on a particular topic in terms

of what is explained and what is not explained. In his discussion of the application of sociological research and public policy, Burawoy (2005: 1605) asserts that sociologists can support a specific policy “on the basis of an accumulated body of evidence” Certainly, a body of relevant criminological research can be provided to demonstrate the likely outcomes of a policy or absence of a policy (i.e., educational achievement for incarcerated youths, imprisonment for drug offenses, drug courts, and hot spots policing).

Research decisions will continue to be driven by a myriad of factors (e.g., researcher’s interests and expertise, theoretical considerations, funding availability, relevance, time constraints) and, most importantly, the result of a scientific examination of the topic area’s current knowledge and associated need for future empirical inquiry to improve the area’s best available knowledge. The knowledge building process, including the ongoing search for causality, while continual, does not and should not prohibit criminologists from sharing best available research findings for current public policy considerations.

Contents of the Special Issue

The featured manuscript by Robert J. Sampson, Christopher Winship, and Carly Knight provides a set of recommendations on how to advance, or “translate,” the best available research knowledge into effective public policy in the absence of causal certainty while simultaneously employing rigorous standards of scientific methodology. The authors advocate for the use of mechanisms and causal pathways, effect heterogeneity, and contextualization in criminological research. In the effort to integrate empirical and theoretical findings to policy, Sampson, Winship, and Knight (2013) call for the use of “policy graphs” that can be used to analyze the associated policy implications of causal relationships. These recommendations will be applied to three criminal justice policy areas, namely, crime prevention (by Abigail Fagan), policing (by Daniel S. Nagin and David Weisburd), and incarceration in supermax prisons (by Daniel P. Mears).

The three substantive area essays are aimed at accomplishing four tasks. First, the essays will identify and review the best available knowledge on the subject area. Second, the essays will summarize the existing policies and practices related to the area. Third, the essays will discuss the extent to which the best available research knowledge corresponds with existing policy and practice in the area. And fourth, the essays will conclude with actionable policy and practice recommendations drawing upon the recommendations from the featured article. Alfred Blumstein will conclude the special issue with his assessment of the efficacy of translating the best available research knowledge into public policy by employing the recommendations provided by Sampson, Winship, and Knight.

Conclusion

The assumption underlying this special issue is that criminal justice policy can benefit significantly from the utilization of best available research knowledge in the absence of

causal certainty and it appears likely that criminologists will increasingly be involved in efforts to link their research to public policy. The process will, no doubt, be incremental and involve overcoming various research knowledge limits and policy process impediments. In this regard, translational criminology, as advocated by the National Institute of Justice (NIJ), is directly aimed at addressing these barriers by creating an open exchange of information among researchers, policy makers, and practitioners to create knowledge and help guide actionable policy steps. However, the blurring of professional boundaries and the confrontation between political ideology and best available research knowledge will not be easily accomplished. The alternative is far more problematic: crime and justice policies and practices that are determined largely by politics, ideology, illusion, or mere convenience. We will need professional collaborations and “champions” as noted by Richard Thornburg (2011) to help navigate the political process. To gain these “champions” it is essential for researchers, practitioners, and policy makers to forge ongoing partnerships and professional collaborations aimed at identifying specific research, policy, and practice priorities and associated action strategies.

While criminology’s research and causal reasoning is contingent and incomplete, it is the best basis for public policy. Certainly, it is imperative not to oversell policies that are based upon still evolving research and causal claims and to include specific requirements for implementation and evaluation of recommended policies and practices. We must be transparent about what we know and what we do not know and our policy recommendations must reflect this transparency. As stated before, given our best but incomplete research knowledge, our policy efforts will not necessarily involve identifying a particular policy or practice but, rather, identifying likely consequences involved in particular policy choices.

References

- Allen, Francis. 1964. *The Borderland of Criminal Justice: Essays in Law and Criminology*. Chicago: University of Chicago Press.
- Association of Doctoral Programs in Criminology & Criminal Justice. 2013. *Association of Doctoral Programs in Criminology & Criminal Justice (ADPCCJ): 2013 Survey Report*. Retrieved August 22, 2013, from adpccj.com/documents/2012survey.pdf.
- Blomberg, Thomas G., William D. Bales, Karen Mann, Alex R. Piquero, and Richard A. Berk. 2011. Incarceration, education and transition from delinquency. *Journal of Criminal Justice*, 39: 355–365.
- Blomberg, Thomas G., William D. Bales, Alex R. Piquero. 2012. Is educational achievement a turning point for incarcerated delinquents across race and sex? *Journal of Youth and Adolescence*, 41: 202–216.
- Blomberg, Thomas G. and Karol Lucken. 2010. *American Penology: A History of Control*. New Brunswick, NJ: Aldine Transaction.
- Blomberg, Thomas G. and Michael D. Reisig. 2008. Editor’s Comments. *Criminology & Public Policy*, 7: 1–4.

- Burawoy, Michael. 2005. For public sociology: 2004 Presidential Address. *American Sociological Review*, 70: 4–28.
- Burgess, E. W. and D. J. Bouge. 1967. *Urban Sociology*. Chicago: University of Chicago Press.
- Cloward, Richard A. and Lloyd Ohlin. 1960. *Delinquency and Opportunity: A Theory of Delinquent Gangs*. Glencoe, IL: Free Press.
- Cohen, Albert K. 1955. *Delinquent Boys: The Culture of the Gang*. Glencoe, IL: Free Press.
- Cullen, Francis T. 2005. The twelve people who saved rehabilitation: How the science of criminology made a difference. *Criminology*, 43: 1–42.
- Elster, J. 1991. *Arguing and Bargaining in Two Constituent Assemblies*. The Storrs Lectures, Yale Law School.
- Fagan, Abigail. 2013. Family-focused interventions to prevent juvenile delinquency: A case where science and policy can find common ground. *Criminology & Public Policy*, 12: 617–650.
- Feeley, Malcolm M. and Austin D. Sarat. 1980. *The Policy Dilemma: Federal Crime Policy and the Law Enforcement Assistance Administration 1968–1978*. Minneapolis: University of Minnesota Press.
- Kleck, Gary and J. C. Barnes. 2011. Article productivity among the faculty of criminology and criminal justice doctoral programs, 2005–2009. *Journal of Criminal Justice Education*, 22: 43–66.
- Latessa, Edward, Francis T. Cullen, and Paul Gendreau. 2002. Beyond Correctional Quackery—Professionalism and the Possibility of Effective Treatment. *Federal Probation*, 66: 43–49.
- Lemert, Edwin. 1970. *Social Action and Legal Change: Revolution within the Juvenile Court*. Chicago: Aldine Publishing Company.
- Mears, Daniel P. 2013. Supermax prisons: The policy and the evidence. *Criminology & Public Policy*, 12: 681–719.
- Merton, Robert K. 1938. Social structure and anomie. *American Sociological Review*, 3: 672–682.
- Messinger, Sheldon L. 1969. *Strategies of Control*. Unpublished doctoral dissertation, University of California, Los Angeles, California.
- Mustaine, Elizabeth E. and Richard Tewksbury. 2009. Rainmakers: The most successful criminal justice scholars and departments in research grant acquisition. *Journal of Criminal Justice Education*, 20: 40–55.
- Nagin, Daniel S. and David Weisburd. 2013. Evidence and public policy: The example of evaluation research in policing. *Criminology & Public Policy*, 12: 651–679.
- Nutley, Sandra M., Isabel Walter, and H. T. O. Davies. 2007. *Using Evidence: How Research can Inform Public Services*. Bristol, U.K: The Policy Press.
- Petersilia, Joan. 1991. Policy relevance and the future of criminology. *Criminology*, 29: 1–16.

- Rein, Martin and Christopher Winship. 2000. The dangers of “strong” causal reasoning: root causes, social science, and poverty policy. In (Jonathan Bradshaw and Roy Sainsbury, eds.), *Experiencing Poverty*. Aldershot: Ashgate.
- Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. Translating causal claims: principles and strategies for policy-relevant criminology. *Criminology & Public Policy*, 12: 587–616.
- Skolnick, Jerome H. 1965. The sociology of law in America: Overview and trends. Law and Society Supplement to *Social Problems* 13: 4–39.
- Skolnick, Jerome. H. 1966. *Justice Without Trial: Law Enforcement in Democratic Society*. New York: Wiley.
- Sudnow, David. 1965. Normal crimes: Sociological features of the penal code in a public defender office. *Social Problems*, 12: 255–276.
- Sutherland, Edwin H. 1947. *Principles of Criminology*, 4th Edition. Philadelphia: Lippincott.
- Thornburgh, Dick. 2011. Laudable goals: Practical hurdles. *Criminology & Public Policy*, 10: 179–182.
- Tittle, Charles. 2004. The arrogance of public sociology. *Social Forces*, 82: 1639-1643.
- U.S. 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.
- Weiss, C. H. 1949. The many meanings of research utilization. *Public Administration Review*, 36: 426–431.
- Wellford, Charles. 2009. Criminologists should stop whining about their impact on policy and practice. In (Natasha A. Frost, Joshua D. Frelich, and Todd R. Clear, eds.), *Contemporary Issues in Criminal Justice Policy: Policy Proposals from the American Society of Criminology Conference*. Belmont, CA: Wadsworth.
-

Thomas G. Blomberg is Dean and Sheldon L. Messinger Professor of Criminology at Florida State University’s College of Criminology and Criminal. He has been editor of *Criminology & Public Policy* since 2007. Among his research interests are linking criminology and public policy, criminal justice and social control, and education and crime.

Julie Mestre is a doctoral student in the College of Criminology and Criminal Justice at Florida State University. She is the managing editor of *Criminology & Public Policy* with research interests in criminal justice policy evaluation, corrections, and public opinion.

Karen Mann is affiliated with Florida State University’s College of Criminology and Criminal Justice Center for Criminology and Public Policy Research. Her research interests include criminology and public policy, bullying, and education and crime.

EXECUTIVE SUMMARY

CRIMINOLOGY, CAUSALITY, AND PUBLIC POLICY

Overview of: “Translating Causal Claims: Principles and Strategies for Policy-Relevant Criminology”

Robert J. Sampson

Christopher Winship

Carly Knight

Harvard University

Research Summary

This article reviews the causal turn in the social sciences and accompanying efforts by criminologists to make policy claims more credible. Although there has been much progress in techniques for the estimation of causal effects, we find that the link between evidence and valid policy implications remains elusive. Drawing on criminological theory and research insights from disciplines such as sociology, economics, and statistics, we assess principles and strategies for informing policy in a causally uncertain world. We identify three distinct domains of inquiry that form a part of the translational process from evidence to policy and that complicate the straightforward exportation of causal effects to policy recommendations: (a) mechanisms and causal pathways, (b) effect heterogeneity, and (c) contextualization. We elaborate these three concepts by examining research on broken windows theory, policing, video games and violence, the Moving to Opportunity voucher experiment, incarceration, and especially the rich set of experimental studies on domestic violence that originated in Minneapolis, MN in the early 1980s. We also articulate a set of conceptual tools for advancing the goal of policy translation and offer recommendations for how what we call “policy graphs”—causal graphs used to analyze the policy implications of a system of causal relations—can potentially integrate the theoretical and policy arms of criminology.

Policy Implications

Evidence, even if causal, does not necessarily inform policy. In fact, the question of “what works,” the focus of the growing evidence-based movement in criminology, turns

out to be a different question than, “what will work?” Evidence-based policy research must therefore be concerned with much more than providing policy makers with research on causal effects, however precisely measured. The implication is that we must separate criminology’s increasing focus on causality from its policy turn and formally recognize that the latter requires a different standard of theory and evidence than does the former. In particular, criminologists interested in making policy claims must ask hard questions about the potential mechanisms through which a treatment influences an outcome, heterogeneous effects across people and time, contextual variations, and all of the real-world phenomena to which these challenges give rise—such as unintended consequences, policies that change incentive and opportunity structures, and the scale at which policies change in meaning. Theoretically guided causal graphs enhance this goal and help inform policy in a causally uncertain world. Translational criminology is ultimately a process that entails the constant interplay of theory, research, and practice.

Keywords

causality, mechanisms, effect heterogeneity, policy graphs

Translating Causal Claims

Principles and Strategies for Policy-Relevant Criminology

Robert J. Sampson

Christopher Winship

Carly Knight

Harvard University

Criminology has deliberately and increasingly turned its attention to influencing public policy. The founding of *Criminology & Public Policy* as an official publication of the American Society of Criminology is a major recognition of this trend. At the same time, criminology and the social sciences at large have undergone what some have termed a “causal revolution,” a movement characterized by increased attention to, and higher standards for, causal claims. These two trends are controversial and seemingly in conflict. As Blomberg, Mestre, and Mann (2013) noted in their introduction to this special issue, many criminologists eschew making policy recommendations or “public” claims (Tittle, 2004). In a field characterized by uncertain knowledge—especially regarding the root causes of crime—and contested claims for which very few stylized facts are agreed upon, strong policy advice may be premature (Manski, 2013; Rein and Winship, 1999).

We agree with those criminologists who see a social world characterized by contingency, and yet we support criminology’s policy turn. This article seeks to resolve this tension by assessing principles and strategies for informing policy in a causally uncertain world. Our claim for a new mode of “translational criminology” (Laub, 2012) does not double down on the standards for establishing causality, as does the current trend. To accommodate policy-relevant research, for example, many criminologists have shifted their object of study to proximate and malleable factors that are amenable to intervention by the state, and they have made great strides in developing experimental approaches and technical tools for the identification of causal effects. In a causally complex world,

Direct correspondence to Robert J. Sampson, William James Hall, Harvard University, 33 Kirkland St., Cambridge, MA 02138 (e-mail: rsampson@wjh.harvard.edu).

however, policy research requires more than the estimation of causal effects, even if precisely and well identified. Rather, it requires system-level knowledge of how policy is expected to work within a larger social context. Methodological fine-tuning and even technical certainty, we argue, cannot substitute for theory, substantive knowledge, and attention to context.

To translate criminological findings into policy recommendations instead requires a set of strategies that move us beyond the narrow confines of causal identification. In this article, we identify three distinct topics or domains of inquiry that must be part of the translational process and that complicate the straightforward exportation of causal effects to policy: (a) *mechanisms and causal pathways*, (b) *effect heterogeneity*, and (c) *contextualization*. We elaborate each of these in turn, accompanied by conceptual tools for advancing the goal of policy translation. We specifically offer recommendations for how what we call “policy graphs”—causal graphs used to analyze the policy implications of a system of causal relations—can potentially integrate the theoretical and policy arms of criminology.

The Shift in Causal Standards

Over the past several decades, the social sciences have been swept by an interest in causality and in developing a technical apparatus with which to identify it. The very language of causality—from selection bias to endogeneity concerns—has spilled over the borders of experimental research to become part of a common social scientific lexicon. The sources of the causal revolution are complex, but the disciplines of economics and statistics are major drivers that have integrated an epistemological agenda with powerful methodological tools. In economics, Heckman (2005) has drawn on a venerable tradition of structural equation modeling and on the analysis of alternative courses of economic action to put forth what he terms a “scientific model of causality” (to which we return in the subsequent discussion). In statistics, Donald Rubin and others (Holland, 1986; Rubin, 1974, 1978), building on the foundational work of Neyman (1935, 1990 [1923]), have articulated the counterfactual model of causal inference (also referred to as the “potential outcome” model) that supplies an exact definition of a causal effect with implications for how it should be identified. Here, the fundamental problem of causal inference is that for any given unit of treatment we cannot actually observe that unit’s counterfactual outcome—that is, what would have happened if it did or did not receive the treatment. Experiments are able to overcome this problem through the power of randomization. By randomizing a treatment and averaging over observations, experiments provide an unbiased estimate of the *average causal effect*.

It is not difficult to understand, then, how experiments have won the title of the “gold standard” of empirical research. Compared with experiments, methods based on observational data alone are no longer considered by many social scientists to warrant the same level of confidence when estimating causal parameters. The result is that evidence-based policy has largely become equated with evidence from randomized controlled trials

(RCTs).¹ Because experiments often are not possible, however, a litany of quasi-experimental methods (e.g., the use of instrumental variables) and sophisticated modeling (e.g., matching and propensity scores) have been developed that attempt to mimic the classic experimental design. The result of these changes is that empirical standards of evidence today are as high as they have ever been, and concerns with identification of causal estimates now animate much of social science methodology.

Criminology and Policy Today

It is within this general social scientific context that criminology has undergone its own causal turn. Criminology has historically been focused on “backward-looking causality” or what often is termed “the causes of an effect.” In the case of the causes of crime, classic criminological subjects such as poverty or subcultural values are typically considered root causes. Yet the turn toward causality and policy has pushed much of criminology away from this kind of focus. As famously maligned by James Q. Wilson in *Thinking About Crime* (1975), root causes are not only steeped in causal uncertainty but also, from the perspective of policy, may be irrelevant. Rather than investigate causes of crime that governments (at least in criminal justice departments) are generally powerless to change, Wilson argued that criminology should seek solutions elsewhere and in essence turn its back on theory.

The rise of the counterfactual paradigm, coupled with policy demands, has pushed research toward trying to identify possible interventions the government would be better equipped to undertake. Policy-based criminology has thus largely adopted a “forward-looking” approach—“the effects of a cause.”² Root causes have been replaced with a focus on treatments, such as policing tactics. The question has shifted from “what causes crime” to “did a program work?” In addition, causal standards have been raised throughout the field, such that it is now common in the criminological literature to see the use of propensity scoring and instrumental variable approaches in warranting causal claims, even for classic questions on root causes.

Many aspects of this turn are clearly salutary. Identifying treatments that “work” is no easy matter, and causal clarity can hardly be considered a bad thing. But increasingly, it seems, the causal turn and the policy turn have led to the posing of increasingly narrow questions and to the conflation of “what policy works” to the issue of “did the treatment have a causal effect?” This emphasis is perhaps best indicated by the Department of Justice’s new “clearinghouse” website for the assessment of existing research—“crime solutions” (see crimesolutions.gov/about.aspx). The idea is to offer policy makers guidance on what to do based on prior research that has been deemed by review panels to meet rigorous standards of causal evidence. As stated on the website, “crimesolutions.gov uses rigorous research to

-
1. See the discussion in Cartwright and Hardie (2012) and Ludwig, Kling, and Mullainathan (2011); see also coalition4evidence.org/.
 2. Statistics has largely abandoned the causes of effects, as Holland (1986) proposed.

inform practitioners and policy makers about *what works* in criminal justice, juvenile justice, and crime victim services” (accessed May 23, 2013, emphasis in original). Individual studies are ranked, with randomized experiments getting the highest rating in terms of ensuring the internal validity of results.

Yet internal validity and what works in terms of policy are two separate, and only loosely connected, questions. No causal estimate, however precise, is the same as a policy prescription for “what *will* work.” In the next section, we will expand on what we believe are additional and necessary questions that must be taken into account when translating empirical results into policy recommendations. Establishing internal validity is only a first step.

Lost in Translation? From Causal Claims to Policy Intervention

Despite the obvious importance of experimental and quasi-experimental evidence in causal analysis, some scholars, most notably the economist James Heckman, have pointed to their underappreciated limitations for policy analysis (Heckman, 2005, 2008; Heckman and Smith, 1995). Heckman (2008: 4–5) argued that three types of questions are involved in policy analysis:

- (P1) Evaluating the effects of historical interventions on outcomes, including their impact on the treated and society at large.
- (P2) Forecasting the effects (constructing counterfactual states) of interventions implemented in one environment in other environments.
- (P3) Forecasting the effects of interventions (constructing counterfactual states associated with interventions) never historically experienced to various environments.

Ideally, experiments can inform the answers to the first question (P1). According to Heckman (2008), theory and ultimately structural equation modeling are needed to provide answers to the second (P2) and third (P3) questions. Heckman’s argument looms large when we consider that in most policy contexts, questions P2 and P3 are what really matter—that is, what will happen in contexts different from where an experiment has occurred and where the intervention differs in important ways from that carried out in the experiment.

Similar to Heckman (2008), we argue that there is a large gulf between the kinds of information causal analysis typically provides and the kind of information that well-informed policy demands. To traverse this gap, we advance three broad topics that must form a part of the translational process from experimental results to policy recommendations: (a) the identification of mechanisms and pathways, (b) effect heterogeneity, and (c) contextualization. In each case, theory is essential to assessing the importance and possible policy implications of experimental or other causally based evidence.

Consider first the gold standard of an RCT. The experiment establishes the existence of a link between a treatment and an outcome for a particular population in a particular context. Within this circumscribed context, the goal of the experiment is to generate an internally valid estimate. We depict this bare-bones causal model in Figure 1. Establishing

FIGURE 1

Treatment Affects Outcome Conditional on Experimental Context

$$C : \textit{Treatment} \longrightarrow \textit{Outcome}$$

this link is no small feat. It is of great importance to policy makers as well as to academics to be able to identify with precision that some program has reduced recidivism or juvenile offending, or any number of seemingly intractable criminological outcomes.

Yet once this link is established, the translational process has only just begun. First, policy requires information about what will happen in *different* contexts, the very thing for which experiments make no claim to be able to estimate. Although the problem of external validity is well known in the social sciences, it has not been fully confronted by criminologists. It is remarkable, for example, that whereas the website of crimesolutions.gov contains clear descriptions of causal evidence and internal validity, there is currently no entry whatsoever for external validity.³ Causal evidence is instead defined in terms of “evidence that documents a relationship between an activity, treatment, or intervention (including technology) and its intended outcomes, including measuring the direction and size of a change, and the extent to which a change may be attributed to the activity or intervention.” This definition may nicely fit the requirements for causal evidence for internal validation. However, causal evidence for *policy recommendations* outside this context should be held to a more demanding evidentiary—and theoretical—standard.

Second, and more fundamentally, in most cases a policy is not a treatment. Thus, to recommend policy requires more than considering how a treatment would be expected to work across diverse locales. When one considers policy not as a randomized trial but as a change in institutional structure, it becomes clear that theory must be brought to bear for prediction. A policy is, by definition, a change in the rules of the game. As a result, “policy translation” involves both the problem of what happens when “C” in Figure 1 changes *and* the problem of accounting for changes in organizational, political, or wider social structure when the treatment in Figure 1 scales up into official policy.

The point merits repeating: Even the most internally valid RCT, one that provides near incontrovertible evidence as to the existence of a link between treatment and effect, can only be uncertainly applied to a formal policy context. This limitation arises with equal force to nonexperimental designs.⁴ The fundamental disjuncture of evidence and policy application raises the problem of causal *interpretation*. No matter how experiments may be

3. Accessed May 23, 2013. The classic treatment of internal and external validity is from Cook and Campbell (1979).

4. We focus here on experiments because they are held up as the gold standard of research design and because other methods typically face the limitations of experiments and many additional problems as well.

understood by researchers, experimental results often are interpreted by policy makers as a direct test of policy, as evidential support that “clinches” a conclusion of “what works” (Cartwright, 2007). But the simplicity of the causal graph in Figure 1 is deceptive, such that causal analyses tend to carry with them an implicit exportability claim (Barinboim and Pearl, 2013). Conditionality and constraints fall away, and experiments all too often are misread as a general statement of how the world works. To move beyond these limitations we must consider what is inside the experimental black box—to look at why and how T is linked to Y .

Strategies for Moving Forward

The call for a holistic and contextual approach to understanding causal relationships may sound intractable. How are we to reconcile complex causality—with multiple pathways, heterogeneous effects, and interdependent systems—with policy recommendations that are useful and manageable? What are the strategies for improving criminological research? Recognition of complexity need not imply nihilism in practice. The purpose of this article is thus not to discourage policy-relevant criminological research or critique experiments but to suggest topics that a translational criminology must address, topics that should feature in any discussion of “what works” or, perhaps more accurately, “what will work.” In doing so, we note that experiments are part of the solution; especially those conducted within criminal justice agencies and that test mechanisms (Ludwig et al., 2011). Another part of the solution involves modes of inquiry that, under the name of causal rigor, often have been branded as inferior: descriptive data and “speculative” theory. We can move forward by remaining cognizant of and theorizing how the relationships identified by experiments fit within a larger social structure.

Although a growing number of tools are used to model complexity, we use a strategy that can aid in giving proper attention to the sorts of complexity that we identify: *causal graphs*. Causal graphs have, for decades, been used by social scientists to understand systems of causal relationships. They are at the core of path analysis (Duncan, 1966, 1975) and are an important component of structural equation modeling (Bollen, 1989). In the past decade they have gained a renewed importance with Judea Pearl’s (2009 [2000]) influential work on directed acyclic graphs (DAGs).⁵ For Pearl, the fundamental purpose of DAGs is that they provide a simple but powerful tool that allows for the analysis of the conditions under which an observed association can be identified as a causal effect. Although translation, not identification, is our key concern, we draw from Pearl and others the broader point that causal graphs are a useful way of specifying the theoretical structure of a problem and illustrating the interdependencies in causal systems. From this perspective, making policy

5. See Bollen and Pearl (2013) for a thorough discussion of the complementarity and interrelationship between structural equation models and DAGs as modes of representing causal structures.

requires knowledge of an interlocking system of organizations and actors, demanding a bird's-eye view of the wider causal picture.

Because we are not concerned with identification but with how causal relationships fit within wider systems, we do not need the extensive mathematical machinery associated with DAGs.⁶ Moreover, we will label causal effects as positive or negative in our various examples, which is something not generally done with DAGs.⁷ As our focus is on policy, we will term our causal graphs “policy graphs” and broaden the initial example in Figure 1.

In using causal graphs to understand policy implications, we also revisit the contrast between forward-looking causation and backward-looking causation. Although useful, this distinction may obscure the fact that *both* kinds of causation must refer to the causal system in which effects are located. That is, to apply forward-looking causation to policy, the constellation of mechanisms surrounding it must be elaborated. To understand how a specific treatment affects an outcome, and thus what the effects of a policy intervention are likely to be, one needs to understand the causal system of which a specific estimated effect is a part. Responsible policy analysis cannot be done theory free (Laub, 2004: 14–19).

In short, the use of causal graphs in policy analysis provides an explicit and concrete way to bring theory into one's analysis or even better, one's research design. Specifically, causal graphs are a way of representing a theoretically derived causal system related to an outcome of interest. They are important both for research design done prior to the collection of data and for the interpretation of empirical findings derived from statistical analyses. As such, we should keep in mind the interaction of various causal processes in the production of an outcome, rather than focusing on the effect of a singular treatment abstracted from its setting. We now examine in depth how this strategy might work with respect to the three pragmatic challenges a translational criminology must address: mechanisms and pathways, effect heterogeneity, and contextualization.

Mechanisms and Pathways

Experiments provide a causal graph of the world in which a treatment leads to an outcome: $T \rightarrow Y$. Mechanisms disaggregate this relationship and in doing so provide what Heckman (2005, 2008) would term a scientific understanding of causality. VanderWeele (2009), drawing on work by Aalen and Frigess (2007), helps us understand this argument by making the distinction between *counterfactual-based causality* and *mechanistic causality*. As VanderWeele (2009: 222) describes the distinction:

Counterfactual-based causality is essentially concerned with the effects of a particular intervention or exposure without regard to the mechanisms by which

6. See Morgan and Winship (2007) for a basic introduction to DAGs and Elwert (2013) for a more thorough presentation of their workings.

7. Because DAGs are nonparametric, the labeling of effects is not appropriate.

these effects arise. Conclusions about causal effects are drawn either through randomized trials or through the careful design and analysis of observational data in which the researcher attempts to control for all the variables that confound the exposure-outcome relationship.

This approach is described as “black box” causality because the methods used to estimate causal effects can be valid *irrespective of how the exposure produces its effect*.

Mechanistic causality is different. We quote VanderWeele again at length (2009: 222):

Mechanistic causality, on the other hand, attempts to understand the mechanisms governing the various processes which give rise to particular outcomes. Assessing mechanistic causality requires closer observations and a good deal of scientific knowledge. The model for mechanistic causality is the natural sciences in which attempts are made to identify the natural laws and precise workings behind the phenomena we observe. In the mechanistic approach, one attempts to “look inside the black box.” A similar distinction to that made by Aalen and Frigess is also made by Heckman (2005, 2008). Heckman calls the counterfactual-based causality described above “statistical causality.” He contrasts this with what he calls the “scientific model of causality” or “economic causality.” Heckman criticizes the statistical literature on causality for not making use of theory and for not taking into account agent choice, equilibrium processes and feedback which he argues are the mechanisms by which outcomes are generated.

There are three reasons why the theoretical specification and analysis of mechanisms are a fundamental part of policy analysis (see also Ludwig et al., 2011; Rosenzweig and Wolpin, 2000). First, mechanisms are necessary for interpretation of what is a cause and what is merely a risk factor in crime (Wikström, 2011). Second, policy generally is concerned with achieving a particular causal *process*, not simply a causal *effect*. Finally, policy efficacy requires considering alternative, cost-effective processes for bringing about the desired outcome; mechanisms can identify these processes.⁸

To elaborate on these points, we begin by considering the “broken windows” theory of crime. Introduced by James Q. Wilson and George L. Kelling in 1982 in the *Atlantic Monthly*, the theory contends that targeting minor forms of disorder—broken windows, loitering, and graffiti—reduces serious crime. Disorder is hypothesized to provide visual cues of the state of neighborhood social control from which potential offenders derive

8. Increasingly, funders of interventions have come to a similar conclusion and are pushing back against the emphasis on the black box of average causal effects. The President of the William T. Grant Foundation recently wrote: “In today’s vernacular, we need more research attention paid to *why* and *under what conditions* things work as the missing ingredients in the ‘what works’ agenda” (Granger, 2011: 29).

FIGURE 2
Classic Broken Windows

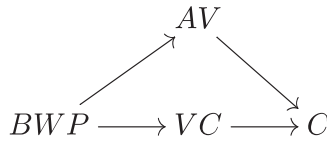
$$BW \longrightarrow VC \longrightarrow C$$

expectations of whether further antisocial behavior will be tolerated or reported. Thus, the theory goes, the appearance of disorder begets further disorder: Broken windows breed an environment conducive to crime. Represented in Figure 2, broken windows (*BW*) provides visual cues (*VC*) of social disorganization that in turn incentivizes crime (*C*).

Broken windows theory has proven to be one of the most influential theories in criminology, prompting a wave of “zero-tolerance” and “order-maintenance” policing in New York City, Chicago, and Los Angeles (Duneier, 1999; Harcourt, 2001). For example, New York City Mayor Rudy Giuliani’s crackdown on misdemeanor crimes was widely touted as a principal cause of the drop in crime in the 1990s. Even today, the NYC Police Department continues to cite aggressive policing as a decisive factor in the crime drop. However, although there is some empirical support that broken windows policing lowers crime (Kelling and Sousa, 2001), findings have been mixed overall (Harcourt, 1998, 2001; Harcourt and Ludwig, 2006; Sampson and Cohen, 1988). Even the evidentiary basis in Zimbardo’s 1969 vandalism experiment, often cited as the foundation of the broken windows thesis, is questionable. In that study, cars were purposively left abandoned in the Bronx, NY, and Palo Alto, CA. In the Bronx, the car was immediately vandalized, whereas in Palo Alto, the car was left untouched for a week. Only after Zimbardo himself smashed in the windows did the Palo Alto car succumb to further destruction. The causal link from visual cues to crime was thus contextually dependent even at the outset (Sampson, 2013: 17).

Mechanisms feature prominently in critiques of broken windows. This brings us to our first point: Mechanisms are a necessary part of the *interpretation* of causal claims (Rosenzweig and Wolpin, 2000). To the extent that we can identify an empirical relationship, mechanisms tell us why that relationship exists. Consequently, they also warn us when a relationship may be spurious, or caused by some other process than the one generally touted to be at play (Knight and Winship, 2013; Wikström, Oberwittler, Treiber, and Hardie, 2012). For example, Sampson and Raudenbush (1999, 2004) called into question the mechanism linking visual cues to crime according to the original broken windows theory, considering both spurious pathways (common causes) and interpretive processes. In analyzing the social-psychological processes behind the formation of perceptions of disorder, they find only a modest association between *perceived* disorder and *actual* disorder. Rather, stereotypes about the neighborhood, based in large part on its racial composition, have a much larger effect on perceptions. In other words, there is no guarantee that repaired windows will be noticed. Neighborhoods with large minority populations also generate enhanced perceptions

FIGURE 3

Broken Windows, Alternative Mechanism

of disorder among all race/ethnic groups, suggesting a general *cultural* process, not an unmediated effect of visual cues.

By testing the broken window's hypothesized mechanism, Sampson and Raudenbush (2004) complicated the story. If broken windows policing works and if their findings are correct, then this style of policing may work through a different process than changes in perception of actual disorder. Or, to the extent that broken windows theory is correct, then policies targeting disorder may be least likely to work in some of the most disadvantaged areas with the most entrenched neighborhood stereotypes. To go back to Zimbardo's (1969) experiment, broken windows might be an important signal in some communities while they may be meaningless in others. The wider point that emerges from this observation is that the identification of mechanisms is inextricably tied to interpretation of the causal claim. Answers to when, where, and for whom broken windows policing works are incoherent without recourse to mechanisms.

Our second point builds on the first: Mechanisms are necessary for the *justification* of public policy. In many cases, it is of great normative importance whether the treatment exerts its causal force through one or another mechanism. Policy, in other words, is not always concerned simply with achieving an outcome by manipulating a cause: Rather, it aims to achieve an outcome via *a certain route*. Consider, for example, that there are a number of pathways through which broken windows policing, with its targeting of misdemeanor crimes, could affect the crime rate. As depicted in the enhanced causal graph presented in Figure 3, the standard mechanism proposed by broken windows theory is given by the pathway in which broken windows policing (*BWP*) affects visual cues (*VC*) that in turn affects crime (*C*): ($BWP \rightarrow VC \rightarrow C$). However, as Sampson and Cohen (1988) suggested, this is hardly the only pathway by which policy could act: Broken windows policing may directly increase police presence, raising the likelihood of arrest for serious crimes. This possibility is represented by a pathway from broken windows policing (*BWP*) to arrests for violent crime (*AV*) to crime (*C*): ($BWP \rightarrow AV \rightarrow C$). In this case, what seems to be deterrence would actually be the result of a higher arrest rate for serious crime and subsequent incapacitation, rather than the result of a higher arrest rate of misdemeanors or a change in offender perceptions. Harcourt (1998) discussed a related possibility, in which

aggressive policing may lead to the arrest of individuals the police would not have grounds to arrest otherwise, preempting potential offenders through increased surveillance.

Which of these pathways is responsible for the observed effect matters for policy because the political attractiveness of a certain intervention may depend on the mechanism being used to change behavior. Zero-tolerance policing may be politically efficacious if it works by sending a signal to offenders that the neighborhood will not tolerate crime. It is less so, however, if it works through the police indiscriminately abusing the power to arrest. This latter mechanism is doubly pernicious given that police legitimacy would likely be undermined by aggressive and discriminatory behavior (Fagan and Davies, 2000), perhaps leading to increased crime in the long run.

Mechanisms are also an important part of identifying *efficient* policy. Even granting that a study has identified evidence of a cause, without an understanding of the often competing underlying mechanisms, we cannot determine whether that treatment is a more justified policy response than another, potentially effective and cheaper intervention. For example, if zero-tolerance policies promote higher rates of arrest for violent crime and it is these arrests that are doing the causal work, then arresting individuals for relatively minor misdemeanors would be a waste of resources. In other words, causal certainty about the effect of a treatment does not in itself imply certainty that a policy change will yield effects in ways that follow the logic of the intervention.

Despite the importance of mechanisms, they rarely feature in policy recommendations of “what works.” For instance, an RCT conducted by Braga and Bond (2008) to test broken windows theory with respect to hot-spot policing was given the highest evidentiary rating by crimesolutions.gov. Yet what qualified as broken windows policing consisted of multiple interventions, including increased misdemeanor arrests, better lighting, providing youth with recreational opportunities, working with local shelters to provide housing for homeless individuals, and connecting problem tenants to mental health services. The latter components were “intended to create opportunities for high-risk individuals to assist police efforts to promote social order.”⁹ A benefit of the experimental approach was that a particular broken windows policing “package” could be randomized. The drawback, of course, is that it is impossible to know which component of this strategy worked. To our mind, also it is problematic to assume that the purpose of giving social services is to help recipients assist police officers—indeed, this assumption is loaded with unexamined theoretical baggage. Even if an increase in social services does promote cooperation with the police, this is a very different mechanism than that typically associated with increased misdemeanor arrests for social disorder.

9. See also crimesolutions.gov/ProgramDetails.aspx?ID=208. In another evaluation of “broken windows” policy, the treatment included better technology for detecting crime patterns (commonly known as COMPSTAT)—see crimesolutions.gov/ProgramDetails.aspx?ID=87. Improved technology is far removed from the mechanisms posited in the original broken windows theory.

In sum, the debate over broken windows theory serves to illustrate the basic point that contention over policy rarely concerns only causal identification. Rather, it must grapple with how policy works. In saying this, we are making a stronger point than that criminological research also should be concerned with identifying mechanisms. Going further, we have argued that policy claims often make strong theoretical assumptions, or are themselves unarticulated theories about social processes. In advocating for broken windows policing, policy makers are not simply interested in a causal effect but in a causal *process*. Criminological research, then, should be understood as adjudicating among competing models of social processes.

Heterogeneous Effects

Mechanisms complicate our simple causal model by identifying the various pathways through which a treatment works. Effect heterogeneity, on the other hand, implies that multiple causal models exist that correspond to different subpopulations. Effect heterogeneity occurs when a given treatment has a different causal effect for different individuals or subgroups (e.g., males or females). In extreme cases, a treatment may be beneficial for some individuals, detrimental for others, and have no effect for others. Medical researchers are increasingly aware of the ubiquitous nature of effect heterogeneity, to the point that a lengthy *New York Times* essay recently posed the provocative question: “Do Clinical Trials Work?” (Leaf, 2013).

This question matters for criminal justice policy too but in a way that goes beyond the specific method of an RCT. At the most basic level, policy is concerned not only with *how* effects are brought about but also *for whom* they are brought about. Given the centrality of questions concerning quality and fairness to policy, the issue of differential effects must be a primary part of policy analysis. For example, a policy that benefits Whites while harming Blacks is not an example of a policy that most would consider “works” even if its average causal effect was salutary. Moreover, behavioral assumptions implicit in unaccounted heterogeneous effects may cause us to *understate* the effect of a policy. When a random experiment provides a treatment both to those for whom the treatment will give a high return as well as to those for whom the treatment will give a low return, forced treatment leads to a lower average outcome than would be obtained if the treatment were limited to those likely to benefit.

Consider the case of the Moving to Opportunity experiment (Kling, Liebman, and Katz, 2007; Kling, Ludwig, and Katz, 2005) in which families were randomly assigned vouchers to relocate to low-poverty neighborhoods. Subsequent analysis revealed that although the move was beneficial for young females (lowered arrests for property crime and better mental health), the move had deleterious effects on young male delinquency (heightened aggressive behavior and probability of arrest for property crime). This sort of effect heterogeneity implies separate graphs for each gender (Figure 4). Here, a move (M)

FIGURE 4

Heterogeneity in the Moving to Opportunity Experiment

$$\text{Male} : M \xrightarrow{-} BA \xrightarrow{-} A$$

$$\text{Female} : M \xrightarrow{+} BA \xrightarrow{-} A$$

facilitated behavioral adjustment (BA) for females while hindering it for males. This distinction is represented by the positive arrow for females ($M \xrightarrow{+} BA$) and the negative arrow for males ($M \xrightarrow{-} BA$) and is responsible for different probabilities of arrest in each group (A). In assessing policy, this differential effect must be a part of the conversation. A single causal effect would mask real differences and all the normative and political implications they entail.

Heterogeneity is not just a quality of subgroups. It also is a temporal phenomenon. That is, the efficacy of a treatment depends on its location within the cluster of activities, networks, and institutions that define the actor at any given point. A treatment given to a young man will differ from a treatment given to an old man; a first treatment will differ from a recurring one. Moreover, a treatment at a given age may show an instantaneous effect (or null result) that changes or reverses in a later follow-up. As a consequence, the population-level effects of a policy change, one that individuals may confront daily, will likely diverge from the effects of a single randomized intervention. Our point is not to say that experiments are not useful in estimating treatment effects. Rather, and again, the point is to separate the policy question of what “works” or “will work” from the effects identified by researchers. Policies unfold over a time horizon more expansive than those accommodated by a single empirical study.

Perhaps the best example here is noncriminological but which nonetheless has implications for crime over the life course. The famous Perry Pre-school Project was a randomly assigned treatment that initially boosted IQ in children, but the effect quickly faded. It was only years later that researchers discovered significant treatment effects on lifetime outcomes up to 40 years of age. Accounting for this temporal heterogeneity but also probing the causal mechanisms that produced it, Heckman, Pinto, and Savelyev (in press) showed that the Perry Project significantly enhanced adult outcomes including education, employment, earnings, marriage, participation in healthy behaviors, *and* reduced participation in crime and welfare. They also argued that experimentally induced changes in noncognitive personality traits, rather than IQ, explain a sizable portion of the adult treatment effects. This example shows that the treatment was temporally heterogeneous, had unanticipated spillover effects, and operated through a particular causal pathway—what criminologists would call “self-control.” In the next section we provide a more detailed example of how

the effect of an intended criminological treatment—arrest for domestic violence—similarly varies over the life course and through mechanisms not anticipated by the original design.¹⁰

Contextualization

Academics often dutifully note the importance of social context, typically in terms of boundary conditions that limit causal claims. Yet context implies more than an unarticulated background or boundary against which to generalize causes and effects. To contextualize is to consider an entrenched causal web that intervenes and shapes every point of an unfolding causal process, dictating the nature of incentives, opportunities, and institutional relationships that define the policy world. As such, policy researchers must rethink their understanding of the role of social context, moving it from the periphery to the center of analysis.

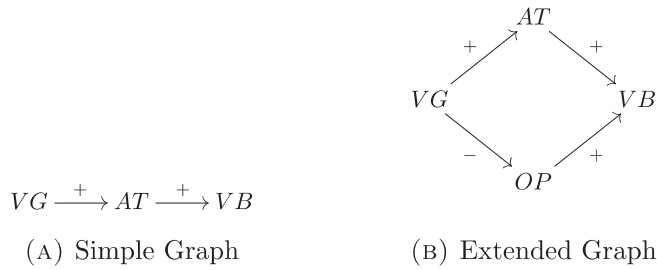
In the subsequent discussion, we elaborate three dimensions of context, from most basic to most demanding, that must be kept in mind when translating empirical results into policy. First, and most straightforwardly, context can be understood as the macro-environment that directly affects the actor, whether neighborhood, city, or country. Consider a study on recidivism by Kubrin and Stewart (2006). The authors pointed out, rightly, that most studies tend to examine the individual-level characteristics associated with recidivism, ignoring how the effects of these characteristics may be dependent on local factors, in this instance, place. Using nonexperimental data, they found that neighborhood context accounted for nearly 13% of the variance in recidivism, with offenders returning to disadvantaged communities reoffending more, net of individual-level factors. In another study showing the power of place-based context to influence recidivism, Kirk (2009) employed a quasi-experimental approach that used the residential destruction resulting from Hurricane Katrina as an exogenous source of variation that influences where a parolee will end up once released from prison. He found that a forced move away from a parolee's former geographic home substantially lessens the likelihood of re-incarceration. Such studies thus have suggested that neighborhood context influences the efficacy of criminal justice policies. More generally, context can be understood as the macro-level environment influencing behavior.

Second, and expanding further, context can be understood as the opportunity structure within which an actor exists. Context shapes incentives and limits choice. As such, the context of the actor in the real world often differs markedly from that of the actor in the experimental one. For example, psychological laboratory experiments have established a positive relationship between video game use and violent behavior. Video games are often said to increase aggressive tendencies through a process of social learning that is posited to support a violent personality and encourage criminal behavior. Yet extrapolating from the

10. In another example, temporal heterogeneity combines with effect heterogeneity. The gender interaction in delinquency uncovered in the interim follow-up of the Moving to Opportunity experiment seems to have eroded in the long-term follow-up (Sanbonmatsu et al., 2011).

FIGURE 5

Video Games and Violent Behavior



lab to the street is a difficult endeavor and ignores the trade-offs between video game playing and other options youths have to occupy their time. Considered against the whole gamut of possible activities, video game playing may not be so socially deleterious. In particular, a recent study by Cunningham, Engelstatter, and Ward (2011) found that playing video games may *incapacitate* violent activity by taking users off the streets, on the whole, decreasing criminal behavior on aggregate.

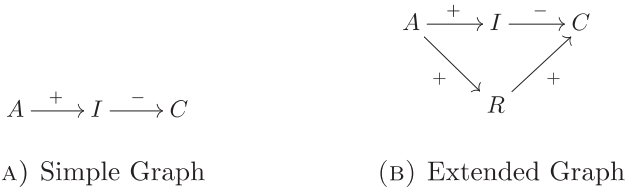
The causal graph given in Figure 5a represents the causal pathway suggested by psychology experiments in which video gaming (*VG*) leads to violent behavior (*VB*) through the creation of aggressive tendencies (*AT*). But when the causal graph is amended to take into account opportunity structure, it is no longer clear whether the positive association identified in the laboratory should hold. As shown in Figure 5b, video game playing (*VG*) may still increase the likelihood of aggressive tendencies (*AT*) and thus increase violent behavior (*VB*), as in the pathway $VG \xrightarrow{+} AT \xrightarrow{+} VB$. However, this criminogenic pathway may be mitigated by a prosocial pathway in which video game playing lowers opportunities for delinquency and, thus, violent behavior ($VG \xrightarrow{-} OP \xrightarrow{+} VB$). Given a positive relationship between opportunities for violence and violent behavior, any drop in opportunity for violence should lessen observed violent behavior.

What we observe in the causal graphs in Figure 5 is that if the effect of video gaming along the prosocial pathway ($VG \xrightarrow{-} OP \xrightarrow{+} VB$) is stronger than that along the antisocial pathway ($VG \xrightarrow{+} AT \xrightarrow{+} VB$), then video game playing will actually lead to a *decrease* in violent behavior. This finding recalls the criminological literature on routine activities theory that explains crime not by reference to singular treatments but by the intersection of activities, opportunities, and environment (Cohen and Felson, 1979).

Not least, context concerns the interdependence of institutions and societal responses—a policy intervention in one part of the criminal justice system will have reverberations, quite possibly changing the intended outcome of the intervention. This result is likely to develop in any social context, but it is a particularly salient feature in criminology, where

FIGURE 6

Countervailing Effects of Incarceration



interlocking institutions and interdependent social networks are the norm. As far back as the 1960s, the President’s Commission on Law Enforcement and the Administration of Justice introduced the influential concept of the criminal justice “system.”

The implication is that feedback loops and unintended causal consequences are theoretically expected. For example, consider America’s grand “natural experiment”—mass incarceration. Although incapacitation may reduce violent crime by removing offenders from the community, Sampson (2011) suggested that removal also decreases the ratio of males to females, which in turn increases family disruption and rates of violence. Based on research that has shown that imprisonment has negative effects on employment, especially the marginalization of Black men from the labor market (Western, 2006), Sampson argued that imprisonment may indirectly lead to future crime through its disruptive effect on Black family structure.

We can illustrate this theoretical model in Figure 6. Figure 6a represents the simple relationship among arrest (*A*), incapacitation (*I*), and crime (*C*): ($A \xrightarrow{+} I \xrightarrow{-} C$). If Sampson’s (2011) contention is right, then estimating only this simple pathway is incomplete without knowledge of the pathway from incarceration to increased family disruption and, hence, increased violence. The alternative pathway of incarceration’s effect through removal (*R*) is represented in the causal graph of Figure 6b ($A \xrightarrow{+} R \xrightarrow{+} C$). This and other potential pathways are especially salient when interpreting a randomized experiment on enhanced policing. An experiment that randomizes by neighborhood or block will capture the causal pathway from arrest to lower crime via incapacitation. But it is less likely to capture the effect of increased incarceration on employment opportunities and family structure, which operates in a temporally and geographically broader context (Western, 2006). Yet the overall effect of increased arrest for a community will depend on the balance between the two pathways, a calculation of which informed policy must assess. Unlike heterogeneous treatment effects and mechanisms, these differences cannot be modeled within an experiment alone, and they require recourse to theory to estimate their effects.

In short, contextualization challenges a framework that assumes that we can manipulate a treatment, *ceteris paribus*—that we can isolate an intervention that is exogenous to the system and assume that incentive structures or practices among individuals and organizations

will not change. A concern for modeling context also points to the importance of *replications*, *meta-analyses*, and *large-scale observational studies* specifically geared toward investigating macro-level factors.

Putting it All Together: The Minneapolis Domestic Violence Experiment and Follow-ups

Up to this point, we have discussed three complexities to the bare-bones RCT causal model of the world: mechanisms, causal heterogeneity, and contextualization. These subjects, we argue, are as necessary to policy considerations or recommendations as are precise causal estimates. Yet they typically do not feature in criminological conversations about policy that works and, when they do, not in a formal or systematic way. We thus turn to a final extended example to illustrate further our argument and provide guidelines. We chose the Minneapolis (MN) Domestic Violence Experiment (hereafter MDVE) and a group of further studies inspired by the original MDVE as a case of experimental criminology that was ultimately “done right.” Through experiments, replications, observational studies, and importantly, criminological theory, the MDVE and its aftermath accumulated information about how mechanisms, heterogeneity, and the importance of context would shape policy on the ground.

Prior to the MDVE, law enforcement had been reluctant to arrest or intervene in cases of domestic violence. This changed during the 1980s after *Thurman v. City of Torrington* (1985) established police liability in the case of domestic assault, thereby incentivizing states to enact policies that eliminated officers’ discretion in cases of domestic abuse. By randomizing police response, the MDVE found that the arrest of an alleged offender caused a marked drop in rates of reoffending (as compared with counseling or separation from the partner). Seizing on the seemingly strong results, legislators, policy experts, and academics began supporting mandatory arrest policies (Mignon and Holmes, 1995).

Although the original researchers repeatedly cautioned against premature extrapolation of the experimental results, the findings drew national attention and quickly ushered in rapid and perhaps unprecedented change in how states and cities administered police responses to domestic violence. Protestations to the limitations of external validity notwithstanding, 24 states adopted mandatory arrest policies (Miller, 2005), an illustration of the power of the experiments’ implicit “exportability” claim. Indeed, the MDVE often has been cited as evidence that mandatory arrest, as a general policy instrument for crimes other than domestic violence, is an effective tool (Davis, 2008).

It is useful, however, to go back to the seminal *American Sociological Review* paper in 1984 that reported the results of the MDVE (Sherman and Berk, 1984).¹¹ The experiment

11. This article is perhaps the most cited criminology paper (with more than 1,100 citations) to appear in the *American Sociological Review* in modern times. A policy brief also was published around the same time as the *ASR* paper.

FIGURE 7

Domestic Violence and Deterrence

$$A \xrightarrow{+} D \xrightarrow{-} RV$$

was presented as a test between two competing theories of the effect of punishment. On the one hand, deterrence theory suggests that punishment will lower recidivism, especially when punishment is certain, swift, and severe. On the other hand, labeling theory predicts that arrest may be criminogenic (Becker, 1963; Lemert, 1951). By randomly assigning arrest, the MDVE sought to adjudicate the effect of arrest. In the end, labeling theory was not supported. At least for the particular population under study, the causal relationship was represented by the simple causal graph in Figure 7: Randomized arrest (A) increases deterrence (D) (i.e., $[A \xrightarrow{+} D]$), which in turn lessens the probability of repeated violence (RV), or $D \xrightarrow{-} RV$.

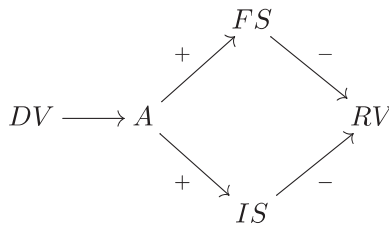
Yet over the course of further research, this simple causal graph began to fray. In the 6 years after MDVE, the National Institute of Justice funded five replication studies, whose results varied widely. On the one hand, studies in Omaha, NB; Milwaukee, WI; and Charlotte, NC, not only found no evidence for the deterrent effect of arrest, but also they reported *increases* in subsequent crimes. Colorado Springs, CO, and Dade County, FL, on the other hand, did find evidence of deterrence.

Our framework suggests that this kind of difference can be expressed by widening the causal graph. As we discussed with regard to mechanisms and effect heterogeneity, an average causal effect may mask important differences in the routes through which, and the groups for which, an outcome is realized. Analyzing the data from the replication studies 8 years after the original MDVE article was published, Sherman, Smith, Schmidt, and Rogan (1992) offered a possible explanation for these diverse findings: differences in how subsamples of individuals responded to arrest as a result of the operation of two possible mechanisms. They argued that although arrest serves as a formal sanction, arrest also is linked to informal sanctions that decrease the likelihood of reoffending for a subset of the population. In particular, individuals who were more socially and institutionally embedded (employed or married) were hypothesized to experience more informal sanctions and greater social controls from their partners and social networks (Sampson and Laub, 1993). Arrest was thus arguably more effective in this group, whereas for men with fewer social bonds, arrest lost its crime-reducing sting.

Put in the language of this article, differences in the domestic violence experiments could be explained by a combination of the heterogeneity of effects and differences in mechanisms. From this perspective, the causal arrow that represents “deterrence” in the simple causal graph from Figure 7 must be further broken down. In Figure 8, we demonstrate how the effect of arrest works through two separate pathways, one involving formal sanctions

FIGURE 8

Domestic Violence: Mechanisms and Effect Heterogeneity



($A \xrightarrow{+} FS \xrightarrow{-} RV$) and the other an effect of informal sanctions on socially embedded men ($A \xrightarrow{+} IS \xrightarrow{-} RV$), both of which lead to decreases in violence (RV). We can then hypothesize that the efficacy of each pathway may well vary with individual characteristics of the offender.

Reexamining the data and consistent with Figure 8, Sherman et al. (1992) found that offenders with a greater number of social ties and thus greater “stakes in conformity” were less likely to reoffend than those missing such ties. To the extent that “the effectiveness of legal sanctions rests on a foundation of informal control” (p. 688), differences in a city’s economic situation and strength of individual ties are predicted to result in different directions of the effect of arrest. That these results are dependent on the degree of “social bonds” suggested a different causal graph from the one posited in the initial experiment. Similar heterogeneous effects of arrest were found in a reanalysis by Berk, Campbell, Klap, and Western (1992). In a subsequent work, Sherman (1993) offered a theory to account explicitly for how arrest “either reduces, increases, or has no effect on future crimes, depending on the type of offenders, offenses, social settings, and levels of analysis” (p. 445). His theory of “defiance” helps explain the conditions under which punishment, or in this case arrest for domestic violence, increases crime.¹²

Within our framework, policy analysis also must be cognizant of temporal heterogeneity. Indeed, a recent 23-year follow-up study by Sherman and Harris (2013) investigated the effect of arrest from the Milwaukee domestic violence experiment. In what is probably the longest follow-up ever of a randomized trial testing the effect of criminal sanctions, they found that arrest had no effect for employed individuals and actually *increased* the prevalence of reoffending for unemployed individuals. Crucially, the harmful effect of arrest was only apparent after 6 years and continued growing for the next 20 years. Much like the Heckman

12. Sherman (in press) also notes the widespread heterogeneity in estimates derived from “hot-spot” policing research—“While the average effect is beneficial, the range of effects is very great. Whether another agency implementing some form of hot spot policing will achieve a large or small effect remains highly uncertain from the available research.”

et al. (in press) article cited earlier on the Perry Preschool Project, the authors conclude that short-term evaluations are inappropriate for understanding longer life-course outcomes. We would generalize further and emphasize the logical conclusion, often overlooked in criminological discussion of policy implications, that any single empirical result must not be conflated with having predicted the long-term outcomes of a policy regime change.

Scaling Up and the Importance of Context

“If I called the police to get him out of my house, I’d get evicted.” Victim of domestic violence in Milwaukee on reluctance to call 911 (quoted in Eckholm, 2013: A1)

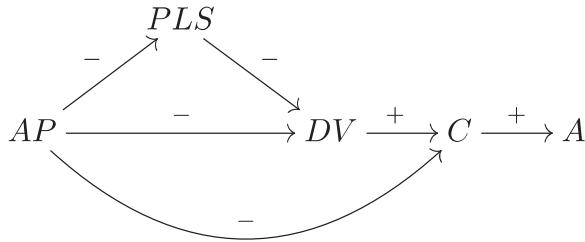
Even with no effect heterogeneity and full knowledge of the mechanisms operating within a particular study, the context challenge implies that a single experiment cannot provide evidence of the consequences of scaling up. Although this challenge might not matter in medical trials, the canonical example of an experimental science, crime, and criminal justice are quintessentially social phenomena. In regard to the MDVE, the treatment was randomly assigned arrest in a study of 314 individuals in a city of nearly 400,000, not an alternative policy universe in which arrest was mandatory. Yet there are many reasons to assume that scaling up will change the nature of the intervention, altering both offenders’ understanding of the likely consequences and, importantly, also the victim’s likelihood of reporting.

Using nonexperimental methods, Iyengar (2007) has found evidence that suggests that the rise in mandatory arrest laws may be associated with a greater probability of spousal homicide, an effect mediated by the lowered probabilities that victims will call the police when it is known that their partner will be arrested. As Iyengar stated, the MDVE’s use of randomized assignment provided no evidence as to whether mandatory arrest would lower recidivism under a policy regime change of larger scale. Once it was known that a victim calling in would unequivocally result in the offender’s arrest, there was a marked drop in reports of domestic violence. Iyengar’s work suggests that this drop was associated with an increase in spousal homicides. Taking into account the effects of scaling up makes clear *policy itself* also must be a part of our causal graph of the world (Figure 9). By doing so, one creates what one might call a “policy graph”—a causal graph in which the policy itself is a variable. This move responds to Heckman’s (2005) call for policy research to take into account agent choice and feedback processes based on expectations.

By making the arrest policy (*AP*) part of the graph in Figure 9, we draw attention to two important causal processes that link the policy of mandatory arrest to actual instances of arrest: (a) the perceived likelihood of sanctions (*PLS*), here on the part of the offender that his or her partner will report domestic abuse, and (b) the likelihood that the partner will make the call reporting domestic abuse (*C*). First, arrest policy (*AP*) may lessen the likelihood that the offender believes the partner will call the police (*PLS*). Second, arrest

FIGURE 9

Domestic Violence: Graphing the Policy Intervention



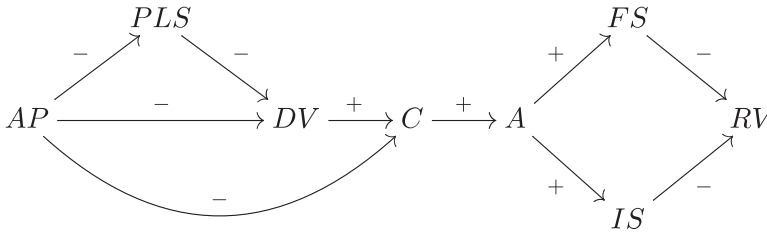
policy may lessen the actual probability that the victim will report abuse (negative arrow between *AP* and *C*). To the extent that these two mechanisms are in play, we could then witness an overall *decrease* in calls (*C*), while witnessing an *increase* in domestic violence (*DV*). Even if we assume that an increase in domestic violence will be associated generally with an increase in reports of abuse (positive between *DV* and *C*), the overall outcome will depend on the strength of the negative relationship between arrest policy and calls.

For example, if it is certain that the potential offender would be arrested if a call were made, that offender may not believe his or her partner would take such an action, especially if an arrest would lead to a loss of household income or even eviction for victims who are also tenants. In many cities including Milwaukee, the site of one of the domestic violence experiments, landlords can evict tenants who frequently call the police or house criminals. Based on so-called “crime-free housing” ordinances, the idea is to put responsibility on landlords to weed out disruptive tenants (Eckholm, 2013). Although 911 calls may in themselves be defined as disruptive, arrest is more so and leads to a criminal record of someone in or associated with the household. As the victim quoted at the beginning of this section reveals, these ordinances can thus dampen the willingness of citizens to call on the police for help, in turn leading to unintended consequences. In fact, Desmond and Valdez (2013: 117) found that a third of all evictions in Milwaukee involve domestic violence. Desmond (2012: 88) reaches a strong conclusion overall: “In poor black neighborhoods, eviction is to women what incarceration is to men: a typical but severely consequential occurrence contributing to the reproduction of urban poverty.”

It stands to reason that in addition to potential monetary losses (e.g., child support) there are real incentives for victims to *not* call the police under mandatory arrest regimes. It is further reasonable to assume that potential offenders are aware of these incentives and potential effects on victims’ behavior. With respect to our model, if potential offenders do not believe there will be a cost associated with their actions, then a decrease in perceived likelihood of calling may serve to increase domestic violence (negative arrow between *PLS* and *DV*). Therefore the behavior of both victims and offenders is potentially altered in

FIGURE 10

Combined Causal Graph of Domestic Violence



previously unanticipated and possibly countervailing ways once domestic violence policies are scaled up.

Perhaps no other result provides a more powerful reminder of the importance of considering the potential unintended consequences of moving from experiment to policy. In the case of the MDVE, it is clear that the policy intervention, and its unintended effects on the incentives and opportunities available to individuals within the new policy regime, must be theorized as part of the translation from experiment to policy. Noting the difference between the causal graph implicit in the initial experiment (Figure 7) and the processes graphed in Figures 8 and 9, we draw a combined causal graph in Figure 10. A result of the accumulation of replication, theory, and observational data, Figure 10 underscores that the relationship identified by the experiment—that between arrest and reoffending—constitutes just one part of an interlocking causal web. The effect of arrest policy (*AP*) is necessarily different from that of randomized arrest (*A*). It is conditioned by how arrest policy alters the legal structure in which offenders and victims act, by the mechanisms that impact how offenders react to arrest, and by the differences among individuals and their local contexts.

To summarize, the results of the initial MDVE must be understood as part of a larger research program. The first, single experiment provided evidence of a time-bound relationship between arrest and rates of reoffending. Yet, as research about the boundary conditions of this effect mounted, the contingencies of moving from treatment to policy became clearer. The MDVE body of research should thus be considered, from the point of view of research, a success. Multiple methods and theory were brought to bear on many independent streams of data in a sequential process. Through the accumulation of both experimental and observational studies, the simplistic causal graph of a treatment influencing an outcome was gradually transformed into a policy graph, one that gives information regarding mechanisms, heterogeneous effects, and the potential unintended consequences of scaling up.

In terms of policy, by contrast, the MDVE proved a mixed blessing. The initial experiment was taken as causal evidence of “what works.” Mandatory arrest was expanded before its potential effects were clear and follow-up studies revealed the contingency of results. Because policy makers did not pay sufficient attention to the accumulation of multiple data or the formulation of good theory, they implemented measures that, in many cases, proved to be counterproductive (Sherman, 1993; Sherman and Harris, 2013). It is probably unknowable how much net harm (or good) was done as a result of the early adoption of mandatory arrest policies.

Toward Complex Parsimony

Throughout this article, we have argued for a theoretically informed approach to policy, one that gives as much weight to understanding the social structure of criminal justice policy as it currently does to identifying causal effects within it. Although our focus has been on the interpretation of experimental results, our argument applies equally to problems in observational research design. Experiments and related designs will be most informative if researchers seriously consider theory even before data are collected. We have argued specifically that three key topics—*mechanisms*, *effect heterogeneity*, and *contextualization*—are central to the project of policy creation, although they often are addressed, if at all, as an afterthought, typically as either boundary conditions or details of the causal narrative. By contrast, we believe that these topics merit the core of our attention as criminologists at both the design and the analysis stage.

In exploring the role of mechanisms, heterogeneity, and context, we have adopted the use of causal graphs because we believe they make more explicit the causal claims and theory that are implicated in policy research. Just the drawing of a graph is an informative exercise for what at first seems to be a simple relationship. We also use these graphs because our principal goal in this article is to complicate underlying causal models—to make the policy intervention a part of the graph itself—that is, to create a “policy graph.” This representation can, we believe, eliminate or at least temper the implicit exportability claim of experimental results, making clear the conditionality of causal results and their place within a wider causal system. Furthermore, we agree with Ludwig et al.’s (2011) claim that experiments testing causal mechanisms can yield generalizable, policy-relevant information even if they test interventions that do not correspond to realistic policy options. Testing the visual cue hypothesis of broken windows theory by randomly assigning disorder (e.g., abandoned cars) in a field experiment is an example of this strategy. More generally, although it may seem counterintuitive, the best way to inform policy is not always to test policy (Ludwig et al., 2011: 30).

The ultimate principle is that to provide effective policy, causal effects must be understood within a larger organizational, political, and social structure. Causal graphs transformed into policy graphs provide one way of representing that structure and can

complement ongoing causal inquiry: Once we have a theoretical model of a given causal structure, we can go about the complicated task of estimating parts of the causal web and understanding how its different components are related and interact with each other.¹³ Causal graphs and theory also can be used to design experiments themselves and test hypothesized policy mechanisms (Ludwig et al., 2011). A key strength of our recommended approach is that to understand whether a particular causal effect is identified, *we do not need to estimate the whole system*. What we do need to do is to draw on theory to gain an understanding of how parts of the system are stitched together. Otherwise, we are left with static, detached segments of purported causal relationships; such a balkanized view of reality tells very little about the dynamic relationships of complex causal systems. Moreover, it is only by asking questions about the larger causal structure that we can examine when and how a set of experimental results generalizes to other situations. In the context of policy, we must thus reject the separation of forward and backward causality—to understand policy going forward cannot be divorced from a “backward-looking” understanding of the causes of effects.

It should be emphasized again that descriptive data and noncausal analysis constitute a crucial part of the construction and evaluation of a policy graph. Consider, for example, that much of our knowledge on “broken windows” noted earlier was based on careful observation rather than on experimentation or counterfactual causal analysis. Or consider the important gains derived from meta-analyses, demographic-like analysis of stocks and flows through the criminal justice system, and research on the previously neglected links between high rates of incarceration and inequality (Western, 2006) that motivated the hypothesized links in Figure 6. More generally, noncausal analysis of offender patterns derived from longitudinal research on the life course is important for understanding temporal heterogeneity and informing sentencing policy. Blumstein and Nakamura (2009), for example, used criminal-history information to develop longitudinal estimates of expected career lengths and what they called “redemption times.” These estimates make no causal claims but are relevant to building a rational policy on setting release times from prison and for employment policy concerning ex-offenders. Descriptive data and theoretically driven analysis thus form essential building blocks of causal policy graphs.

Conclusion

Policy research must be concerned with much more than providing policy makers with information about average causal effects. As we have argued, “what works” does not reduce

13. A mature criminological science that combines causal evidence with theory and the accumulation of knowledge across multiple studies and different contexts may lead to areas of consensus that can yield reasonably strong policy inferences. This seems to be occurring in policing research, as reviewed by Nagin and Weisburd (2013, this issue) and Sherman (In press).

to the estimation of a causal effect, however precisely measured. *We must instead separate criminology's increasing focus on causality from its policy turn, recognizing that the latter requires a different standard of theory and evidence than does the former.* We must in turn be willing to ask questions related to mechanisms, heterogeneous effects, and social context, and all of the real-world phenomena to which these difficulties give rise—such as the possibility of unintended consequences, of policies that change incentive and opportunity structures, and more. There are tools that can enhance this goal, hopefully leading to a set of topics, all of which must be addressed, by researchers who wish to inform policy in a causally uncertain world.¹⁴

We thus agree with Blomberg et al. (2013) that causal uncertainty does not negate criminological contributions to policy. On the contrary, causal uncertainty can be the subject of investigation and criminologists can do a better job of making explicit their assumptions. It is here that concepts, theory, and descriptive analyses—including insights from practitioners themselves—are essential (Laub, 2012; Sherman, in press). Indeed, it may be that for some of the issues raised in this article, practitioners (e.g., cops on the beat) may be better “theorists” of what policy changes will trigger on the ground than academic criminologists who theorize at a considerable remove. Along with a commitment to conducting ongoing systematic research or experiments, criminal justice agencies can be co-producers of the sort of feedback information that is necessary to address questions of heterogeneity, mechanisms, and context.

Finally, although James Q. Wilson might have been right in 1975 to note the inability of criminal justice institutions to change root causes, he was wrong, we think, to suggest the casual irrelevance of criminological theory. Even the most “root”-like causes, such as concentrated disadvantage, unemployment, and legacies of racial inequality are central to our understanding of how policies will likely be received by those subjected to intervention by the state—and, therefore, how the policies seemingly recommended by experiments or other causal analyses are likely to work in practice and in the future. The domestic violence results are a clear case in point. Observed in this light, translational criminology is a *process* that entails the constant interplay of theory, research, and practice. Evidence, even if causal, does not “speak for itself” to policy.

References

- Aalen, Odd O. and Arnaldo Frigessi. 2007. What can statistics contribute to a causal understanding? *Scandinavian Journal of Statistics*, 34: 155–168.

14. We set aside the disturbing reality that in today's political climate, many policy makers are openly hostile to science and seek to avoid any policy-relevant criminology that conflicts with prior world views. The assertion of values over evidence is a deep issue that needs to be confronted by criminologists, but it is well beyond the scope of our article.

- Barinboim, Elias and Judea Pearl. 2013. Causal transportability with limited experiments. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press.
- Becker, Howard. 1963. *Outsiders: Studies in the Sociology of Deviance*. New York: Free Press.
- Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western. 1992. The deterrent effect of arrest in incidents of domestic violence: A Bayesian analysis of four field experiments. *American Sociological Review*, 57: 698–708.
- Blomberg, Thomas G., Julie Mestre, and Karen Mann. 2013. Seeking causality in a world of contingency: Criminology, research, and public policy. *Criminology & Public Policy*, 12: 571–584.
- Blumstein, Alfred and Kiminori Nakamura. 2009. “Redemption” in an era of widespread criminal background checks. *Criminology*, 47: 327–359.
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, Kenneth A. and Judea Pearl. 2013. Eight myths about causality and structural equation models. In *Handbook of Causal Analysis for Social Research*. New York: Springer.
- Braga, Anthony A. and Brenda J. Bond. 2008. Policing crime and disorder hot spots: A randomized controlled trial. *Criminology*, 46: 577–607.
- Cartwright, Nancy. 2007. Are RCTs the gold standard? *Biosocieties*, 2: 11–20.
- Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford, U.K.: Oxford University Press.
- Cohen, Lawrence E. and Marcus Felson. 1979. Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44: 588–608.
- Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- Cunningham, Scott, Benjamin Engelstatter, and Michael Ward. 2011. Understanding the effects of violent video games on violent crime. Retrieved from ssrn.com/abstract=1804959.
- Davis, Richard L. 2008. *Domestic Violence: Intervention, Prevention, Policies and Solutions*. Boca Raton, FL: CRC Press.
- Desmond, Matthew. 2012. Eviction and the reproduction of urban poverty. *American Journal of Sociology*, 118: 88–133.
- Desmond, Matthew and Nicol Valdez. 2013. Unpolicing the urban poor: Consequences of third-party policing for inner-city women. *American Sociological Review*, 78:117–141.
- Duncan, Otis D. 1966. Path analysis: Sociological examples. *American Journal of Sociology*, 72: 1–16.
- Duncan, Otis D. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.
- Duneier, Mitchell. 1999. *Sidewalk*. New York: Farrar, Straus & Giroux.
- Eckholm, Erik. 2013. Victims’ dilemma: 911 calls can bring eviction. *New York Times*. A1, August 17.

- Elwert, Felix. 2013. Graphical causal models. In (Steven L. Morgan, ed.), *Handbook of Causal Analysis for Social Research*. New York: Springer.
- Fagan, Jeffrey and Garth Davies. 2000. Street stops and broken windows: Terry, race, and disorder in New York City. *Fordham Urban Law Journal*, 28: 457–504.
- Granger, Robert C. 2011. The Big Why: A learning agenda for the scale-up movement. *Pathways Magazine*, Winter: 28–32.
- 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.
- Harcourt, Bernard E. 2001. *Illusion of Order: The False Promise of Broken Windows Policing*. Cambridge, MA: Harvard University Press.
- Harcourt, Bernard E. and Jens Ludwig. 2006. Broken windows: New evidence from New York City and a five-city social experiment. *University of Chicago Law Review*, 73: 271–320.
- Heckman, James J. 2005. The scientific model of causality. *Sociological Methodology*, 35: 1–97.
- Heckman, James J. 2008. Econometric causality. *International Statistical Review*, 76: 1–27.
- Heckman, James J., Rodrigo Pinto, and Peter Savelyev. In press. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*.
- Heckman, James J. and Jeffrey Smith. 1995. Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9: 85–110.
- Holland, Paul. 1986. Statistics and causal inference. *Journal of the American Statistical Association*, 81: 945–970.
- Iyengar, Radha. 2007. Does the certainty of arrest reduce domestic violence? Evidence from mandatory and recommended arrest laws. NBER Working Paper No. 13186. Cambridge, MA: National Bureau of Economic Research.
- Kelling, George L. and William H. Sousa Jr. 2001. Do police matter? An analysis of the impact of New York City's police reforms. Civic report. New York: CCI Center for Civic Innovation at the Manhattan Institute.
- Kirk, David S. 2009. A natural experiment on residential change and recidivism: Lessons from Hurricane Katrina. *American Sociological Review*, 74: 484–505.
- Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz. 2007. Experimental analysis of neighborhood effects. *Econometrica*, 75: 83–119.
- Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz. 2005. Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics*, 120: 87–130.
- Knight, Carly and Christopher Winship. 2013. The causal implications of mechanistic thinking: Identification using directed acyclic graphs (DAGS). In (Steven L. Morgan, ed.), *Handbook of Causal Analysis for Social Research*. New York: Springer.

- Kubrin, Charis E. and Eric A. Stewart. 2006. Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology*, 44: 165–197.
- Laub, John H. 2004. The life course of criminology in the United States: The American Society of Criminology 2003 Presidential Address. *Criminology* 42: 1–26.
- Laub, John H. 2012. Translational criminology. *Translational Criminology*, Fall: 4–5.
- Leaf, Clifton. 2013. Do clinical trials work? *New York Times*. SR1, July 14.
- Lemert, Edwin M. 1951. *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior*. New York: McGraw-Hill.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25: 17–38.
- Manski, Charles F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.
- Mignon, Sylvia I. and William M. Holmes. 1995. Police response to mandatory arrest laws. *Crime & Delinquency*, 41: 430–442.
- Miller, Susan. 2005. *Victims as Offenders: The Paradox of Women's Violence in Relationships*. New Brunswick, NJ: Rutgers University Press.
- Morgan, Stephen and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Nagin, Daniel S. and David L. Weisburd. 2013. Evidence and public policy: The example of evaluation research in policing. *Criminology & Public Policy*, 12: 651–679.
- Neyman, Jerzy. 1935. Statistical problems in agricultural experimentation. *Journal of the Royal Statistical Society Series B*, 2: 107–180.
- Neyman, Jerzy. 1990 [1923]. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5: 465–480.
- Pearl, Judea. 2009 [2000]. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K.: Cambridge University Press.
- Rein, Martin and Christopher Winship. 1999. The dangers of “strong” causal reasoning in social policy. *Society*, 36: 38–46.
- Rosenzweig, Mark R. and Kenneth I. Wolpin. 2000. Natural “natural experiments” in economics. *Journal of Economic Literature*, 38: 827–874.
- Rubin, Donald B. 1974. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66: 688–701.
- Rubin, Donald B. 1978. Bayesian-inference for causal effects: The role of randomization. *Annals of Statistics*, 6: 34–58.
- Sampson, Robert J. 2011. The incarceration ledger: Toward a new era in assessing societal consequences. *Criminology & Public Policy*, 10: 819–828.
- Sampson, Robert J. 2013. The place of context: A theory and strategy for criminology's hard problems: 2012 Presidential Address to the American Society of Criminology. *Criminology*, 51: 1–31.

- Sampson, Robert J. and Jacqueline Cohen. 1988. Deterrent effects of the police on crime: A replication and theoretical extension. *Law & Society Review*, 22: 163–190.
- Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points through Life*. Cambridge, MA: Harvard University Press.
- Sampson, Robert J. and Stephen W. Raudenbush. 1999. Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology*, 105: 603–651.
- Sampson, Robert J. and Stephen W. Raudenbush. 2004. Seeing disorder: Neighborhood stigma and the social construction of broken windows. *Social Psychology Quarterly*, 67: 319–342.
- Sanbonmatsu, Lisa, Jens Ludwig, Lawrence F. Katz, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade, and Stacy Tessler Lindau. 2011. *Moving To Opportunity For Fair Housing Demonstration Program: Final Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development.
- Sherman, Lawrence W. 1993. Defiance, deterrence, and irrelevance: A theory of the criminal sanction. *Journal of Research in Crime and Delinquency*, 30: 445–473.
- Sherman, Lawrence W. In press. The rise of evidence-based policing: Targeting, testing, and tracking. *Crime and Justice in America: 1975–2025*.
- Sherman, Lawrence W. and Richard A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. *American Sociological Review*, 49: 261–272.
- Sherman, Lawrence W. and Heather Harris. 2013. Effects of arrest over the life-course: a 24-year follow-up of the Milwaukee domestic violence experiment. Paper presented at The Stockholm Criminology Symposium, Stockholm, Sweden, June 11.
- Sherman, Lawrence W., Douglas A. Smith, Janell D. Schmidt, and Dennis P. Rogan. 1992. Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review*, 57: 680–690.
- Tittle, Charles R. 2004. The arrogance of public sociology. *Social Forces*, 82: 1639–1643.
- VanderWeele, Tyler J. 2009. Mediation and mechanism. *European Journal of Epidemiology*, 24: 217–224.
- Western, Bruce. 2006. *Punishment and Inequality in America*. New York: Russell Sage Foundation.
- Wikström, Per-Olof H. 2011. Does everything matter? Addressing the problem of causation and explanation in the study of crime. In (J. M. McGloin, C. J. Sullivan, and L. W. Kennedy, eds.), *When Crime Appears: The Role of Emergence*. London, U.K.: Routledge.
- Wikström, Per-Olof H., Dietrich Oberwittler, Kyle Treiber, and Beth Hardie. 2012. *Breaking Rules: The Social and Situational Dynamics of Young People's Urban Crime*. Oxford, U.K.: Oxford University Press.
- Wilson, James Q. 1975. *Thinking About Crime*. New York: Random House.
- Wilson, James Q. and George Kelling. 1982. Broken windows: The police and neighborhood safety. *Atlantic*, 127: 29–38.

Court Case

Thurman v. City of Torrington 595 F.Supp. 1521 (1985).

Robert J. Sampson is Henry Ford II Professor of the Social Sciences at Harvard and Past President of the American Society of Criminology. His research focuses on crime, urban inequality, the life course, neighborhood effects, and the social structure of cities. His most recent book, *Great American City: Chicago and the Enduring Neighborhood Effect*, was published in paperback in 2013 by the University of Chicago Press.

Christopher Winship is Diker-Tishman Professor of Sociology at Harvard and a faculty member in the Kennedy School of Government. In criminology, he has studied The Ten Point Coalition, a group of black ministers working with the Boston police to reduce youth violence, and changes in the racial differential in imprisonment rates. With Stephen Morgan, he is the author of *Counterfactuals and Causal Inference* (Cambridge, 2007).

Carly Knight is a PhD student in the Department of Sociology at Harvard and a Doctoral Fellow with the Multidisciplinary Program in Inequality and Social Policy. Her research interests include comparative/historical sociology, economic sociology, criminology, social mechanisms and causality, and the philosophy of science.

Family-Focused Interventions to Prevent Juvenile Delinquency

A Case Where Science and Policy Can Find Common Ground

Abigail A. Fagan

University of Florida

Research Summary

Several criminological theories prioritize the role of the family in influencing juvenile delinquency, and both observational studies (Derzon, 2010; Hoeve et al., 2009; Steinberg, 2001) and experimental evaluations of family-focused interventions (Farrington and Welsh, 2003; Sandler, Schoenfelder, Wolchik, and MacKinnon, 2011) have demonstrated that parenting practices affect the degree to which children and adolescents will engage in substance use, delinquency, and/or violence. Moreover, research regarding the ability of family-focused interventions to reduce offending when implemented in “real-world” conditions is mounting.

Policy Implications

Scientific experiments are increasingly being used to identify why, for whom, and under what conditions treatments are most effective, information that Sampson, Winship, and Knight (2013) in the main article of this special issue emphasize is needed to provide assurances that programs will work when replicated in communities. Based on this evidence, should criminologists advocate for policies and practices intended to increase the dissemination of effective family-based programs? This essay contends that there is sufficient evidence to do so. To ensure that such policies produce desired reductions in juvenile delinquency, however, criminologists also must reach consensus on the evidentiary standards needed to identify “what works,” advocate for increased

The author would like to acknowledge Dr. Chris Gibson and the anonymous reviewers for their helpful feedback on the substantive issues discussed in this essay. Direct correspondence to Abigail A. Fagan, Sociology and Criminology & Law, 3219 Turlington Hall, P.O. Box 117330, University of Florida, Gainesville, FL 32611-7330 (e-mail: afagan@ufl.edu).

attention to implementation fidelity during program replication, and increase support for preventive interventions among policy makers and the public.

Keywords

juvenile delinquency, delinquency prevention, evidence-based programs, family-based interventions, parent training

Blomberg, Mestre, and Mann (2013) note in their introduction to this special issue that “for criminology . . . it has not yet been possible to isolate a single cause of crime.” Instead, a variety of factors have been posited by criminological theories and shown in empirical research to affect delinquency. Moreover, these factors are likely to co-occur and interact with one another, making it difficult for criminologists to present a simple, parsimonious explanation of offending or to offer easy solutions for how to reduce crime. In addition, the degree to which any one factor is “causally” related to offending is unclear, leading many criminologists to avoid making public policy recommendations for how to lower rates of criminal behavior.

Rather than be disappointed by our inability to isolate the root cause(s) of crime and reluctant to help shape crime-control policies, I advocate in this policy essay that criminologists make immediate and “rational” (Mears, 2007) policy recommendations based on the best available scientific evidence (Blomberg et al., 2013) we have regarding the factors that contribute to juvenile offending and its prevention. Specifically, I argue for policies intended to increase the use of evidence-based preventive interventions that seek to alter one or more of the individual, school, peer, family, and community risk and protective factors shown in scientific research to affect delinquency (Farrington, 2000; Hawkins, Catalano, and Miller, 1992; Herrenkohl et al., 2000; Lipsey and Derzon, 1998). Much progress has been made in the development and evaluation of such strategies, and growing numbers of interventions have been shown to reduce juvenile substance use, delinquency, and violence by targeting criminogenic influences (Andrews et al., 1990; Farrington and Welsh, 2005; Sherman, Gottfredson, MacKenzie, Reuter, and Bushway, 1998; U.S. Department of Health and Human Services, 2001).

Rather than reviewing the extensive body of evaluation research in this area, this policy essay focuses on the development, testing, and potential policy implications of family-focused delinquency prevention programs. The essay begins by discussing the theoretical and evidentiary basis of these programs; that is, the degree to which family-focused programs are based on criminological theories and empirical evidence regarding family influences on delinquency, and findings from experimental evaluations testing the degree to which targeted parenting practices¹ are associated with reductions in delinquency and substance use.

1. Although structural features of the family (e.g., family structure, size, and socioeconomic status) also have been shown to affect juvenile delinquency, this article focuses on parenting practices because

The second section of this policy essay reviews additional evidence from experimental evaluations of family-focused interventions to address the concerns raised by Sampson, Winship, and Knight (2013) in this issue. They assert that treatments designed to target the “root causes” of crime must establish not only their technical ability to reduce offending (e.g., showing effectiveness via experiments with strong internal validity), but also their ability to produce results when replicated widely, under “real-world” conditions. Specifically, they contend that, prior to the dissemination of interventions, there must be demonstration of the mechanisms or causal pathways that produce reductions in delinquency, as well as identification of the populations for whom and conditions under which treatments have their anticipated effects (i.e., “effect heterogeneity” and “context”). Although Sampson et al. (2013) express some doubts that experimental research can meet these criteria, experiments are increasingly being used to examine these issues, and this progress will be discussed using examples from family-focused interventions.

The conclusion of this policy essay identifies policies that have already been put into place to increase the use of family-focused programs and discusses the necessary next steps for ensuring that such policies lead to desired reductions in delinquency. Recommendations include establishing consensus among scientists regarding the evidentiary standards needed to determine intervention effectiveness, advocating for implementation integrity during program replication, and increasing support for preventive interventions among policy makers and the public.

Theoretical Explanations of Family Influences on Delinquency

Delinquency prevention programs are most likely to be effective when they target for change the factors identified in criminological theories and shown in empirical studies to influence offending (Farrington, 2000). Although they may disagree on the family factor(s) that is *most important* in affecting delinquency, several criminological theories emphasize the importance of parenting practices in directly and indirectly influencing delinquency. To summarize briefly, social learning theories (Akers, 1985, 2009; Bandura and Walters, 1959; Patterson, 1982) contend that children learn positive and negative behaviors via interaction with others, and that parents are particularly important for influencing children’s prosocial and antisocial behavior. Parents who endorse attitudes favoring deviant behavior or who fail to correct children’s misbehavior increase the likelihood that children will view delinquent activities as acceptable to achieve certain outcomes, particularly when they perceive more benefits than consequences from engaging in delinquency. Moreover, parents

they tend to have a more prominent role in criminological theories compared with structural influences, have been shown to mediate the impact of family structure on delinquency (Kotchick & Forehand, 2002; Sampson & Laub, 1994; Stern & Smith, 1995), and are likely to be more amenable to change than are structural factors.

who engage in deviance or law-breaking provide children with models of illicit behavior to emulate.

Self-control theory (Gottfredson and Hirschi, 1990) contends that family influences are largely indirect: Parents who actively monitor children's behavior, set and communicate clear expectations that delinquency is not acceptable, and reward compliance (and punish transgressions) instill high levels of self-control in children, which in turn reduces the likelihood of youth offending. Social bonding theory (Hirschi, 1969) places more emphasis on direct effects of parenting practices, particularly affective bonds between parents and children. When parents show affection for their children, communicate effectively with them, provide them opportunities to be involved in the family, and positively recognize their children for displaying positive behaviors and refraining from delinquency, parents create strong, positive relationships with their children (Catalano and Hawkins, 1996). Youth who are attached to parents refrain from problem behaviors because they do not want to risk losing their approval and affection (Hirschi, 1969).

General strain theory (Agnew, 1992, 2006) includes family stress such as parent/child conflict and child maltreatment among the forms of strain most likely to lead to youth deviance. These negative experiences are likely to evoke strong negative emotions (e.g., anger, anxiety, or depression) among children. In turn, youth may engage in substance use, aggression, or other forms of delinquency to try to minimize and/or alleviate these feelings (Brezina, 1998; Hay, 2003).

Finally, integrated and life-course theories of crime (Catalano and Hawkins, 1996; Elliott, Huizinga, and Ageton, 1985; Sampson and Laub, 1993; Thornberry, 1987) assert that the degree to which parents model positive behavior; supervise and effectively discipline children; and establish close, affective bonds with youth are all important in shaping delinquency. Nonetheless, as children age, seek independence from parents, and form relationships with others, parental influences wane in importance. In a similar vein, Moffitt (1993) posited that adverse family conditions during the prenatal and early childhood periods (e.g., maternal drug use during pregnancy or child maltreatment), as well as genetic influences, are most important in the development of life-course persistent offending, while exposure to deviant peers is the most salient predictor of (less serious) delinquency limited to the adolescent period. Patterson (1982) contended that parenting practices have an immediate effect on children's misbehavior and longer term, indirect influences on delinquency, via children's school and peer experiences. When parents ignore or inappropriately react to children's misbehavior, they reinforce and increase the likelihood of its continuation. Such children may be more likely to act out in school, which in turn increases the probability of negative interactions with teachers, poor school performance, rejection from positive peers, and increased contact with delinquent peers. These are all risk factors in their own right, and the accumulation of these negative circumstances makes delinquency during mid-to-late adolescence much more likely.

Empirical Evidence of Family Influences on Delinquency

A vast body of research has investigated the degree to which parenting practices identified as important in criminological theories impact youth delinquency. Two meta-analyses (Derzon, 2010; Hoeve et al., 2009) of cross-sectional and longitudinal studies each reported average effect sizes on delinquency ranging from about 0.15 to 0.20 for the parenting practices under consideration, including child maltreatment; parental monitoring, supervision, and discipline of children; parental affection and support; and parental criminal behavior and/or deviant attitudes. Although small to moderate in size, these effects are comparable with those found in meta-analyses of other primary risk factors for delinquency, including low self-control (Pratt and Cullen, 2000), differential association and reinforcement (Pratt et al., 2010), and macro- or neighborhood-level predictors (Pratt and Cullen, 2005). Moreover, the studies included in the meta-analyses varied considerably in terms of sample characteristics, measurement instruments, analysis strategies, and inclusion of other risk and protective factors, indicating support for a general relationship between parenting behaviors and delinquency.

Nonetheless, findings based on meta-analyses cannot be used to make claims regarding the “causal” impact of family influences on offending. Even longitudinal studies, which can analyze the impact of parenting practices on children’s *subsequent* behaviors and/or investigate the degree to which changes in the family affect changes in delinquency, are limited in their ability to establish definitively parenting as a “root cause” of crime. Causal claims are predicated on the ability to account in full for and rule out other potential explanations of delinquency. Because multiple factors have been shown to impact adolescent offending, this task is challenging. Furthermore, children’s individual characteristics (e.g., temperament) are likely to impact both parenting practices and delinquency (Harris, 1995; Thornberry, 1987), making it difficult to identify cause and effect. In addition, some research has suggested that the relationship between parenting practices and delinquency is attributable more to heritability than to socialization practices (Rowe, 1986; Wright and Beaver, 2005), and failure to take genetics into account may produce inflated estimates of the impact of parenting practices on delinquency. Finally, the ability to isolate the impact of parenting practices occurring early in life is further challenged by the fact that many other important events (i.e., “turning points” and “transitions”; see: Sampson and Laub, 1993) can intervene and affect offending.

Thus, even the most well-conducted observational study is limited in its ability to ascertain causality, as “no amount of statistical control can completely eliminate the possibility that an unknown third variable . . . causes . . . antisocial behavior” (Conduct Problems Prevention Research Group, 2002: 926). Experimental studies, which seek to change parenting practices and then assess subsequent changes in children’s behaviors, are better equipped to investigate causality. Experiments, particularly randomized controlled studies, not only establish temporal ordering between the independent and dependent variables of interest

but also minimize contamination by other potential influences, which should be equally distributed across individuals in treatment and control groups (assuming successful randomization to conditions) (Sherman, 2003; Weisburd, 2010). Observed changes in behaviors among those participating in the experimental versus control condition can then be attributed with much more certainty to the intervention (i.e., to the change in the targeted parenting practice), rather than to other preexisting or intervening biological, individual, social, or contextual factors (Zhou, Sandler, Millsap, Wolchik, and Dawson-McClure, 2008). Although there are certainly limitations on the ability of randomized trials to establish causality (Hough, 2010; Mears, 2007; Sampson, 2010), experiments likely represent the most scientifically rigorous approach to identifying the “root causes” of crime (Farrington and Welsh, 2005; Sampson et al., 2013; Sandler, Schoenfelder, Wolchik, and MacKinnon, 2011).

Although not all crime-related problems can be prevented via experimental manipulation, much progress has been made in developing and testing in quasi-experimental and randomized trials interventions that prevent juvenile delinquency by targeting for change modifiable risk and protective factors (Catalano et al., 2012; Fagan and Eisenberg, 2012). This research includes substantial attention to the development and testing of family-focused interventions. The specific content of these interventions varies depending on the program type and the developmental age of the child, but on the whole, these programs seek to alter the family-related risk and protective factors discussed in theories and shown in empirical studies to influence delinquency and related problem behaviors.² Such strategies include home visitation programs (in which trained health-care providers support new mothers in the care and nurturing of their babies), early childhood education (which encourages parents’ active participation in children’s learning), parent training interventions (typically emphasizing effective limit-setting, monitoring, and reinforcement of positive and negative behaviors, as well as parent/child bonding and attachment), and intensive family therapy programs for youth offenders (which may target family as well as peer and environmental influences on offending).

Programs falling into each of these types of services have been demonstrated to reduce youth delinquency in high-quality research evaluations. For example, the Blueprints for Healthy Youth Development website (colorado.edu/cspv/blueprints/), known for nominating programs as effective only if they meet rigorous evaluation criteria, includes home visitation, early childhood education, parent training, and family therapy programs. Of the 47 programs on the list, 16 (34%) are family-based interventions, and these programs have each shown significant effects in reducing substance use, delinquency, and/or violence. Moreover, some have demonstrated very long-term benefits for participants. For example, the Nurse-Family Partnership has shown reductions in arrests and convictions among

2. For more detailed reviews of the content and effectiveness of family-focused and parent training interventions, see Kumpfer and Alvarado (2003) and Sandler et al. (2011).

children whose mothers were visited by nurses 13–15 years earlier compared with those whose families did not receive treatment (Olds, Henderson, and Cole, 1998). The Perry Preschool program has evidenced even longer sustained effects, with reductions in criminal arrests when research participants were 27 and 40 years of age, two to three decades after the intervention was delivered (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010).

Meta-analyses of evaluations of family-based programs also provide evidence of their ability to reduce youth offending. Among the 18 family-focused interventions that assessed delinquency, most of which were evaluated using randomized experiments, Farrington and Welsh (2003)'s review reported an overall effect size of 0.31. A meta-analysis (Piquero, Farrington, Welsh, Tremblay, and Jennings, 2009) of 55 randomized evaluations of programs that targeted parents of youth aged less than 5 years old reported an average mean effect size of 0.35 on child antisocial behavior or delinquency.

Although these meta-analyses indicated that family-focused interventions have the potential, on average, to reduce youth offending moderately, these effect sizes may be somewhat underestimated. Most evaluations in this area collect information on outcomes from only one child in the family, but it is likely that changes in parenting practices will benefit multiple children living in the household. An evaluation of the Functional Family Therapy program, for example, reported that siblings of youth who received intervention had significantly fewer contacts with the juvenile court compared with siblings of those in the control condition (Klein, Alexander, and Parsons, 1977). Also, it is possible that enhancements to the family environment can reduce risk factors and improve protective factors in other domains of influence (e.g., by reducing exposure to delinquent peers; Patterson, 1982; Steinberg, 2001), which also could contribute to reductions in youth offending, particularly over the long term.

Evidence Supporting “Real-World” Effectiveness of Family-Focused Interventions

Given their ability to establish temporal ordering and reduce contamination of other potential influences on delinquency, findings from well-conducted evaluation trials offer perhaps the most convincing evidence that parenting practices affect youth delinquency. Even more importantly, this information has significant practical value in providing specific methods, tools, and resources to assist parents in their efforts to socialize children in a positive way. Based on this evidence, then, it seems reasonable and rational to advocate for policies that will increase the use of effective family-focused interventions in communities.

Yet, is more caution needed prior to making such a recommendation? Randomized trials have been critiqued for having low external validity and for demonstrating results under artificial conditions that may bear little resemblance to the “real world” (Hough, 2010; Sampson, 2010). In addition, Sampson and colleagues (2013) in this special issue caution that experiments offer little insight into “what *will* work” if replicated in communities because they “make no claim to be able to estimate” three important conditions

that can affect replication results: *mechanisms and causal pathways*, *effect heterogeneity*, and *contextualization*.

Although it is true that there is much to be learned about how interventions work and their ability to maintain effects when widely replicated (Glasgow, Lichtenstein, and Marcus, 2003; Welsh, Sullivan, and Olds, 2010), there is more room for optimism than Sampson et al. (2013) suggest. In fact, since 2003, when the National Institutes of Health (Zerhouni, 2003) declared “translational research” a research priority, there has been increasing demand for experimental research to go beyond efficacy studies, which establish the basic impact of an intervention on an outcome, and focus more on effectiveness studies, which investigate treatment effects under real-world conditions, including the examination of mediating mechanisms and differential effects across individuals and contexts (Flay et al., 2005; Glasgow et al., 2003). The Institute of Medicine reinforced this call to action in a 2009 report, asking that evaluations of treatments designed to promote mental, emotional, and behavioral health be conducted “in a variety of settings in order to increase the knowledge base of what works, for whom, and under what conditions” (O’Connell, Boat, and Warner, 2009: 7). Following these directives, experiments designed to reduce delinquency and related problem behaviors have begun to examine more carefully the issues raised here by Sampson et al. (2013), so that we can be more confident that they *will* work when replicated in communities. The next section discusses this progress using examples from recent evaluations of family-focused interventions.

Mechanisms and Causal Pathways

To understand better the causal processes that lead to desired outcomes of an intervention, mediation analyses are required (Sampson et al., 2013; Sandler et al., 2011). Related to the current topic, such tests would examine the degree to which changes in parenting practices produce reductions in delinquency for participants in intervention versus control conditions. To do so, an evaluation would, for example, show that (a) the intervention reduced delinquency, (b) the intervention affected the targeted parenting practice(s), (c) the targeted parenting practice affected delinquency, and (d) the effect of the intervention on delinquency was completely or partially reduced (e.g., becomes nonsignificant) when the parenting practice was taken into account (Baron and Kenny, 1986).³

Table 1 provides illustrative examples of evaluations that have tested the degree to which targeted parenting practices mediated the effect of the preventive intervention on children’s antisocial behavior and/or delinquency. These programs were selected to represent as much diversity as possible in the developmental age of the target child (from preschool to adolescence), program type and content (e.g., early childhood education programs and

3. This methodology for testing for mediation is rather simplistic, and more methodologically advanced techniques are available; see, for example, MacKinnon (2008) and Preacher and Kelley (2011).

T A B L E 1

Examples of Mediation Analyses in Evaluations of Family-Focused Delinquency Prevention Programs

Program (Citation)	Program Description	Target Population (at Baseline)	Statistically Significant Effects on Parenting Practice(s) for Participants in Intervention vs. Control Conditions	Significant Child Behavior Outcomes (for Intervention vs. Control Conditions) Mediated by Changes in Parenting Practices
Child-Parent Center Program Study 1: Reynolds, Ou, and Topitzes (2004) Study 2: Reynolds and Ou (2011)	2 years of intensive preschool and family support services	Low income, mostly African American families in Chicago; children aged 3	<i>Study 1</i> : Increased parent and teacher reported parental participation at school at 8–12 years of age <i>Study 2</i> : Reduced child maltreatment at 4–17 years of age	<i>Study 1</i> : Reduced likelihood of any juvenile arrests by 18 years of age <i>Study 2</i> : Reduced likelihood of any felony arrest from 18 to 24 years of age
Oregon Divorce Study Study 1: DeGarmo and Forgatch (2005) Study 2: Forgatch, Patterson, DeGarmo, and Beldavs (2009)	14 weekly small group parent training sessions	Low income, mostly White single mothers with sons in grades 1–3	<i>Studies 1 and 2</i> : Increased observed "effective parenting" (support, monitoring, and discipline) in grades 2–4	<i>Study 1</i> : Reduced growth in teacher-reported delinquency (grades 4–7) <i>Study 2</i> : Reduced growth in teacher-reported delinquency and the number of juvenile arrests by grade 12
Coping Power Program Lochman and Wells (2002)	1.25-year school curriculum to improve child social skills; 16 small group parent training sessions	Boys identified by teachers as having behavioral problems in grade 6; 66% African American	Increased child-reported inconsistent parental discipline at posttest	Reduced self-reported delinquency at grade 7

Continued

T A B L E 1

Continued

Program (Citation)	Program Description	Target Population (at Baseline)	Statistically Significant Effects on Parenting Practice(s) for Participants in Intervention vs. Control Conditions	Significant Child Behavior Outcomes (for Intervention vs. Control Conditions) Mediated by Changes in Parenting Practices
Multidimensional Treatment Foster Care Chamberlain et al. (2008)	16 weekly small group parent training sessions	Foster families with children aged 5–12; ethnically diverse	Increased foster parent reported use of positive reinforcement and discipline at posttest	Reduced parent reported child behavior problems at posttest
Strong African American Families (SAAF) Brody, Kogan, Chen, and McBride Murry (2008)	7 weekly small group sessions for parents and children	Primarily low-income African American families in rural GA with children in grade 6	Increased positive parenting (management, supervision, attachment, communication) at posttest	Reduced conduct problems (delinquency, delinquent peers, and low self-control) at grade 8
Familias Unidas Study 1: Prado et al. (2007) Study 2: Pantin et al. (2009)	1-year small group parent training sessions, home visits and parent/child sessions	Low-income Hispanic families with children in grade 8	Study 1: Increased positive parenting (involvement, support, communication) from grades 7–10 Study 2: Increased positive parenting (involvement, support, communication, peer monitoring) at posttest	Study 1: Reduced growth in self-reported smoking and hard drug use (but not alcohol use) from ages 14–17 Study 2: Reduced growth in self-reported substance use (alcohol and marijuana use) from ages 14–16

parent training interventions), and the types of mediators and outcomes assessed.⁴ All of the examples showed *some* evidence of mediation, although not all targeted parenting practices and/or outcomes may have changed. Also, it is important to note that, although progress has been made in the number of evaluations that test for mediation (Eddy and Chamberlain, 2000), many evaluations still do not report such tests, particularly for programs targeting parents of very young children or universal interventions intended for the majority of the population (Sandler et al., 2011). Furthermore, it is possible that null effects are less likely to be published (though, for exceptions, see Bernat, August, Hektner, and Bloomquist, 2007; Hiscock et al., 2008; Mason et al., 2009; Zhou et al., 2008).

To highlight two examples shown in Table 1, both of which are parent training programs but which target for recruitment different types of families, first consider the Oregon model of parent management training (PMTO) tested in the Oregon Divorce Study (DeGarmo and Forgatch, 2005; Forgatch et al., 2009). Given evidence that marital stress and disruption can significantly impair effective parenting practices, this program is intended for recently separated, low-income single mothers with sons in grades 1–3. The studies listed in Table 1 involved random assignment of 238 families to a 14-week intervention group and a no-contact control group, with the former meeting in small groups to discuss with other mothers strategies for reducing parent/child conflict and increasing the use of positive reinforcement (DeGarmo and Forgatch, 2005). Program participants were observed to show more positive interactions with their sons compared with the control group participants, including improvements in their use of positive encouragement, negative reinforcement, and supervision skills, over time, from baseline to 2.5 years after the intervention (DeGarmo and Forgatch, 2005). These skills mediated the effect of the intervention on teachers' reports of boys' delinquent behaviors 3.0 years (DeGarmo and Forgatch, 2005) and 8.5 years postintervention, and in the total number of arrests through 17 years of age (Forgatch et al., 2009).

Given the importance of familism and family support in the Hispanic culture, as well as elevated levels of substance use among Hispanic adolescents, the Familias Unidas program was designed to reduce adolescent substance use via family-focused sessions delivered to Hispanic parents and adolescents (Prado et al., 2007). The 1-year intervention includes small group sessions and home visits that allow discussion and practice of methods for increasing involvement, bonding, and communication between parents and youth. The content integrates Hispanic cultural norms, expectations, and values, such as the belief that family is most important and that parents are the leaders of the family with authority to make decisions (Prado et al., 2007). Two evaluations of the program involving primarily

4. All but one of the listed programs were evaluated using experimental research designs with random assignment to intervention and control conditions. The Child-Parent Center Program (Reynolds & Ou, 2011) used a quasi-experimental research design without random assignment to conditions. This program is cited because it is one of few early childhood programs to assess mediating mechanisms.

low-income Hispanic families from Miami with children in middle school have shown mediating effects for targeted parenting practices. In the first study (Prado et al., 2007), the intervention significantly improved family functioning (especially positive parenting and parent/child communication) for parents in the intervention versus control condition, and this effect mediated the intervention's impact on growth in self-reported smoking and hard drug use (but not alcohol use) from 13 to 16 years of age. In the second trial (Pantin et al., 2009), family functioning also improved for intervention versus control group parents, and this effect mediated positive intervention effects on substance use from 14 to 16 years of age, assessed by comparing self-reports from youth in intervention and control conditions.

These examples and the results of other analyses of mediation shown in Table 1 are important not only because they provide more confidence that these interventions work as designed, but also because they can provide policy makers and practitioners with more specific information regarding the family processes that produce changes in delinquency. Policy makers may be reluctant to endorse interventions that seek to "interfere" in family processes, but evidence that improvements in parenting practices can reduce delinquency and substance use should increase support for these interventions. Second, evidence showing that particular program elements (e.g., content that focuses on enhancing positive parenting and parent/child communication) are responsible for producing effects can be shared with practitioners to reinforce the importance of delivering and not making alterations to these program components.

Heterogeneity in Effects

As discussed by Sampson et al. (2013) in this special issue, investigating effect heterogeneity, or the variation in treatment outcomes across individuals and/or groups, is important for both research and policy/practice. Evaluations that focus on estimating *average* causal effects may fail to detect results if some populations are benefitted by the intervention while others are not affected or are harmed. In addition, given limited human and financial resources, policy makers and practitioners will want to know whether a program works for all, some, or none of their populations of interest. If a program's effects are not highly generalizable, it may not be financially practical to implement, and if some groups are better served than others, recruitment efforts can be tailored to those individuals.

Effect heterogeneity is typically assessed by using statistical interaction terms to compare the degree to which targeted outcomes changed among those in intervention and control conditions for two or more groups. Many family-focused interventions have assessed heterogeneity in this manner; the examples in Table 2 represent only a subset of research in this area. The programs listed were purposively selected to represent different interventions from those listed in Table 1 and to illustrate the types of moderators usually assessed in such evaluations.

The most common approach to date has been to investigate differences in program effects across males and females, different racial/ethnic groups, or groups that differ in levels

T A B L E 2

Examples of Evaluations Investigating Heterogeneity of Outcomes Among Participants in Family-Focused Delinquency Prevention Programs

Program (Citation)	Program Type	Targeted Population (at Baseline)	Types of Moderators Investigated	Heterogeneity in Effects
Nurse-Family Partnership Eckenrode et al. (2010)	Home visitation program	Primarily single, low-income mothers in rural NY	Child gender	Greater reductions in self-reported arrests and convictions for females versus males at age 19
Strong African American Families (SAAF) Study 1: Brody et al. (2008) Study 2: Brody et al. (2009)	Parent training	Primarily low-income African American families in rural GA; youth in grade 6	Study 1: Low self-control exposure to delinquent peers Study 2: Genetic risk	Study 1: Greater reductions in conduct problems (delinquency, delinquent peers, and low self-control) for youth with lower self-control and more delinquent peers at baseline Study 2: Stronger effects on substance use for those at genetic risk
New Beginnings Wolchik et al. (2002)	Parent training for divorced mothers	Recently divorced, mostly White mothers in AZ; youth aged 9–12	High-risk youth: externalizing behaviors at baseline	Greater reductions in externalizing behaviors and substance use at ages 15–19 for those with greater externalizing at baseline

Continued

T A B L E 2

Continued

Program (Citation)	Program Type	Targeted Population (at Baseline)	Moderators Investigated	Heterogeneity in Effects
Strengthening Families Program (SFP 10–14) Spoth, Shin, Gylli, Redmond, and Azevedo (2006)	Parent training	White families in rural Iowa; youth in grade 6	High-risk youth: score of >4 on a summated measure of 10 family and youth risk factors assessed at baseline	Similar effects in reducing the initiation of alcohol use through age 18 for high- and low-risk groups
Staying Connected with Your Teen Haggerty, Skinner, Mackenzie, and Catalano (2007)	Parent training	African American and White families in Seattle; youth aged 13–14	Race/ethnicity Intervention delivery mode (self-administered program vs. group-based sessions)	At ages 15–16, the program reduced violence only among African American youth who received the self-administered version of the program
Multisystemic Therapy (MST) Henggeler, Melton, Brondino, Scherer, and Hanley (1997)	Intensive family therapy	Families of offenders referred to the SC juvenile justice system; mostly African American youth aged 11–17	MST treatment integrity	Greater reductions in offending and arrests for youth receiving treatment from therapists who more strongly adhered to MST protocols

of “risk” at the start of the intervention (e.g., high levels of risk factors or involvement in targeted behaviors at baseline). Although not depicted in Table 2 given space limitations, programs with long-term evaluations have sometimes shown inconsistent evidence of heterogeneity across time. For example, an evaluation of the Nurse-Family Partnership program indicated significant gender differences in intervention effects (e.g., comparing those whose mothers received services with those who did not) when children were 19 years of age, with stronger reductions in self-reported arrests and convictions for females than for males (Eckenrode et al., 2010; see Table 2). However, no gender differences in offending were evidenced when children were 15 years of age (Olds et al., 1998). Sampson and colleagues (2013) note here the importance of assessing heterogeneity in effects across development, as some outcomes may only emerge at particular developmental periods. This recommendation is particularly important for family-focused interventions that target parents of young children, who may not display behaviors such as substance use or violence until many years postintervention.

Because policy makers and practitioners often are interested in targeting the highest risk individuals with intervention services, evaluations sometimes test for differential effectiveness across youth considered to be at high and low risk for delinquency. For example, an evaluation of the Strong African American Families (SAAF) program found that the intervention was more effective for youth at higher risk for developing conduct problems and delinquency, operationalized as those who had elevated levels of low self-control and exposure to delinquent peers at the start of the study (Brody et al., 2008). Reductions in youth conduct problems (a latent variable based on child and parent reports of delinquency, exposure to delinquent peers, and low self-control) approximately 2 years postintervention for intervention versus control youth were stronger for those with lower levels of self-control and more affiliations with delinquent peers at baseline (Brody et al., 2008).

A second evaluation of the SAAF program (Brody, Beach, Philibert, Chen, and McBride Murry, 2009) with 11-year-old African American youth and their parents tested whether program effects varied according to genetic risk for problem behaviors, operationalized as those with one or two copies of the short allele (compared with those with the long allele) of the 5-HTTLPR, which is found in the promoter region of the SLC6A4 gene and which helps regulate the transmission of serotonin. The results demonstrated a stronger intervention impact for those who possessed this genetic vulnerability, who evidenced a slower rate of increase in a combined measure of drinking, smoking, marijuana use, and sexual intercourse from 11 to 14 years of age, compared with those with genetic risk who did not receive the program (Brody et al., 2009). This is one of very few interventions to test for gene-by-environment interactions in effects, but more such evaluations are likely to occur as interest in biosocial criminology—and its implications for policy and practice—increases.

Also, there is growing attention to the need to compare program outcomes produced under different implementation conditions. Most evaluations assess this type of impact in post hoc analyses. That is, if a process evaluation indicates significant variation in

treatment adherence across implementers, outcomes are compared for those who received the intervention under high-fidelity versus low-fidelity conditions. For example, Henggeler et al. (1997) found that mental health clinicians differed in the degree to which they adhered to Multisystemic Therapy (MST) treatment protocols, and the impact on participants' subsequent index offending, arrests, and incarceration was weaker for those who received therapy from low-adherence versus high-adherence therapists (see Table 2). A more rigorous test of implementation conditions occurs when individuals are assigned at baseline to different versions of the treatment. For example, Haggerty et al. (2007) randomly assigned families to receive a self-administered parent training program and a group-based program. Comparisons of teenagers in each intervention group with the control condition showed reductions in the perpetration of violence only among teenagers who received the self-administered program (and only among African Americans in this treatment group).

To summarize, heterogeneity in effects across participants and implementation conditions can and has been rigorously examined in experimental research. Importantly, although most evaluations in Table 2 evidenced some heterogeneity in effects, this pattern is not representative of all delinquency prevention or family-focused programs. In fact, there is much evidence of generalizability in effects across populations (Elliott and Mihalic, 2004). For example, the Multisystemic Therapy (MST) program, originally designed for families of youth already involved with the juvenile justice system, has been subject to numerous efficacy and effectiveness trials (Henggeler, 2011). Although these evaluations have demonstrated some disparity in outcomes (Farrington and Welsh, 2005), positive intervention effects have been found across studies taking place over three decades, in different regions of the United States, in different countries, using different types of control/comparison groups, and when implemented by highly supervised clinical graduate students and community therapists. Moreover, there is preliminary evidence of positive health and behavioral outcomes when the model has been implemented with families of physically abused children and those with serious health problems (Henggeler, 2011).

Context

The fact that MST has been replicated with very diverse populations and under very different implementation conditions also can be used to illustrate the importance of "context" discussed in this special issue by Sampson et al. (2013). They note that the larger environment in which youth reside can impact the likelihood of offending and the effectiveness of interventions designed to reduce offending in myriad ways. As it relates to family-focused interventions, evidence has shown that parents' ability to monitor children and provide them with emotional support is likely to be influenced by neighborhood context (Leventhal and Brooks-Gunn, 2000; Sampson and Laub, 1994), and that the impact of parenting practices on delinquency may vary according to the presence of risk and protective factors experienced in other areas of children's lives (Fagan, Van Horn, Jaki, and Hawkins, in press). These findings suggest that contextual factors may affect the ability of family-focused

interventions to reduce delinquency, and replications of such treatments under different environmental conditions can increase our understanding of how and when benefits will be realized, and when they will not.

One other illustration from MST intervention research is applicable to this discussion. In the 1970s, when MST was developed, most treatment services for delinquent youth involved one-on-one interactions between therapists and children and took place either in clinics or in detention facilities (Henggeler, 2011). MST, in contrast, involves the entire family in the treatment process and provides services in the home by therapists who are “on-call” during the 3 months of program delivery. This very large shift in the mode and context of delivery may be related to the program’s effectiveness, given that many “treatment” services offered by the juvenile justice system (e.g., probation, parole, or placement in a detention center) have not evidenced strong effects on youth recidivism (Howell, 2003).

The SAAF program provides an alternative example in which *homogeneity* in effects has been demonstrated across different contexts. The structure and format of this intervention was heavily based on the Strengthening Families Program (SFP) 10–14, an intervention originally designed for and shown to be effective at reducing adolescent substance use among White families in rural Iowa (Spoth and Redmond, 2002). Although some changes were made to the program content to be more culturally relevant for African American families living in the rural South (e.g., attention to the negative impact of racial discrimination), the SAAF program involved the same number, frequency, and format of program sessions as the original model, and involved both parents and adolescents (11–14 years of age) in program sessions. Positive program impacts on adolescent substance use have been found for the SAAF participants (Brody et al., 2009), just as they were for youth participating in the original SFP program.

Additional examples are needed to add to our understanding of how context may affect treatment effectiveness, and randomized evaluations can and should be used to provide this type of information. In fact, innovative research methods and analysis strategies are increasingly being developed to allow more of a comprehensive and nuanced understanding of how interventions work and how they are likely to work when replicated in communities (Spoth et al., 2006). For example, emerging “systems science” approaches are designed to identify the complex interactions that lead to behavioral change, including how changes produced at an individual level affect system-level processes or how social networks impact intervention effects (Hassmiller Lich, Ginexi, Osgood, and Mabry, 2013).

Moving from Science to Policy and Practice

Translational research seeks to ensure that scientific information is used to inform policies and practices to promote public health and well-being (Spoth et al., 2013). In criminology, the prevailing view several decades ago was that criminologists could and should do nothing to assist policy makers in their efforts to reduce crime largely because we had nothing to offer (Andrews et al., 1990; Cullen and Jonson, 2011). Although some of this sentiment

may remain (Blomberg et al., 2013), the scientific basis for the claim that “nothing works” (Martinson, 1974) and that criminologists can provide little information of value to the public has changed substantially. Especially when considering juvenile delinquency, we have come very far. Although criminologists may debate the extent to which they are “causal” predictors of crime, there is substantial evidence that family as well as individual, peer, school, and community risk and protective factors precede and affect the likelihood of youth offending (Hawkins et al., 1992; Herrenkohl et al., 2000; Lipsey and Derzon, 1998). Moreover, a growing number of interventions have been developed and shown in high-quality research trials to affect these factors and to reduce rates of substance use, delinquency, and violence (Farrington and Welsh, 2005; Sherman et al., 1998; U.S. Department of Health and Human Services, 2001).

As Blomberg et al. (2013) indicate in their introduction to this special issue, criminologists have a choice in deciding whether to make policy recommendations based on such information or waiting until we can identify with complete certainty what causes individuals to break the law. Sampson et al. (2013) in their main article go on to recommend even more caution prior to taking action, asserting that even when a crime-prevention program shows evidence of effectiveness in a well-conducted scientific trial, there is no certainty that effects will be replicated when implemented in the real world, when many other factors are likely to affect offending.

This policy essay contends that now is the time to make policy recommendations. Although it is important to proceed with gathering information regarding what causes crime and how to best prevent it, this knowledge will take time to accumulate. Mandating a full investigation of such issues prior to taking action would result in a potentially endless waiting period, for when would we know that we had discovered all the true mechanisms responsible for offending and all of the actual conditions that impinge on treatment effectiveness? In the meantime, policy makers and practitioners will move forward, and it is likely that their decisions regarding how to reduce crime will be based not on science, but rather on anecdotal evidence, what “seems right,” what has always been done, pressure from special interest groups, political or financial constraints, or any number of other reasons (Latessa, Cullen, and Gendreau, 2002). Moreover, some of the proposed solutions may actually harm youth, as shown in evaluations of “get-tough” policies and practices like waiving juveniles to adult courts (Task Force on Community Prevention Services, 2007) and Scared Straight programs (Petrosino, Turpin-Petrosino, and Buehler, 2003).

Ethically, if for no other reason, it seems incumbent on criminologists to call for increased use of science-based treatments that have been proven to reduce delinquency. As it pertains to this policy essay, many family-focused interventions have been shown in high-quality research trials to reduce youth delinquency by reducing risk factors and enhancing protective factors that exist in the family environment (Farrington and Welsh, 2003). Moreover, many of these programs have demonstrated via mediating analysis that effects were produced as a result of changing targeted family characteristics, and that many

interventions have been replicated and shown to be effective across diverse populations and settings. As others have concluded (Farrington and Welsh, 2003; Piquero et al., 2009), this evidence is enough to support a recommendation that policy makers call for increased dissemination of effective family-focused interventions across communities.

It is encouraging that some progress has already been made in this area, particularly given that, historically, there has been much reluctance to interfere in “private” family matters (Gelles, 1983). In 1999–2000, the Strengthening America’s Families initiative launched by the Office of Juvenile Justice and Delinquency Prevention (OJJPD) and assisted by the Center for Substance Abuse Prevention (CSAP) funded communities to replicate effective family-focused programs (Kumpfer and Alvarado, 2003). More recently, in 2011, the Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV) was established by the U.S. Department of Health and Human Services to fund states (via the Affordable Care Act) to implement effective home visitation programs. The federal government also has allotted significant recurring funding for early childhood education programs, such as Head Start and Early Head Start, to increase the dissemination of these services to high-risk families. Finally, states can now apply for a waiver to use the Title IV-E federal funding for foster care systems to support the implementation of innovative *prevention* services, such as the Multisystemic Therapy and Multidimensional Treatment Foster Care programs, which seek to avoid and/or minimize out-of-home placements by bolstering the family environment.

These examples provide some optimism that policy makers can be persuaded to enact legislation that increases the use of science-based, family-focused interventions. Nonetheless, a few words of caution are necessary. Most legislation of this type has advocated for the implementation of particular “types” of family-based services, such as early childhood education or home visitation programs. However, meta-analyses of programs have indicated substantial heterogeneity in the impact on delinquency across programs using the same general approach to family intervention (Farrington and Welsh, 2003; Piquero et al., 2009). For example, although some home visitation programs reduce criminal involvement, others have not shown evidence of reductions (Task Force on Community Prevention Services, 2005). That is, not all family-based services “work,” not all have provided evidence that parenting practices mediated the effects of the intervention on children’s outcomes, and not all have shown an ability to sustain effects over time (Sandler et al., 2011).

As a result, it is important that policies stipulate the use of *particular* interventions, not *types* of interventions, and that these programs have *strong evidence of effectiveness*. Although increased investment in effective, science-based treatments is an important step in producing widespread reductions in crime, achieving this goal requires additional action. As outlined in the final section of this policy essay, other critical work that needs to be undertaken while simultaneously lobbying for increased dissemination of evidence-based programs includes (a) agreeing on the standards of evidence required to establish that an intervention “works” to reduce crime; (b) requiring that interventions replicated in

community settings target appropriate populations and are implemented with integrity; and (c) increasing public support for crime *prevention*, as opposed to reactive measures like punishment and incarceration.

Reaching Consensus About “What Works”

One indicator of the significant progress made in recent years regarding the causes and prevention of offending is the proliferation of databases that seek to make accessible to the public information regarding “best practices” or “evidence-based interventions.” Ideally, policies intended to increase the use of effective interventions would simply require that agencies invest in the programs identified on lists such as the Office of Justice Program’s *Crime Solutions* database or the Substance Abuse and Mental Health Services Administration’s *National Registry of Evidence-based Programs and Practices* (NREPP). Unfortunately, the solution is not so simple because some lists include programs that do not have strong evidence of effectiveness (Gruner Gandhi, Murphy-Graham, Petrosino, Schwartz Chrismer, and Weiss, 2007). Because there is no consensus among scientists regarding the standards that must be met to demonstrate effectiveness, each list relies on different evidentiary standards and no two lists promote the same set of prevention practices.

The disparity makes it difficult for policy makers and practitioners to know which interventions should be used or to place their faith in the scientists whose work helped to create these lists. Rather than allowing such disagreements to impede policy recommendations, it is imperative that scientists strive to reach consensus on the standards for determining program effectiveness, and to ensure that standards are rigorous enough to achieve desired outcomes and avoid harmful consequences (Catalano et al., 2012; Elliott, 2012; Flay et al., 2005). As it pertains to evaluation research, this requires at a minimum agreeing on the methodological standards necessary to demonstrate the internal validity of an evaluation. Some of the most contentious issues debated by researchers are whether a randomized experiment is necessary (or if well-conducted quasi-experimental studies are sufficient), the degree to which baseline differences in participants exist and how these differences are treated, how much attrition is allowed and how this can vary across intervention conditions, and whether long-term effects on participants must be demonstrated (Biglan, Mrazek, Carnine, and Flay, 2003; Elliott, 2012; Gruner Gandhi et al., 2007). Other criteria that vary across lists focus on issues of external validity, such as whether interventions must be replicated (and how many times this must occur), mediating analyses must be conducted, effects heterogeneity or moderating analyses must be examined, and cost/benefit information calculated (Flay et al., 2005; Valentine et al., 2011).

Although it is important to agree on a set of standards that is rigorous enough to ensure that recommended interventions have the best possible chance of producing positive effects and of avoiding harmful outcomes, there is a trade-off in limiting the number of interventions available to communities. The Blueprints for Healthy Development website (colorado.edu/cspv/blueprints/), which arguably has the strictest criteria for determining

effectiveness, reserves the designation of “model” program for interventions that have at least one replication and sustained effects of at least 1 year postintervention. Because relatively few programs can meet this very high evidentiary standard, however, a “promising” designation is given to programs that have not yet been replicated or demonstrated sustained effects (Mihalic, Irwin, Elliott, Fagan, and Hansen, 2001). Nonetheless, programs in both categories must clearly describe the populations with whom they have been tested, the risk and protective factors targeted for change, the core components of the intervention that are theorized to produce change, and their ability to be disseminated in communities. Thus, the criteria include attention to how well programs “will work” when replicated in communities, as advocated by Sampson et al. (2013) in this special issue. Although detailed analysis of mediating and moderating effects is not required by Blueprints, all available information regarding how and for whom interventions have been demonstrated effective is delineated on the website to ensure that replications mirror evaluation conditions (and associated outcomes) as much as possible.

Emphasizing the Importance of Implementation Integrity

In addition to the factors discussed by Sampson et al. (2013) in this issue, an important contributor to weak replication results is the failure of communities to implement prevention programs with fidelity (i.e., in adherence to the core components of the intervention). For a variety of reasons, practitioners may alter the content, method of delivery, or timing of interventions during service delivery, and these deviations can reduce the likelihood of achieving positive outcomes (Durlak and DuPre, 2008; Gottfredson and Gottfredson, 2002; Lipsey, 2009). One of the most extreme examples of this relationship was shown in a state-wide replication of the Functional Family Therapy program in Washington (Washington State Institute for Public Policy, 2002). In this effectiveness trial, youth whose families received services from FFT therapists who strongly adhered to program specifications had significantly lower rates of recidivism after the intervention compared with the control group, but those who received services from therapists who substantially deviated from FFT guidelines had *higher* rates of recidivism compared with the control group (Washington State Institute for Public Policy, 2002). As another example, an evaluation of the Nurse-Family Partnership program found that the intervention had weaker effects on families when delivered by paraprofessionals compared with registered nurses (Olds, 2002); yet paraprofessionals often are used by communities instead of RNs, particularly when resources are limited.

These findings suggest that policy makers must not only mandate the use of effective interventions but also specify that implementation delivery must match the requirements of these programs. A state-wide initiative in Pennsylvania demonstrates how this might work (Rhoades, Bumbarger, and Moore, 2012). In this case, the Pennsylvania Commission on Crime and Delinquency funded communities to replicate programs on the Blueprints list (including five family-focused interventions), but to continue receiving funding after

the start-up phase, agencies had to receive a certification from program developers that interventions were being delivered with high fidelity. Importantly, a state-level resource center also was established to support local agencies in their efforts to implement and monitor the quality of the new programs (Greenwood and Welsh, 2012; Rhoades et al., 2012). Problems are likely to arise when programs are implemented in the “real world,” and the provision of technical assistance can help ensure program integrity. Evaluations of the Blueprints Initiative have shown that when communities are provided with intensive and proactive technical assistance from program developers and/or scientific consultants, they can replicate science-based programs with integrity (Elliott and Mihalic, 2004; Fagan and Mihalic, 2003).

Increasing Support for Prevention

Federal and/or state mandates to increase the dissemination (and high-quality implementation) of effective delinquency prevention programs should help increase the uptake of these services. For example, the Safe and Drug Free Schools initiative and No Child Left Behind Act both required schools to use effective curricula to prevent student substance use, and use of effective programs increased after passage of such legislation (Ringwalt et al., 2011). Uptake has been slow, however, and many schools still implement programs with little evidence of effectiveness (Hallfors and Godette, 2002; Ringwalt et al., 2008; Ringwalt et al., 2011). Regarding family-focused services, only a small percentage of families who can benefit from effective family-focused interventions receive them, either because they are not offered such services or because they do not take advantage of opportunities that are made available (Kumpfer and Alvarado, 2003; Printz and Sanders, 2007). Both cases suggest a lack of support for family-focused interventions from both policy makers and the general public. Until greater enthusiasm for using science-based preventive interventions is engendered, widespread use of services and reductions in offending are unlikely to be realized.

Biglan and Taylor (2000) also noted a lack of support, as well as a lack of organized efforts by scientists to increase support, when contrasting the substantial reduction in cigarette smoking evidenced in the past few decades with the failure to impact rates of violent crime significantly. They attribute the former to collaborative efforts by researchers and government agencies to make the public aware of the harms of tobacco, the factors that lead to smoking, and the strategies that have been shown to reduce cigarette use, and the latter to the lack of consensus among scientists regarding the causes of violence (even though many risk and protective factors have been identified) and the lack of effort to market successful violence prevention programs (even though many such interventions exist). Furthermore, Biglan and Taylor (2000) pointed out that those charged with reducing violence have little scientific training in how to investigate why crime occurs, large budgets dedicated to policing and locking up offenders, and some pressure from the public to “get

tough on crime.” How, then, can support for science-based, preventive interventions be increased among policy makers and the public?

Blomberg et al. (2013) in their introduction emphasize the need for better communication between scientists and policy makers and increased efforts to build partnerships between these groups, such as the American Society of Criminology’s sponsorship of annual Congressional luncheons and the appointment of academic scholars to head public agencies such as the National Institute of Justice and the Bureau of Justice Statistics. Additional examples can be gleaned from the Society for Prevention Research, a multidisciplinary organization dedicated to improving health and well-being, which routinely includes representatives from multiple federal agencies (e.g., the National Institutes of Health, Center for Disease Control, and the U.S. Department of Health and Human Services) on its Board of Directors, as planners and participants in its annual conference, and as co-authors on position papers (e.g., Flay et al., 2005; Spoth et al., 2013). These interactions provide ample and meaningful opportunities for scientists and policy makers to discuss ways to work together on common areas of interest.

A key to effective partnerships is creating “win/win” situations, such that both parties perceive that proposed actions are necessary to achieve their goals. As it pertains to juvenile delinquency and family influences on offending, it is imperative that scientists communicate to policy makers and practitioners the benefits of investing in prevention. Research has established that most offenders initiate law-breaking during adolescence (Farrington, 2003; Sampson and Laub, 1993); that adolescent substance use and delinquency can result in many physical, social, and behavioral problems that extend into adulthood (Catalano et al., 2012); and that the costs of these outcomes are significant (Aos, Lieb, Mayfield, Miller, and Pennucci, 2004; Cohen, Piquero, and Jennings, 2010). On the other hand, lowering rates of juvenile offending will reduce adult crime, and investment in effective delinquency prevention programs can be cost beneficial (Aos et al., 2004). Research by the Washington State Institute of Public Policy indicated that although family-focused programs can be costly to implement, they can achieve “significantly more benefits than costs” (Aos et al., 2004). For example, for every dollar invested, the Nurse-Family Partnership can realize \$2.88 in return and the Multisystemic Therapy program \$2.64 in benefits.

Such information should increase support for prevention efforts from policy makers, but even more evidence may be needed to convince the general public, and parents in particular, that family services are worthwhile. Even well-funded scientific trials have challenges recruiting families to participate in programs, as parents may not perceive their family’s need for services (particularly if their children are not currently engaging in delinquency), may avoid programs for fear of being labeled a “bad parent,” or may lack the time or financial means to attend sessions (Printz and Sanders, 2007; Spoth and Redmond, 2000). Moreover, parents of teenagers may view strained relationships as inevitable, and they may view participation in a prevention program as counter to their goal of helping

adolescents seek independence (Dishion, Nelson, and Bullock, 2004; McGue, Elkins, Walden, and Iacono, 2005). Increasing public awareness that parents do matter, and that parents can actively help their children refrain from substance use and delinquency, even during adolescence, can likely help increase the uptake of family-focused services.

Another strategy to bolster support for preventive interventions is to identify the level of need for such services in a particular community. Communities have been shown to vary in levels of risk and protective factors (Hawkins, Van Horn, and Arthur, 2004), such that in some communities, youth may experience, on average, careful parental monitoring and strong emotional attachment to parents. In these areas, there will not be a strong need for parent training interventions. However, in other settings, there may be high levels of family risk and low levels of family protection. Information on such factors can be collected via population surveys, used to determine specific areas of need, and publicized to help garner support for services that will target these needs (Greenwood and Welsh, 2012).

To illustrate how such a process would work, it is useful to consider the Communities That Care (CTC) prevention system (Hawkins and Catalano, 1992; Hawkins, Catalano, and Arthur, 2002). CTC advocates for the collection of local epidemiologic data regarding the prevalence of risk and protective factors in a local area via anonymous, school-based, scientifically validated surveys of students (specifically, the Communities That Care Youth Survey; Arthur, Hawkins, Pollard, Catalano, and Baglioni, 2002; Glaser, Van Horn, Arthur, Hawkins, and Catalano, 2005). The survey should be completed by as many students as possible to obtain a community-wide “profile” of risk and protective factors most commonly faced by the majority of youth. Services are then directed at the most elevated risk factors and most depressed risk factors. CTC advocates a universal approach to prevention services. Instead of focusing on and possibly stigmatizing “at risk” youth and families, interventions are delivered broadly across the community. For example, local schools might implement violence prevention programs for all students in a particular grade(s), or local agencies would offer parent training programs to as many parents as possible in the community. Recognizing that multiple factors affect offending, communities are encouraged to implement an array of services that target as many risk and protective factors as possible, so long as interventions focus on changing the factors identified as most important in the local needs assessment (Fagan and Hawkins, 2012). Thus, some communities might implement family-focused interventions, but if they are not needed, programs targeting peer or school risk factors might be implemented. The active involvement of community members in the needs assessment process and the careful matching of effective prevention programs to local needs are what is most important, as these elements should increase the likelihood that preventive actions are broadly supported, well implemented, and sustained over time (Hawkins et al., 2002).

Conclusion

This policy essay has endeavored to present an optimistic view regarding the potential for criminologists to contribute to public policy, but also a credible and feasible approach that is based on scientific information about what contributes to the onset and prevention of youth delinquency. Although there is much more to be learned, criminologists have made significant progress in these areas, and it seems incumbent on us to share this information with those who are best positioned to effect change in society. To refrain from doing so risks even more youth becoming involved in crime, which is likely, in turn, to result in increased offending among adults, more imprisonment, and greater social and financial costs to society.

References

- Agnew, Robert A. 1992. Foundation for a general strain theory of crime and delinquency. *Criminology*, 30: 47–88.
- Agnew, Robert A. 2006. *Pressured into Crime: An Overview of General Strain Theory*. Cary, NC: Roxbury.
- Akers, Ronald L. 1985. *Deviant Behavior: A Social Learning Approach*, 3rd Edition. Belmont, CA: Wadsworth.
- Akers, Ronald L. 2009. *Social Learning and Social Structure: A General Theory of Crime and Deviance*. New Brunswick, NJ: Transaction.
- Andrews, Don A., Ivan Zinger, Robert D. Hoge, James A. Bonta, Paul Gendreau, and Francis T. Cullen. 1990. Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28: 369–404.
- Aos, Steve, Roxanne Lieb, Jim Mayfield, Marna Miller, and Annie Penucci. 2004. *Benefits and Costs of Prevention and Early Intervention Programs for Youth*. Olympia: Washington State Institute for Public Policy.
- Arthur, Michael W., J. David Hawkins, John A. Pollard, Richard F. Catalano, and A. J. Baglioni. 2002. Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review*, 26: 575–601.
- Bandura, Albert and Richard H. Walters. 1959. *Adolescent Aggression*. New York: Ronald Press.
- Baron, Reuben M. and David A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173–1182.
- Bernat, Debra H., Gerald J. August, Joel M. Hektner, and Michael L. Bloomquist. 2007. The Early Risers preventive intervention: Testing for six-year outcomes and mediational processes. *Journal of Abnormal Child Psychology*, 35: 605–617.
- Biglan, Anthony, Patricia J. Mrazek, Douglas Carnine, and Brian R. Flay. 2003. The integration of research and practice in the prevention of youth problem behaviors. *American Psychologist*, 58: 433–440.

- Biglan, Anthony and Ted K. Taylor. 2000. Why have we been more successful at reducing tobacco use than violent crime? *American Journal of Community Psychology*, 28: 269–302.
- Blomberg, Thomas G., Julie Mestre, and Karen Mann. 2013. Seeking causality in a world of contingencies: Criminology, research, and public policy. *Criminology & Public Policy*, 12: 571–584.
- Brezina, Timothy. 1998. Adolescent maltreatment and delinquency: The question of intervening processes. *Journal of Research in Crime and Delinquency*, 35: 71–99.
- Brody, Gene H., Steven R. H. Beach, R. A. Philibert, Yi-fu Chen, and Velma McBride Murry. 2009. Prevention effects moderate the association of 5-HTTLPR and youth risk behavior initiation: Gene X environment hypotheses tested via a randomized prevention design. *Child Development*, 80: 645–661.
- Brody, Gene H., Steven M. Kogan, Yi-fu Chen, and Velma McBride Murry. 2008. Long-term effects of the Strong African American Families Program on youths' conduct problems. *Journal of Adolescent Health*, 43: 474–481.
- Catalano, R. F., Abigail A. Fagan, Loretta E. Gavin, Mark T. Greenberg, Charles E. Irwin, Jr., David A. Ross, and Daniel T. L. Shek. 2012. Worldwide application of prevention science in adolescent health. *Lancet*, 379: 1653–1664.
- Catalano, R. F. and J. David Hawkins. 1996. The Social Development Model: A theory of antisocial behavior. In (J. David Hawkins, ed.), *Delinquency and Crime: Current Theories*. New York: Cambridge University Press.
- Chamberlain, Patricia, Joe Price, Leslie D. Leve, Heidemarie Laurent, John A. Landsverk, and John B. Reid. 2008. Prevention of behavior problems for children in foster care: Outcomes and mediation effects. *Prevention Science*, 9: 17–27.
- Cohen, Mark A., Alex R. Piquero, and Wesley G. Jennings. 2010. Studying the costs of crime across offender trajectories. *Criminology & Public Policy*, 9: 279–305.
- Conduct Problems Prevention Research Group. 2002. The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology*, 30: 1–17.
- Cullen, Francis T. and Cheryl L. Jonson. 2011. Rehabilitation and treatment programs. In (James Q. Wilson and Joan Petersilia, eds.), *Crime and Public Policy*. New York: Oxford University Press.
- DeGarmo, David and Marion S. Forgatch. 2005. Early development of delinquency within divorced families: Evaluating a randomized preventive intervention trial. *Developmental Science*, 8: 229–239.
- Derzon, James H. 2010. The correspondence of family features with problem, aggressive, criminal, and violent behavior: A meta-analysis. *Journal of Experimental Criminology*, 6: 263–292.
- Dishion, Thomas J., Sarah E. Nelson, and Bernadette M. Bullock. 2004. Premature adolescent autonomy: Parent disengagement and deviant peer process in the amplification of problem behaviour. *Journal of Adolescence*, 27: 515–530.

- Durlak, Joseph A. and Emily P. DuPre. 2008. Implementation matters: A review of the research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41: 327–350.
- Eckenrode, John, Mary Campa, Dennis W. Luckey, Charles R. Henderson Jr., Robert Cole, Harriet Kitzman, Elizabeth Anson, Kimberly Sidora-Arcoleo, Jane Powers, and David L. Olds. 2010. Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Archives of Pediatric & Adolescent Medicine*, 164: 9–15.
- Eddy, J. Mark and Patricia Chamberlain. 2000. Family management and deviant peer associations as mediators of the impact of treatment condition on youth antisocial behavior. *Journal of Consulting and Clinical Psychology*, 68: 857–863.
- Elliott, Delbert S. 2012. Crime prevention and intervention over the life course: Emerging trends and directions for future research. In (Chris L. Gibson and Marvin D. Krohn, eds.), *Handbook of Life-Course Criminology*. New York: Springer.
- Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton. 1985. *Explaining Delinquency and Drug Use*. Beverly Hills, CA: Sage.
- Elliott, Delbert S. and Sharon Mihalic. 2004. Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5: 47–53.
- Fagan, Abigail A. and Nicole Eisenberg. 2012. Latest developments in the prevention of crime and anti-social behaviour: An American perspective. *Journal of Children's Services*, 7: 64–72.
- Fagan, Abigail A. and J. David Hawkins. 2012. Preventing substance use, delinquency, violence, and other problem behaviors over the life-course using the Communities That Care System. In (Chris L. Gibson and Marvin D. Krohn, eds.), *Handbook of Life-Course Criminology*. New York: Springer-Verlag.
- Fagan, Abigail A. and Sharon Mihalic. 2003. Strategies for enhancing the adoption of school-based prevention programs: Lessons learned from the Blueprints for Violence Prevention replications of the Life Skills Training Program. *Journal of Community Psychology*, 31: 235–254.
- Fagan, Abigail A., M. Lee Van Horn, Thomas Jaki, and J. David Hawkins. In press. Differential effects of parental controls on adolescent substance use: For whom is the family most important? *Journal of Quantitative Criminology*.
- Farrington, David P. 2000. Explaining and preventing crime: The globalization of knowledge—The American Society of Criminology 1999 Presidential Address. *Criminology*, 38: 1–24.
- Farrington, David P. 2003. Developmental and life-course criminology: Key theoretical and empirical issues—The 2002 Sutherland Award address. *Criminology*, 41: 221–255.
- Farrington, David P. and Brandon C. Welsh. 2003. Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology*, 36: 127–151.
- Farrington, David F. and Brandon C. Welsh. 2005. Randomized experiments in criminology: What have we learned in the last two decades. *Journal of Experimental Criminology*, 1: 9–38.

- Flay, Brian R., Anthony Biglan, Robert F. Boruch, Felipe G. Castro, Denise C. Gottfredson, Sheppard G. Kellam, Eve K. Moscicki, Steven Schinke, Jeffrey C. Valentine, and Peter Ji. 2005. Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. *Prevention Science*, 6: 151–176.
- Forgatch, Marion S., Gerald R. Patterson, David S. DeGarmo, and Zintars G. Beldavs. 2009. Testing the Oregon delinquency model with 9-year follow-up of the Oregon Divorce Study. *Development and Psychopathology*, 21: 637–660.
- Gelles, Richard J. 1983. An exchange/social control theory. In (David Finkelhor, Richard J. Gelles, Gerald T. Hotaling, and Murray A. Straus, eds.), *The Dark Side of Families: Current Family Violence Research*. Beverly Hills, CA: Sage.
- Glaser, Renita R., M. Lee Van Horn, Michael W. Arthur, J. David Hawkins, and Richard F. Catalano. 2005. Measurement properties of the Communities That Care Youth Survey across demographic groups. *Journal of Quantitative Criminology*, 21: 73–102.
- Glasgow, Russell E., Edward Lichtenstein, and Alfred C. Marcus. 2003. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, 93: 1261–1267.
- Gottfredson, Denise C. and Gary D. Gottfredson. 2002. Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*, 39: 3–35.
- Gottfredson, Michael R. and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.
- Greenwood, Peter W. and Brandon C. Welsh. 2012. Promoting evidence-based practice in delinquency prevention at the state level: Principles, progress, and policy directions. *Criminology & Public Policy*, 11: 493–513.
- Gruner Gandhi, Allison, Erin Murphy-Graham, Anthony Petrosino, Sara Schwartz Chrismer, and Carol H. Weiss. 2007. The devil is in the details: Examining the evidence for “proven” school-based drug abuse prevention programs. *Evaluation Review*, 31: 43–74.
- Haggerty, K. P., Martie L. Skinner, Elizabeth P. MacKenzie, and Richard F. Catalano. 2007. A randomized trial of Parents Who Care: Effects on key outcomes at 24-month follow-up. *Prevention Science*, 8: 249–260.
- Hallfors, Denise and Dionne Godette. 2002. Will the “principles of effectiveness” improve prevention practice? Early findings from a diffusion study. *Health Education Research*, 17: 461–470.
- Harris, Judith R. 1995. Where is the child's environment? A group socialization theory of development. *Psychological Review*, 102: 458–489.
- Hassmiller Lich, Kristen, Elizabeth M. Ginexi, Nathaniel D. Osgood, and Patricia L. Mabry. 2013. A call to address complexity in prevention science research. *Prevention Science*, 14: 279–289.
- Hawkins, J. David and Richard F. Catalano. 1992. *Communities that Care: Action for Drug Abuse Prevention*. San Francisco, CA: Jossey-Bass.
- Hawkins, J. David, Richard F. Catalano, and Michael W. Arthur. 2002. Promoting science-based prevention in communities. *Addictive Behaviors*, 27: 951–976.

- 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: 64–105.
- Hawkins, J. David, M. Lee Van Horn, and Michael W. Arthur. 2004. Community variation in risk and protective factors and substance use outcomes. *Prevention Science*, 5: 213–220.
- Hay, Carter. 2003. Family strain, gender, and delinquency. *Sociological Perspectives*, 46: 107–135.
- Heckman, James J., Seong H. Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, 1: 1–46.
- Henggeler, Scott W. 2011. Efficacy studies to large-scale transport: The development and validation of multisystemic therapy programs. *Annual Review of Clinical Psychology*, 7: 351–381.
- Henggeler, Scott W., Gary B. Melton, Michael J. Brondino, David G. Scherer, and Jerome H. Hanley. 1997. Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65: 821–833.
- Herrenkohl, Todd I., Eugene Maguin, Karl G. Hill, J. David Hawkins, Robert D. Abbott, and Richard F. Catalano. 2000. Developmental risk factors for youth violence. *Journal of Adolescent Health*, 26: 176–186.
- Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley: University of California Press.
- Hiscock, Harriet, Jordana Bayer, Anna Price, Obioha C. Ukoumunne, Susan Rogers, and Melissa Wake. 2008. Universal parenting programme to prevent early childhood behavioural problems: Cluster randomised trial. *BMJ*, 336: 318–321.
- Hoeve, Machteld, Judith S. Dubas, Veroni I. Eichelsheim, Peter H. Van der Laan, W. Smeenk, and Jan R. M. Gerris. 2009. The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*, 37: 749–775.
- Hough, Mike. 2010. Gold standard or fool's gold? The pursuit of certainty in experimental criminology. *Criminology and Criminal Justice*, 10: 11–22.
- Howell, James C. 2003. *Preventing and Reducing Juvenile Delinquency: A Comprehensive Framework*. Thousand Oaks, CA: Sage.
- Klein, N. C., James F. Alexander, and Bruce V. Parsons. 1977. Impact of family systems intervention on recidivism and sibling delinquency: A model of primary prevention and program evaluation. *Journal of Consulting and Clinical Psychology*, 45: 469–474.
- Kotchick, Beth A. and Rex Forehand. 2002. Putting parenting in perspective: A discussion of the contextual factors that shape parenting practices. *Journal of Child and Family Studies*, 11: 255–269.
- Kumpfer, Karol L. and Rose Alvarado. 2003. Family-strengthening approaches for the prevention of youth problem behaviors. *American Psychologist*, 58: 457–465.

- Latessa, Edward J., Francis T. Cullen, and Paul Gendreau. 2002. Beyond correctional quackery—Professionalism and the possibility of effective treatment. *Federal Probation*, 66: 43–49.
- Leventhal, Tama and Jeanne Brooks-Gunn. 2000. The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126: 309–337.
- Lipsey, Mark W. 2009. The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*, 4: 124–147.
- Lipsey, Mark W. and James H. Derzon. 1998. Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In (Rolf Loeber and David P. Farrington, eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*. Thousand Oaks, CA: Sage.
- Lochman, John E. and Karen C. Wells. 2002. Contextual social-cognitive mediators and child outcome: A test of the theoretical model in the Coping Power Program. *Developmental Psychopathology*, 14: 945–967.
- MacKinnon, David P. 2008. *Introduction to Statistical Mediation Analysis*. Mahwah, NJ: Erlbaum.
- Martinson, Robert. 1974. What works? Questions and answers about prison reform. *The Public Interest*, 22–54.
- Mason, W. Alex, Rick Kosterman, Kevin P. Haggerty, J. David Hawkins, Cleve Redmond, Richard L. Spoth, and Chungyeol Shin. 2009. Gender moderation and social development mediation of the effect of a family-focused substance use preventive intervention on young adult alcohol abuse. *Addictive Behaviors*, 34: 599–605.
- McGue, Matt, Irene Elkins, Brent Walden, and William G. Iacono. 2005. Perceptions of the parent-adolescent relationship: A longitudinal investigation. *Developmental Psychology*, 41: 971–984.
- Mears, Daniel P. 2007. Towards rational and evidence-based crime policy. *Journal of Criminal Justice*, 35: 667–682.
- Mihalic, Sharon, Katherine Irwin, Delbert S. Elliott, Abigail A. Fagan, and Diane Hansen. 2001. *Blueprints for Violence Prevention*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Moffitt, Terrie E. 1993. Adolescence-limited and life-course persistent anti-social behavior: A developmental taxonomy. *Psychological Review*, 100: 674–701.
- O’Connell, Mary Ellen, Thomas Boat, and Kenneth E. Warner, eds. 2009. *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Washington, DC: National Academies Press.
- Olds, David L. 2002. Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science*, 3: 153–172.
- Olds, David L., Charles R. Henderson Jr., and Robert Cole. 1998. Long-term effects of nurse home visitation on children’s criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *Journal of the American Medical Association*, 280: 1238–1244.

- Pantin, Hilda, Guillermo Prado, Barbara Lopez, Shi Huang, Maria I. Tapia, Seth J. Schwartz, Eduardo Sabillon, C. Hendricks Brown, and J. Branchini. 2009. A randomized controlled trial of Familias Unidas for Hispanic adolescents with behavior problems. *Psychosomatic Medicine*, 71: 987–995.
- Patterson, Gerald R. 1982. *A Social Learning Approach. Volume 3: Coercive Family Practices*. Eugene, OR: Castalia.
- Petrosino, A., Carolyn Turpin-Petrosino, and John Buehler. 2003. “Scared Straight” and other juvenile awareness programs for preventing juvenile delinquency (updated C2 review). *The Campbell Collaboration Reviews of Intervention and Policy Evaluations (C2-RIPE)*. Philadelphia, PA: Campbell Collaboration.
- Piquero, Alex R., David P. Farrington, Brandon C. Welsh, Richard Tremblay, and Wesley G. Jennings. 2009. Effects of early family/parent training programs on antisocial behavior and delinquency. *Journal of Experimental Criminology*, 5: 83–120.
- Prado, G., Hilda Pantin, Ervin Briones, Seth J. Schwartz, Daniel Feaster, Shi Huang, Summer Sullivan, Maria I. Tapia, Eduardo Sabillon, Barbara Lopez, and Jose Szapocznik. 2007. A randomized controlled trial of a parent-centered intervention in preventing substance use and HIV risk behaviors in Hispanic adolescents. *Journal of Consulting and Clinical Psychology*, 75: 914–926.
- Pratt, Travis C. and Francis T. Cullen. 2000. The empirical status of Gottfredson and Hirschi’s General Theory of Crime: A meta analysis. *Criminology*, 38: 931–964.
- Pratt, Travis C. and Francis T. Cullen. 2005. Assessing macro-level predictors and theories of crime: A meta-analysis. *Crime and Justice*, 32: 373–450.
- Pratt, Travis C., Francis T. Cullen, Christine S. Sellers, L. Thomas Winfree Jr., Tamara D. Madensen, Leah E. Daigle, Noelle E. Fearn, and Jacinta M. Gau. 2010. The empirical status of social learning theory: A meta analysis. *Justice Quarterly*, 27: 765–802.
- Preacher, Kristopher J. and Ken Kelley. 2011. Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, 16: 93–115.
- Printz, Ronald J. and Matthew R. Sanders. 2007. Adopting a population-level approach to parenting and family support interventions. *Clinical Psychology Review*, 27: 739–749.
- Reynolds, Arthur J. and Suh-R. Ou. 2011. Paths of effects from preschool to adult well-being: A confirmatory analysis of the Child-Parent Center Program. *Child Development*, 82: 555–582.
- Reynolds, Arthur J., Suh-R. Ou, and James W. Topitzes. 2004. Paths of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. *Child Development*, 75: 1299–1328.
- Rhoades, B. J., Brian K. Bumbarger, and Julia E. Moore. 2012. The role of a state-level prevention support system in promoting high-quality implementation and sustainability of evidence-based programs. *American Journal of Community Psychology*, 50: 386–401.
- Ringwalt, Chris, Sean Hanley, Amy A. Vincus, Susan T. Ennett, Louise A. Rohrbach, and J. Michael Bowling. 2008. The prevalence of effective substance use prevention curricula in the nation’s high schools. *Journal of Primary Prevention*, 29: 479–488.

- Ringwalt, Chris, Amy A. Vincus, Sean Hanley, Susan T. Ennett, J. Michael Bowling, and Susan W. Haws. 2011. The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2008. *Prevention Science*, 12: 63–70.
- Rowe, David C. 1986. Genetic and environmental components of antisocial behavior: A study of 265 twin pairs. *Criminology*, 24: 513–534.
- Sampson, Robert J. 2010. Gold standard myths: Observations on the experimental turn in quantitative criminology. *Journal of Quantitative Criminology*, 26: 489–500.
- Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.
- Sampson, Robert J. and John H. Laub. 1994. Urban poverty and the family context of delinquency: A new look at structure and process in a classic study. *Child Development*, 65: 523–540.
- Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. Translating causal claims: Principles and strategies for policy-relevant criminology. *Criminology & Public Policy*, 12: 587–616.
- Sandler, Irwin N., Erin Schoenfelder, Sharlene A. Wolchik, and David P. MacKinnon. 2011. Long-term impact of prevention programs to promote effective parenting: Lasting effects but uncertain processes. *Annual Review of Psychology* 62: 299–329.
- Sherman, Lawrence W. 2003. Misleading evidence and evidence-led policy: Making social science more experimental. *The ANNALS of the American Academy of Political and Social Science*, 589: 6–19.
- Sherman, Lawrence W., Denise C. Gottfredson, Doris L. MacKenzie, Peter Reuter, and Shawn D. Bushway. 1998. *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: National Institute of Justice.
- Spoth, Richard L. and Cleve Redmond. 2000. Research on family engagement in preventive interventions: Toward improved use of scientific findings in primary prevention practice. *The Journal of Primary Prevention*, 21: 267–284.
- Spoth, Richard L. and Cleve Redmond. 2002. Project Family prevention trials based in community-university partnerships: Toward scaled-up preventive interventions. *Prevention Science*, 3: 203–222.
- Spoth, Richard L., Louise A. Rohrbach, Mark T. Greenberg, Phillip Leaf, C. Hendricks Brown, Abigail A. Fagan, Richard F. Catalano, Mary Ann Pentz, Zili Sloboda, J. David Hawkins, and the Society for Prevention Research Type 2 Translational Task Force. 2013. Addressing core challenges for the next generation of Type 2 translation research and systems: The translation science to population impact (TSci impact) framework. *Prevention Science*, 14: 319–351.
- Spoth, Richard L., Chungeyol Shin, Max Guyll, Cleve Redmond, and Kari Azevedo. 2006. Universality of effects: An examination of the comparability of long-term family intervention effects on substance use across risk-related subgroups. *Prevention Science*, 7: 209–224.
- Steinberg, Laurence. 2001. We know some things: Parent-adolescent relationships in retrospect and prospect. *Journal of Research on Adolescence*, 11: 1–19.

-
- Stern, Susan B. and Carolyn A. Smith. 1995. Family processes and delinquency in an ecological context. *Social Service Review*, 69: 703–731.
- Task Force on Community Prevention Services. 2005. The effectiveness of early childhood home visitation in preventing violence: A systematic review. *American Journal of Preventive Medicine*, 28: 11–39.
- Task Force on Community Prevention Services. 2007. Recommendation against policies facilitating the transfer of juveniles from juvenile to adult justice systems for the purpose of reducing violence. *American Journal of Preventive Medicine*, 32: S5–S6.
- Thornberry, Terence P. 1987. Toward an interactional theory of delinquency. *Criminology*, 25: 863–891.
- U.S. Department of Health and Human Services. 2001. Youth violence: A report of the Surgeon General. Rockville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; Substance Abuse and Mental Health Services Administration, Center for Mental Health Services; National Institutes of Health.
- Valentine, Jeffery C., Anthony Biglan, Robert F. Boruch, Felipe G. Castro, Linda M. Collins, Brian R. Flay, Sheppard G. Kellam, Eve K. Moscicki, and Steven P. Schinke. 2011. Replication in prevention science. *Prevention Science*, 12: 103–117.
- Washington State Institute for Public Policy. 2002. *Washington State's Implementation of Functional Family Therapy for Juvenile Offenders: Preliminary Findings*. Olympia: Author.
- Weisburd, David L. 2010. Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: Challenging folklore in evaluation research in crime and justice. *Journal of Experimental Criminology*, 6: 209–227.
- Welsh, Brandon C., Christopher J. Sullivan, and David L. Olds. 2010. When early crime prevention goes to scale: A new look at the evidence. *Prevention Science*, 11: 115–125.
- Wolchik, Sharlene A., Irwin N. Sandler, Roger E. Millsap, Brett A. Plummer, Shannon M. Greene, Edward R. Anderson, Spring R. Dawson-McClure, Kathleen Hipke, and Rachel A. Haine. 2002. Six-year follow-up of preventive interventions for children of divorce: A randomized controlled trial. *Journal of the American Medical Association*, 288: 1874–1881.
- Wright, John Paul and Kevin M. Beaver. 2005. Do parents matter in creating self-control in their children? A genetically informed test of Gottfredson and Hirschi's theory of low self-control. *Criminology*, 43: 1169–1202.
- Zerhouni, Elias. 2003. The NIH Roadmap. *Science*, 302: 63–72.
- Zhou, Q., Irwin N. Sandler, Roger E. Millsap, Sharlene A. Wolchik, and Spring R. Dawson-McClure. 2008. Mother-child relationship quality and effective discipline as mediators of the 6-year effects of the New Beginnings Program for children from divorced families. *Journal of Consulting and Clinical Psychology*, 76: 579–594.
-

Abigail A. Fagan is an associate professor in Sociology and Criminology & Law at the University of Florida. Dr. Fagan's research focuses on the etiology and prevention of juvenile

delinquency and drug use, with an emphasis on examining the ways in which scientific advances can be successfully translated into effective crime and delinquency prevention practices. Her research includes investigation of the effects of family processes (e.g., parenting practices and sibling relationships), victimization experiences, and community influences on juvenile offending.

Evidence and Public Policy

The Example of Evaluation Research in Policing

Daniel S. Nagin

Carnegie Mellon University

David Weisburd

George Mason University

Hebrew University

Research Summary

In this paper, we argue that both science and the policy process are well served by research with high evidentiary value. We also argue that experimental research is valuable in the policy domain not only because of its high evidentiary content but also because of its transparency. To exemplify this point, we describe how a decade-long research program on hot spots policing overturned the conventional wisdom that police could not affect crime and, in so doing, has profoundly altered police practice. Still, we recognize that for a variety of practical and ethical reasons, policy-relevant research on policing and criminal justice policy more generally cannot be based entirely on experiments. We discuss two key features of randomization—balance in expectation between the treated and the control group on all potential confounders and exogeneity of treatment status—that are the source of the high evidentiary status of randomized experiments. We go on to describe how these same two features can be credibly replicated in quasi-experimental studies that are not subject to the ethical and practical obstacles that may make experiments impractical in some circumstances.

Policy Implications

We urge criminologists to take greater advantage of quasi-experimental research opportunities. We also recommend that criminologists go beyond simply identifying such

Direct correspondence to Daniel S. Nagin, Professor of Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: dn03@andrew.cmu.edu).

opportunities. They should engage practitioners and policy makers and make clear what is required to allow for strong evidence of program effectiveness.

Keywords

evidence, public policy, experiments, quasi-experiments, exogenous, transparency

The Merriam Webster online dictionary defines evidence as “something that furnishes proof.” In the context of research findings, this definition is unequivocal—a finding either does or does not have standing as evidence depending on whether it “furnishes proof” or not. Another online dictionary, Dictionary.com, allows for a more equivocal status. It defines evidence as “that which **tends** to prove or disprove something” (emphasis added). The second definition comes closer to the use of the term “evidence” in scientific settings—the evidentiary content of findings varies across studies from negligible to very high.

The title of this special issue is “Criminology, Causality, and Public Policy.” As the title suggests, it is dedicated to the question of whether the threshold for a research finding to have evidentiary standing in a scientific setting should be different than in a public policy setting. Our answer to this question is an unequivocal “yes.” The criteria for judging the merits of public policies are different than for evaluating the veracity of scientific evidence. Therefore, the standard of proof that should be applied in making judgments about the evidentiary value of a research finding in a scientific setting may be different than in a policy setting. Sometimes the standard-of-proof threshold will be less in a policy setting, especially in circumstances where there is an urgent need for action. It is not hard to imagine circumstances where the threshold will be greater in the policy setting. For example, recent controversy about mandatory vaccination requirements makes clear the importance in some policy settings of evidence with minimal ambiguity, in this instance, that the vaccine is effective and the risk of side effects is extremely small. Thus, the question of required standard of proof for proffering a finding as evidence in a policy setting must be judged on a case-by-case basis.

In this policy essay, we do not take on the question of what criteria should be used in making such case-by-case judgments. Instead, we address the question of how the research enterprise can be better organized so that it can be more effective in influencing public policy. We do so from the lenses of research on policing although we believe our conclusions extend to other domains of criminological research. Our answer to this question begins with the recognition that the completion of scientifically rigorous research is generally very time consuming. As a consequence, completion of individual studies with high evidentiary value is often out of sync with the faster moving policy process. Notwithstanding, we will argue that cumulatively, a rigorously developed research base can profoundly affect policy. In

the next section, we make the case for this premise by recounting the history of hot spots policing experimentation and its impact on policing.

Much of the evidence on the effectiveness of hot spots policing is based on randomized experiments. In scientific settings, high evidentiary value is given to findings from randomized experiments because it is widely understood that, if properly conducted, randomized experiments provide convincing evidence of causality. This feature of experimental findings increases their evidentiary standing in policy settings as well; however, it is not the only reason for their influential status in policy settings. Another is their transparency. Findings from randomized experiments are far easier for policy makers and practitioners to understand than findings from more technically elaborate statistical analyses, for example, various forms of regression analysis of nonexperimental data. The transparency of experiments increases the comfort level of policy makers in acting on experiment-based findings.

Evidence in support of public policy, however, cannot be based exclusively on randomized experiments. This is recognized even by some of the strongest advocates for randomized experiments (Sherman et al., 1997; Welsh, Braga, and Bruinsma, 2013). For example, Welsh and Farrington's (2009) systematic reviews of the effectiveness of closed-circuit television and improved lighting on crime did not include a single randomized experiment. In Weisburd et al.'s (2006) systematic review of problem-oriented policing, they noted that simple pre–post observational evaluations outnumbered more rigorous controlled evaluations by a ratio of nearly 5 to 1. And of the ten evaluations that included control groups, only four were randomized experiments.

For some policy choices, randomized experiments are unethical. For example, ethics alone would rule out the feasibility of a randomized experiment to test the general deterrent effectiveness of capital punishment. Also, it is generally very difficult to carry out experiments that test the effect of policies that have societal-wide impacts. Returning again to the question of the general deterrent effect of the death penalty, even if it were ethical to conduct an experiment to test for a general deterrent effect, a randomized experiment would require randomly assigning the authority to use capital punishment or not to each of the 50 states and the District of Columbia. It does not require much imagination to anticipate the response of state legislators and governors, let alone the courts, to such an experiment.

Our argument that the gradual accumulation of strong evidence will ultimately be influential in the policy process is reminiscent of the tortoise and hare parable. It is our view that in the long run, the tortoise of strong science will outperform the hare of quick-and-dirty findings with low evidentiary value. Notwithstanding, we also believe that experimental research is not the only source of findings with high evidentiary value and that these alternative sources may better meet the timing of the policy process and still add to the cumulative base of strong evidence. Such research can come in many forms, but here we focus on forms of natural quasi-experimental research that plausibly capture the essential features of a randomized experiment.

We focus on this type of research for several reasons. One is to narrow the scope of this essay to a manageable size. However, more fundamentally, we believe that outside of research on policing, naturally occurring quasi-experimental designs have come to be underappreciated in criminology. They also share a feature of randomized experiments—transparency—which makes them attractive to policy makers. We think that criminologists can take advantage of opportunities for natural quasi-experimental research and, perhaps as well, can encourage practitioners and policy makers to implement policies in ways that maximize the causal strength of such evaluations. In this sense, we are not only interested in illustrating why certain types of naturally occurring quasi-experiments can be identified in observational data, but also in identifying ways in which researchers can influence the implementation of specific programs or policies to maximize their evidentiary standing.¹

This essay is organized as follows. The next section provides a brief history of hot spots policing research and its impact on public policy. The third section begins with a discussion of two key features of a randomized experiment that are the sources of the high evidentiary value of experimental findings. It then describes examples of quasi-experimental research that share these features and thereby, in our judgment, have high evidentiary value. Most examples of quality quasi-experimental research involve attentive researchers taking advantage of a natural experiment that they become aware of. In the fourth section, we describe ways in which criminologists can be proactive in taking advantage of, and even in creating, quasi-experimental research opportunities. In the final section, we provide our conclusions.

Hot Spots Policing as an Example of a Successful Experimental Program of Research

As America entered the decade of the 1990s, there was a general consensus among criminologists that traditional police practices did not work in preventing or controlling crime. As Michael Gottfredson and Travis Hirschi (1990: 270) wrote in their classic book *A General Theory of Crime*: “No evidence exists that augmentation of patrol forces or equipment, differential patrol strategies, or differential intensities of surveillance have an effect on crime rates.” David Bayley (1994) wrote even more strongly in the opening paragraphs of his classic book *Police for the Future*:

-
1. In recent years, criminological applications of propensity score matching, a method that has its intellectual roots in experimental design, have become increasingly common. Indeed one of the authors of this essay has contributed several of those applications, including its first, to a criminological topic—the effect of gang membership on violent delinquency (Haviland, Nagin, and Rosenbaum, 2007). We do not highlight propensity score matching in this essay because the method is vulnerable to bias stemming from treatment self-selection that is an important potential source of what Rosenbaum (2002) called hidden bias. In Haviland et al. (2007), they demonstrated how instrumental variables can be used to overcome this problem, but we are not aware of any applications of propensity score matching to the study of police effects on crime that have incorporated instrumental variables.

The police do not prevent crime. This is one of the best kept secrets of modern life. Experts know it, the police know it, but the public does not know it. Yet the police pretend that they are society's best defense against crime . . . this is a myth. First, repeated analysis has consistently failed to find any connection between the number of police officers and crime rates. Secondly, the primary strategies adopted by modern police have been shown to have little or no effect on crime. (Bayley, 1994: 3)

But 10 years later, a National Research Council (2004) committee that included experts on police practices and policies reached a very different conclusion about the potential crime-control effectiveness of police practices. The report concluded:

In contrast to these findings [regarding the standard model of American policing predominant in earlier decades], there is evidence of police effectiveness in each of the remaining cells of the table. In the cell that represents focused policing, we found promising evidence regarding the effects of arrest targeting specific types of people committing specific types of offenses. We found even stronger evidence regarding the use of traditional enforcement strategies that are targeted at specific places. Indeed, studies that focused police resources on crime hot spots provide the strongest collective evidence of police effectiveness that is now available. On the basis of a series of randomized experimental studies, we conclude that the practice described as hot-spots policing is effective in reducing crime and disorder and can achieve these reductions without significant displacement of crime control benefits. Indeed, the research evidence suggests that the diffusion of crime control benefits to areas surrounding treated hot spots is stronger than any displacement outcome. (National Research Council, 2004: 250)

What led to this dramatic change in the perception of the possible effectiveness of policing strategies? Why did the underlying assumption change so radically from the consensus view of 10 years earlier that police were ineffective in preventing crime? Although there are other promising programs in policing that are noted by the panel, the strongest evidence identified is found in the case of hot spots policing strategies where a series of experimental studies had been carried out. What is this experimental program of evidence that led to such a dramatic change in the perceptions of the effectiveness of police strategies?

Although there is a long history of efforts to focus police patrols (Gay, Schell, and Schack, 1977; Wilson, 1967), the emergence of what is often termed "hot spots policing" is generally traced to theoretical, empirical, and technological innovations in the 1980s and 1990s (Braga, 2001; Sherman and Weisburd, 1995; Weisburd and Braga, 2006). Simply defined, hot spots policing is the application of police interventions at very small geographic units of analysis. It does not sound like a very radical innovation, but it represents a major

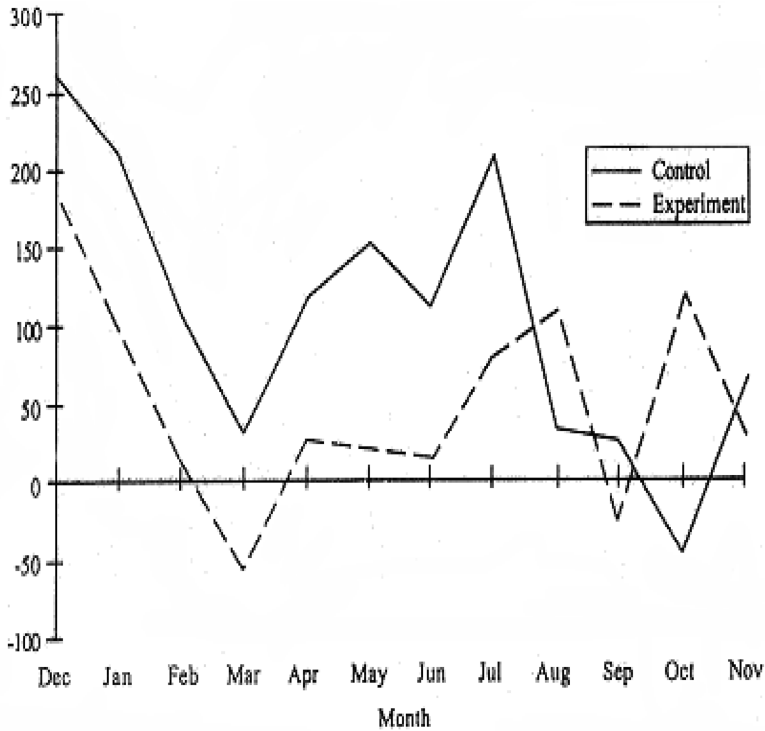
reform in how the police organize to do something about crime. Generalized and unfocused geographic policing strategies, such as random preventive patrol, were the norm in American policing, and these strategies had not fared well in research evaluations in the 1970s and 1980s (Weisburd and Eck, 2004). Indeed, such evaluations were the main source of the skepticism regarding the effectiveness of policing practices to reduce crime. In this context, hot spots policing was not developed by policy makers or practitioners, but rather by scientists who looked at basic research findings about the concentration of crime in cities. Therefore, the pace of the development of evidence about hot spots policing did not proceed at the rapid speed of social policies driven by political and social exigencies. Rather, the program was carried forward by researchers who were able to enlist practitioners willing to experiment with the new ideas.

The approach developed from a collaboration between Lawrence Sherman and David Weisburd that began in the late 1980s. Sherman had recently collected data on crime at addresses in Minneapolis, MN, for a study of problem-oriented policing. The data were startling in that they suggested a tremendous concentration of crime at specific street addresses (see Sherman, Gartin, and Buerger, 1989). This distribution of crime in the city suggested a much greater concentration of crime at specific places than criminologists or the police had assumed. The findings were also consistent with Weisburd's qualitative observations in a pilot community policing program in New York (Weisburd and McElroy, 1988) where police were assigned beats, but tended to focus on just a few problematic blocks. Sherman and Weisburd began to consider how these findings impacted the conclusions reached regarding the effectiveness of police patrol, and the ability of the police to do something about crime more generally. They believed that hot spots policing offered a potential answer to the failures of earlier studies:

The premise of organizing Patrol by beats is that crime could happen anywhere and that the entire beat must be patrolled. Computer-age data, however, have given new support to Henry Fielding's ([1751] 1977) proposal that police pay special attention to a small number of locations at high risk of crime. If only 3 percent of the addresses in a city produce more than half of all the requests for police response, if no police are dispatched to 40 percent of the addresses and intersections in a city over one year, and, if among the 60 percent with any requests the majority register only one request a year (Sherman, Gartin, and Buerger, 1989), then concentrating police in a few locations makes more sense than spreading them evenly through a beat. (Sherman and Weisburd, 1995: 629)

Sherman and Weisburd (1995) sought to design a study that would be persuasive to test their conjecture. They felt that only a randomized experiment could provide causal evidence persuasive enough and transparent enough for changing the widely established pessimism regarding effective policing.

FIGURE 1

The Minneapolis Hot Spots Experiment (Sherman and Weisburd, 1995)

The first experimental evaluation of hot spots policing took place in Minneapolis, MN (Sherman and Weisburd, 1995). Using computerized mapping of crime calls, 110 hot spots of roughly one street in length were identified and randomly assigned to treatment and control status. Police patrol was doubled on average for the experimental sites over a 10-month period. Figure 1 provides a visual summary. The vertical axis measures the monthly change in crime comparing crime calls for service for the pre-experimental year with the experimental year for the experimental and control groups. In each of the first 8 months of the experiment, the high-intensity-patrol hot spots had lower crime counts than the control hot spots. After month 8, treatment conditions were not maintained and, as is apparent from Figure 1, the treatment effect of high-intensity patrol was lost. Still, whether the authors examined only the period when dosage of treatment was maintained or longer time periods, the hot spots assigned to high-intensity-patrol treatment experienced a statistically significant drop in crime counts compared with the control hot spots. Figure 1 also illustrates another important virtue of randomized experiments—their findings are easily communicated.

In another randomized experiment, the Kansas City Crack House Raids Experiment (Sherman and Rogan, 1995a), crackdowns on drug locations also were found to lead to significant relative improvements in the experimental sites, although the effects (measured by citizen calls and offense reports) were modest and decayed in a short period. In yet another randomized trial, however, Eck and Wartell (1998) found that if the raids were immediately followed by police contacts with landlords, crime-prevention benefits could be reinforced and would be sustained for long periods of time. More general crime-and-disorder effects also are reported in three randomized experiments that take a more tailored problem-oriented approach to hot spots policing (Braga et al., 1999; Green Mazerolle, Roehl, and Kadleck, 1998; Weisburd and Green Mazerolle, 1995b). Nonexperimental studies provided similar findings (see Hope, 1994; Sherman and Rogan, 1995b).

If hot spots policing only displaced crime to other locations, its value as a policing strategy would be greatly diminished. Although measurement of crime displacement is complex and a matter of debate (see, e.g., Weisburd and Green Mazerolle, 1995a), in a number of the studies reported earlier, immediate geographic displacement was examined. In the Jersey City Drug Market Analysis Experiment (Weisburd and Green Mazerolle, 1995b), for example, displacement within two block areas around each hot spot was measured. No significant displacement of crime or disorder calls was found. More important, however, the investigators found that drug-related and public morals calls actually declined in the displacement areas in the experimental hot spots as compared with the control hot spots. This “diffusion of crime control benefits” (Clarke and Weisburd, 1994; Weisburd et al., 2006) also was reported in the New Jersey Violent Crime Places Experiment (Braga et al., 1999), the Beat Health study (Green Mazerolle et al., 1998), and the Kansas City Gun Project (Sherman and Rogan, 1995b). In each of these studies, no displacement of crime was reported, and some improvement in the surrounding areas was found. Only Hope (1994) reported direct displacement of crime in a quasi-experimental study. However, the observed displacement occurred only in the area immediate to the treated locations and the overall effect was much smaller than the crime prevention effect. Although the issue of displacement over the longer term or to more distant locations remains unresolved, we note that the evidentiary value of the finding of no displacement to adjacent locations in the shorter term is greatly enhanced by its being the product of a randomized experiment.

The experimental program that underlies hot spots policing also had two elements that make the evidence particularly persuasive. The first is that there was significant replication of the basic premise of the approach across a series of randomized experiments. Replication is the key to the accumulation of evidence of program impacts (Valentine et al., 2011). Less well appreciated is the extent to which the replication across different cities provided a persuasive argument that the effects of hot spots policing could be generalized to a wide array of contexts. This establishes strong external validity, which is not an inherent characteristic of randomized studies or of studies based on observational data, but again reinforces the evidentiary status of this experimental program.

We offer this brief account of the history and impact of the research program on hot spots policing to make several points. One is that the tension between conducting scientifically sound research and responding to the short-term evidentiary demands of the policy process is complex. The hot spots experimental program was conducted over a decade-long period and continues to develop (see Braga, Papachristos, and Hureau, 2012). Over this time frame, police executives could not hold back from making decisions on how to deploy their police forces, and elected officials responsible for funding police operations could not delay their funding decisions in anticipation of the completion of this research program. Still, there is little question that the scientific rigor of the hot spots policing research program was crucial to its overturning the consensus view that the police could have little effect on crime. This in turn had feedback into the domains of policy and practice. For example, in a 2013 address, the Acting Assistant Attorney General in charge of the Office of Justice Programs stated, “Our Smart Policing Initiative, administered by our Bureau of Justice Assistance, pairs law enforcement with researchers to design data-driven responses to neighborhood public safety problems, and it features several hot spots-related projects. . . . It’s [hot spots policing] a strategy we’ve supported through our programs” (Leary, 2013).

This observation brings us to another reason for our recounting the history of hot spots policing. It is our view that the tortoise-like accumulation of findings with high evidentiary value will ultimately be more successful in producing sound public policy than the hare-like provision of findings with low evidentiary value that is intended to be responsive to the short-term demands of the policy process.² Notwithstanding, it is neither practical nor ethical to produce all evidence required to make policy decisions with randomized experiments. We turn next to alternative study designs that have promise to produce research findings with high evidentiary value and strong transparency for policy makers and practitioners.

Examples of Quasi-Experimental Studies That Provide Convincing Causal Evidence

Some of the most convincing nonexperimental evidence on the effect of police presence on crime comes from accounts of the impact of an event unrelated to the crime rate that results in the complete withdrawal of police presence. For example, in September 1944, German soldiers occupying Denmark arrested the entire Danish police force. According to an account by Andenaes (1974), crime rates rose immediately, but not uniformly. The frequency of street crimes such as robbery, whose control depends heavily on visible police presence, rose sharply. By contrast, crimes such as fraud were less affected. Sherman and Eck

2. Robert Boruch commented that there are circumstances in which a single experiment played a hare-like role in expediting public policy. We agree—all analogies have exceptions. An example is the large-scale clinical trials testing the Salk vaccine, which undoubtedly accelerated its universal availability as an approved polio vaccine. Similarly, had it been feasible to do a randomized experiment demonstrating the ill effects of smoking on health, that also would have likely accelerated policy to restrict or discourage smoking.

(2002) summarized studies showing similar impacts of the withdrawal of police presence as a result of strikes.

These studies, however, are not informative about the policy-relevant question: How do incremental changes in police presence, whether by changes in police numbers or policy, affect crime rates? In this section, we describe four quasi-experimental studies that, in our judgment, provide solid evidence in answer to this question. The studies analyze different possible mechanisms by which police may affect crime. One is by legally changing police enforcement powers. This mechanism is illustrated with a classic study by Ross (1973) of the British Road Safety Act, which authorized the use of breathalyzer tests to establish whether a driver exceeded a legally specified blood-alcohol level defined in the Act. Another possible mechanism is by increasing police numbers. This mechanism is illustrated with a study by Klick and Taborrok (2005) of the crime-prevention effect of large changes in police numbers on the Washington, D.C. National Mall prompted by terror alerts. The third mechanism involves changes in the way a set number of police is used to prevent crime. This mechanism is illustrated with a study by Heaton (2010) of the effect of anti-police profiling policies in New Jersey on auto theft rates. The fourth mechanism involves technologies that improve police effectiveness in making apprehensions. This mechanism is illustrated with a study by Ayres and Levitt (1998) of the anti-auto theft Lojack technology.

The studies were also selected because the phenomenon they examined would be difficult to investigate with a randomized experiment because of either the nature of the “treatment” under study or the scale of the treatment response that is measured. Like capital punishment, practical and legal obstacles to the randomization of the authority to use breathalyzer tests across geographic areas makes a randomized experiment in this area difficult to carry out within current legal norms. A randomized experiment of changes in police numbers across geographic areas is certainly feasible as demonstrated by the hot spots experiments, but the National Mall is a far larger geographic area than a hot spot. The cost of doing an experiment involving large changes in police numbers randomized across geographic areas as large as the Mall often make such studies prohibitively expensive to conduct.³ The studies of Heaton (2010) and Ayres and Levitt (1998) examined societal-wide general deterrent effects, which would be extremely difficult and costly to test for with a randomized experiment.⁴

-
3. However, we know of one team of researchers in South America that is working to allocate at random entire police jurisdictions to experimental and control hot spots patrol regimes. Nonetheless, such studies are likely to be extremely rare because of the scale of geography involved.
 4. Another shared characteristic of these studies is that they are all interrupted time-series type studies in which a discrete event creates a “before” and an “after.” So-called “regression discontinuity” designs also exploit abrupt changes by, for example, comparing people who receive different sentences in adjacent blocks of a sentencing grid. Campbell (1969) argued that such designs could provide comparably convincing evidence of causal effects as interrupted time-series studies. We know of no policing studies based on a regression discontinuity design, although in principle the design is applicable.

The studies also were selected because, in our judgment, their findings have high evidentiary content and are transparent to informed laypeople. In turn, they share some important characteristics of experiments regarding the confidence we can have in the causal conclusions that are drawn. Randomization of treatment in an experiment provides two interrelated, but distinct, benefits that form the foundation for why experiments identify causal effects. Randomization ensures balance in the joint distribution of all measured and unmeasured characteristics between treatment and control units. This, in turn, ensures that in *expectation*, there is no difference in the mean of such characteristics between treatment and control. Just as important, it also ensures that all other features of the joint distribution including, for example, the variance and correlation of the universe of characteristics of the experimental population are the same for the treated and control groups. This guarantees that there is no confounding of the treatment effect by some set of factors that is correlated with both treatment status and the outcome of interest.

Treatment self-selection, whereby people self-select treatment based on anticipated benefit, is a leading reason such confounding can occur in nonexperimental data—self-selection generates the very imbalances in treated and control characteristics that a randomized experiment avoids. An example of such self-selection is people choosing whether or not to take depression medication based on their symptoms of depression. In the context of police research, self-selection occurs when police resources or enforcement strategies are altered in response to crime (e.g., assigning more patrol officers to a neighborhood with a crime spike). Self-selection is avoided in an experiment because subjects do not self-select into treatment; they are randomized into treatment.⁵ As a consequence, causal ordering is clear-cut. We will hereafter use the terms “endogenous” and “exogenous” to describe, respectively, circumstances where self-selection is and is not present.

As will be described in the next section, the authors of each of these exemplar studies made a credible case that the treatment under study is exogenous—for example, in the Klick and Tabarrok (2005) study, the change in the number of police officers assigned to the National Mall was not in response to a spike in crime rates on the Mall but to a terror alert. These studies also made a credible case that the measured change in crime was a result of the treatment examined and not of some other factor that coincided with the treatment under investigation.

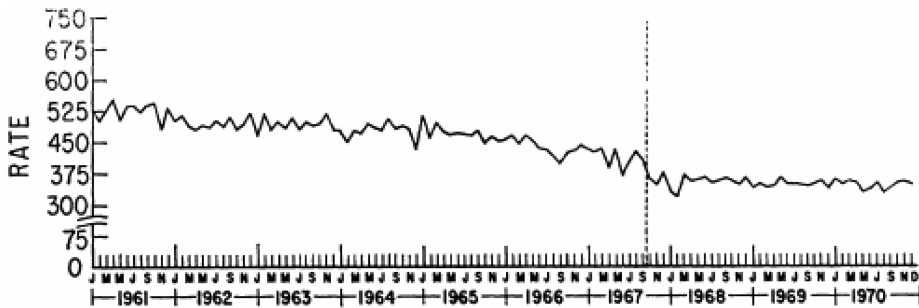
“Law, Science, and Accidents: The British Road Safety Act of 1967” (Ross, 1973)

The British Road Safety Act of 1967 introduced several important expansions in policing powers related to drunk driving. First, it set a blood alcohol level standard for defining alcohol impairment. This replaced subjective judgment by the arresting officer that was

5. In a randomized experiment, all experimental units may be high-crime locations, as in hot spots experiments, but because treatment status is randomized, there will be no difference in expectation of the baseline crime rate between the treated and control locations.

FIGURE 2

Figure 6 of Ross (1973)



Notes. Total casualties per 100 million vehicle miles in Great Britain, corrected for length of months, with seasonal variations removed.

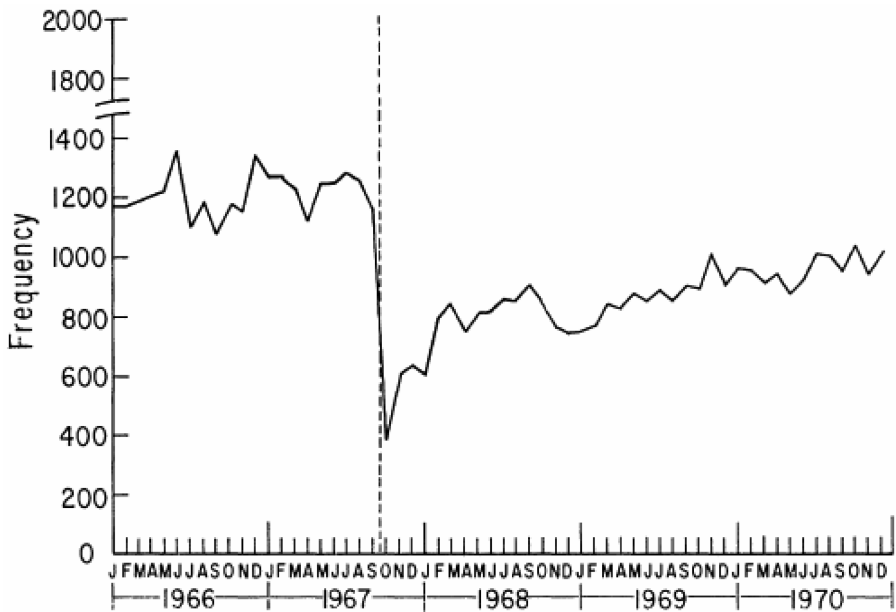
Source. Reproduced with permission from *Journal of Legal Studies* © 1973.

often difficult to sustain in court. It also provided for the use of breathalyzer testing to ascertain whether the standard was exceeded. The bill was first introduced to Parliament in January 1966. It was enacted in May 1967 to become effective on October 1, 1967.

Like randomized experiments, the quasi-experimental evaluation of this program by Ross (1973) was statistically straightforward and transparent. But at the same time, Ross was able to analyze the data in ways that provided convincing evidence that the effects observed were a result of the intervention of the new legislation (and subsequent police activity) rather than of other spurious causes. In turn, Ross established convincingly that the treatment preceded the outcome and that treatment was not a result of the outcomes examined. Figure 2 reproduces Figure 6 in Ross (1973), which reports casualties per 100 million vehicle miles corrected for length of the month and seasonality from 1961 to 1970. Concerning the issue of the exogeneity of the legislated changes, Ross reported graphs of total fatalities and fatalities per vehicle mile with and without seasonal and other adjustments that show no evidence of endogeneity—a spike in traffic deaths during the period 1961 to October 1967. Most important in this regard, there is no evidence of a spike immediately prior to October 1967. This rules out regression to the mean as an explanation for a statistically significant 46.3 casualties per 100 million mile (11.4%) drop in the rate between September and October 1967. Measured on a year-over-year basis, the number of fatalities in October was 12% lower than a year prior. November and December casualties were down by 13% and 21%, respectively, over the year prior. However, as Ross noted, there may have been some other event coincidental to the effective date that may explain the decline. For example, bad weather increases minor accidents but is associated with lower

FIGURE 3

Figure 10 of Ross (1973)



Notes. Fatalities and serious injuries combined for Friday nights, 10 P.M. to midnight; Saturday mornings, midnight to 4 A.M.; Saturday nights, 10 P.M. to midnight; and Sunday mornings, midnight to 4 A.M.; corrected for weekend days per month, seasonal variations removed.

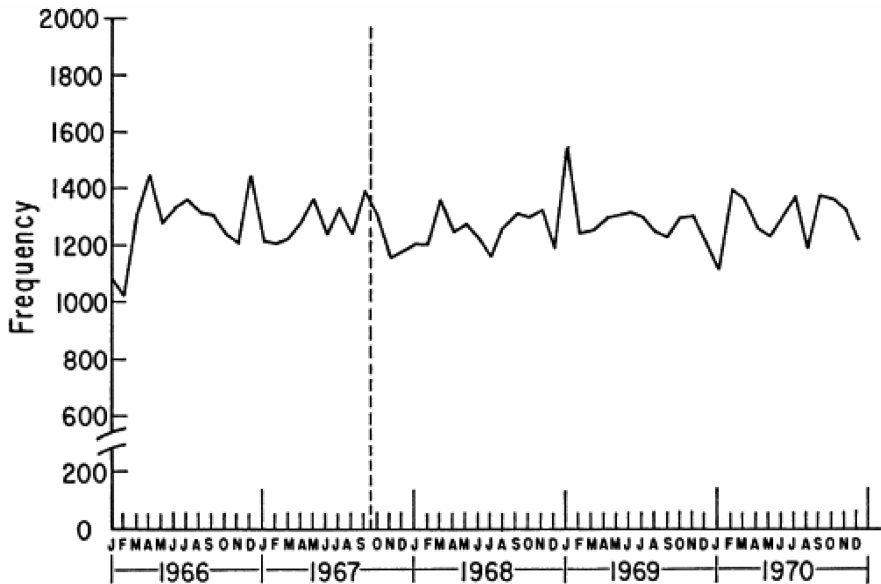
Source. Reproduced with permission from *Journal of Legal Studies* © 1973.

fatalities because people drive more cautiously. Ross reported that even by English standards, the weather in December 1967 was miserable compared with the year prior.

To respond to the possibility that a coincidental event explains the declines, Ross (1973) studied fatalities at night on weekends, when drunk driving is anticipated to be most common, and weekdays during the morning and evening commuting hours when drunk driving is anticipated to be least common. Figures 3 and 4, respectively, reproduce Figures 10 and 11 from Ross (1973). Figure 3, which reproduces the weekend fatality frequency graph, shows a sharp decline right after enactment. In the ensuing 3 years, fatalities trend upward but remain below pre-Act levels. By contrast, Figure 4, which reproduces the graph of fatalities during weekday commuting hours, shows no apparent discontinuity at the effective date. In our judgment, these two graphs constitute convincing evidence that the Act reduced fatalities substantially at least in the short term.

FIGURE 4

Figure 11 of Ross (1973)



Notes. Fatalities and serious injuries for Mondays through Fridays, 7 A.M. to 10 A.M. and 4 P.M. to 5 P.M., corrected for weekdays per month, seasonal variations removed.

Source. Reproduced with permission from *Journal of Legal Studies* © 1973.

“Using Terror Alert Levels to Estimate the Effect of Police on Crime” (Klick and Tabarrok, 2005)

The ongoing threat of terrorism has provided a number of unique opportunities to study the impact of police resource allocation on crime in cities around the world, including the District of Columbia (Klick and Tabarrok, 2005), Buenos Aires (Di Tella and Schargrotsky, 2004), Stockholm (Poutvaara and Priks, 2006), and London (Draca, Machin, and Witt, 2008). The Klick and Tabarrok study examined the effect on crime in Washington, D.C., of the color-coded alert system implemented in the aftermath of the September 11, 2001 terrorist attacks. The purpose of the alerts is to signal federal, state, and local law enforcement agencies to occasions when it might be prudent to divert resources to sensitive locations, such as the District’s National Mall. Klick and Tabarrok used daily police reports of crime for the period March 2002 to July 2003, during which time the terrorism alert level rose from “elevated” (yellow) to “high” (orange) and back down to “elevated” on four occasions. Again, the Klick and Tabarrok analyses are easy to comprehend, but nonetheless provide credible evidence of the causal influence of the increased police patrols.

Klick and Tabarrok (2005) used the changes in police numbers on patrol induced by the alert as the basis for inferring the effect of police numbers on crimes rates. The simplicity of the design is appealing because the changes in police numbers were induced by concerns about a threat of terrorist attack not by concerns about the type of crime (e.g., thefts) under study. Thus, there is good reason to be confident that the changes in police numbers were exogenous.

One important limitation of the study was that Klick and Tabarrok (2005) did not have direct measurements of changes in police numbers. They only knew that the District responded to the alerts by increasing the number of patrols, extending shift times, and activating closed-circuit camera systems that are present in sensitive areas such as the Mall. Their regression analyses of daily level crime data found that District-wide, the crime rate drops by approximately 7% during alerts (see Figure 5). They went on to make a good case that the decline was not a product of a reduction in tourist traffic induced by the alert.

Klick and Tabarrok (2005) also did analyses specifically aimed at assessing the degree to which the Mall area played a prominent role in the decline as a result of its receiving special police attention because it is a prime target for a terror attack. They indeed found this to be case. The crime decline in the Mall area was twice that in the District as a whole, 15%. Klick and Tabarrok also found that auto theft and theft from autos were particularly responsive to the changes in police presence. These types of crime were reduced by 43%. This makes sense because of the visibility of autos to police patrols.

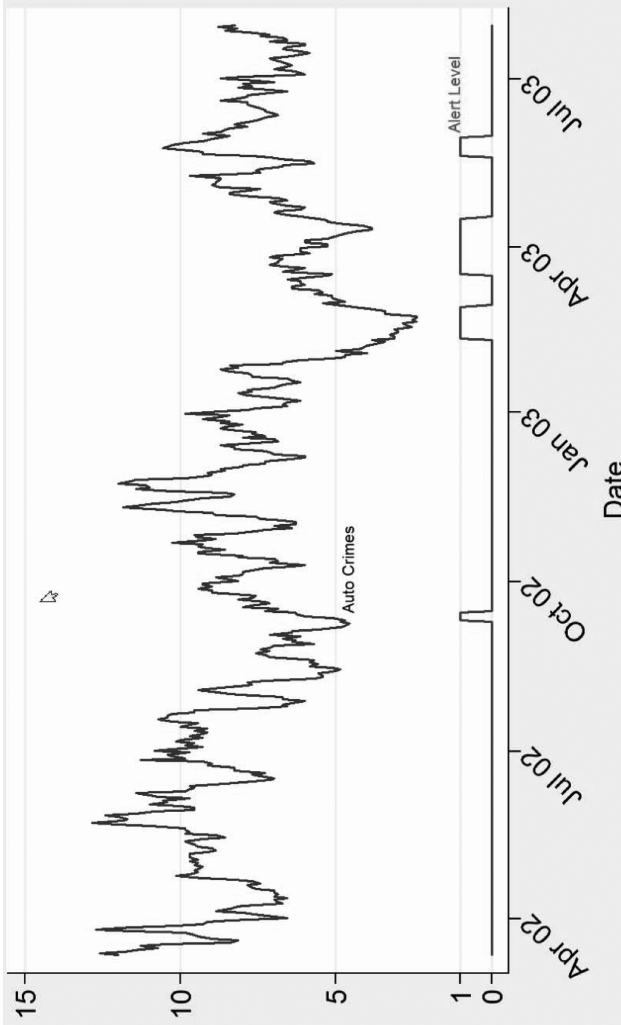
“Understanding the Effects of Antiprofiling Policies” (Heaton, 2010)

Racial profiling by police has been a “hot button” civil rights and crime-control issue for decades. Heaton (2010) examined the impact of a series of profiling-related incidents in New Jersey that occurred in 1998–1999 that dramatically altered police stop-and-interrogate interactions with minorities, particularly those involving auto-related stops by police. The changes were prompted by an incident in April 1998 in which White police officers shot four minority men who they claimed were behaving in a threatening manner, but who in fact were unarmed. This incident prompted a state attorney general report, released in early 1999, that concluded that profiling was a significant problem. The backlash against profiling was further exacerbated by incendiary remarks by the State Police Commissioner about disproportionate minority involvement in drug dealing.

Heaton (2010) convincingly demonstrated that this series of events spanning a relatively short period of time resulted in a sharp reduction of police enforcement activity against Blacks particularly as it related to motor vehicle theft. Specifically, a series of difference-in-difference type regressions that involved differencing of up to four factors make a strong case that the auto-theft-related arrest of Blacks compared with Whites declined sharply after the incidents described. Figures 6 and 7, respectively, reproduce Heaton’s Figures 2 and 3.

FIGURE 5

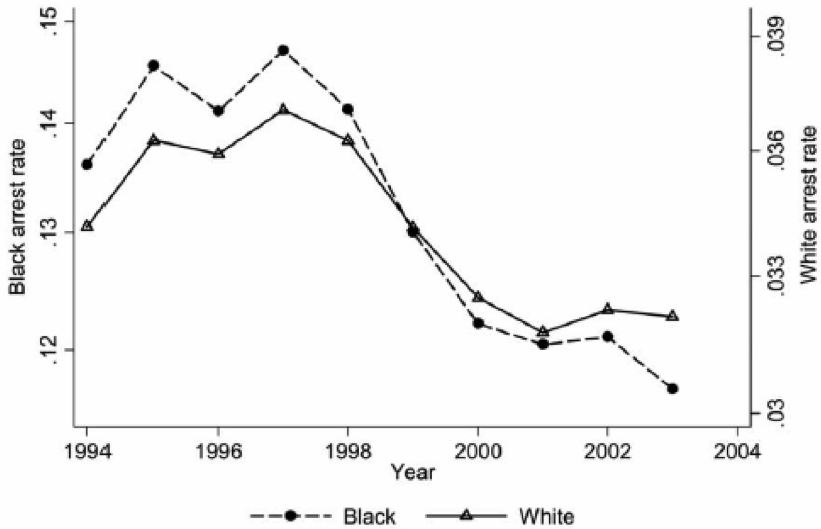
Auto Theft in the National Mall Area and Terror Alert Level



Notes. Crime is a one-week moving average (3,1,3) of stolen automobiles and theft from autos. Alert is 1 during high alert periods. Data is from National Mall region of Washington, DC.
Source. Data courtesy of J. Klick and A. Tabarrok based on Klick and Tabarrok (2005).

FIGURE 6

Figure 2 of Heaton (2010)



Notes. Annual arrest rates in New Jersey, all crimes

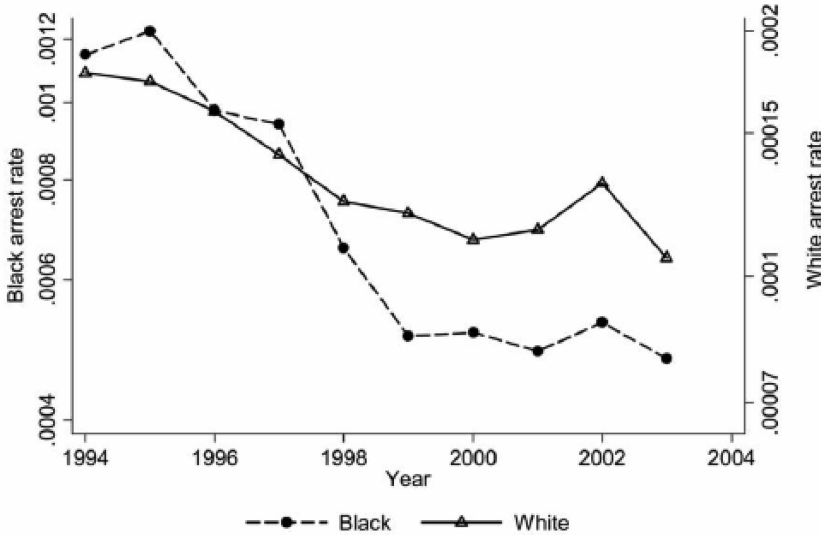
Source. Reproduced with permission from *Journal of Law & Economics* © 2010.

The figures demonstrate in a transparent manner the basis for the regression results. It is evident from these graphs that although the policy changes induced by the scandal had no material effect on overall Black to White arrest rates (Figure 6), they did materially affect arrests rates for auto theft (Figure 7). After 1998, arrests for auto theft of Blacks relative to Whites dropped sharply. Heaton also showed that in nearby Connecticut, no such change occurred in the relative Black to White arrest rates for auto theft.

Heaton (2010) then went on to examine the impact that the changed enforcement activity had on auto theft rates. He first presented evidence showing that the steady decline in auto thefts beginning in 1994 was halted beginning in 1998. By comparison, there was no such change in the trend for burglaries and larcenies. As a further test of the effect on auto theft, Heaton reasoned that the change in auto-related enforcement activity would disproportionately affect auto theft rates in higher minority communities because of the reluctance of the police to make stops against minority drivers. This is, in fact, what he found. The community-based analysis implies, for example, that the changed policing tactics prompted by the scandal resulted in an 18% greater increase in the auto theft rate in a community at the 75th percentile of Black population share compared with a community at the 25th percentile of the Black population.

FIGURE 7

Figure 3 of Heaton (2010)



Notes. Annual arrest rates in New Jersey, motor vehicle theft

Source. Reproduced with permission from *Journal of Law & Economics* © 2010.

“Measuring Positive Externalities from Unobservable Victim Precaution: An Empirical Analysis of Lojack” (Ayes and Levitt, 1998)

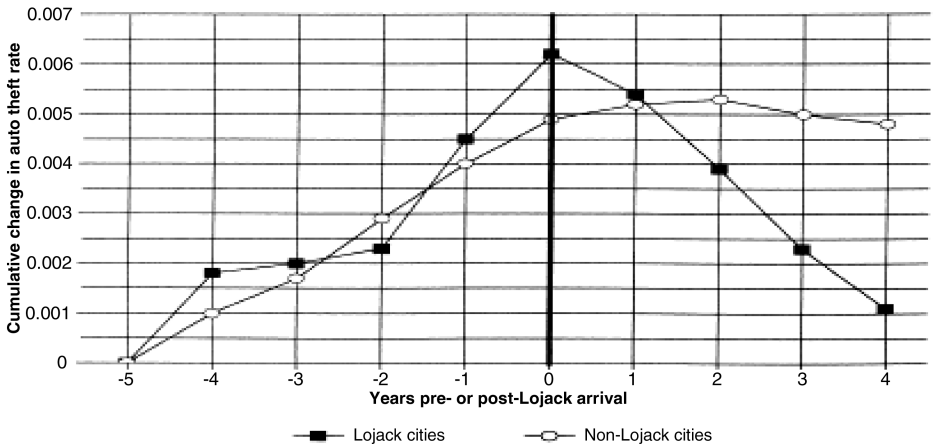
Lojack is an example of a technology that improves the apprehension effectiveness of the police. It is a hidden radio-transmitting device that enables police to locate stolen cars. It thereby increases the likelihood of apprehension of the perpetrator(s) of the theft. Also, because Lojack is hidden, it may have spillover benefits in deterring the theft of vehicles that are not outfitted with the technology. Because would-be auto thieves do not know which vehicles are outfitted with the technology, it may curtail their theft of autos writ large. Ayres and Levitt (1998) tested for this deterrent effect.

The study is based on an analysis of 57 U.S. cities with a population greater than 250,000. Of these cities, 13 had Lojack at some time in the period of the analysis from 1981 to 1994.⁶ Ayres and Levitt (1998) began by making a very credible argument that the introduction of Lojack into these 13 cities was exogenous—namely, that it was not affected by an unusual spike in the auto theft rate at the time of the introduction. The introduction of Lojack required regulatory approval, and data are provided showing that the length

6. Note, however, that identification of the deterrent effect depends primarily on six cities that introduced Lojack prior to 1992.

FIGURE 8

Figure 2 of Ayres and Levitt (1998)



Source. Reproduced with permission from *The Quarterly Journal of Economics* © 1998.

of time to approval and introduction was highly variable—from months to years—which had the effect of making the date of introduction exogenous. They reported a variety of regression analyses indicating that the introduction of Lojack materially reduced auto theft rates. Figure 8, which reproduces Figure 2 from Ayres and Levitt, demonstrates why the regressions measured crime-reduction effects. It pertains to the six cities that introduced Lojack prior to 1992 and measures the average cumulative percentage change in the auto theft rate from 5 years prior to the introduction to up to 5 years after introduction. Also reported for comparison is the composite average of changes in the auto theft rate for non-Lojack cities. The growth in the auto theft rate in Lojack and non-Lojack cities is comparable prior to introduction, but after introduction, there is a pronounced decline in the Lojack cities that is not matched in the non-Lojack cities.

Concluding Thoughts on Quasi-Experiments

In this section, we have summarized four quasi-experimental studies that, in our judgment, provide credible causal evidence on the phenomenon studied, whether it be a change in law, change in police enforcement tactics or numbers, or the introduction of technology that increases the effectiveness of the police. As discussed at the beginning of this policy essay, in each case, it would have been difficult or impossible to have tested for the effects analyzed with a randomized experiment. Like experiments, however, the design and analysis of these studies are straightforward and transparent to policy makers and practitioners.

We also discussed earlier why we judged the findings of these studies to provide good evidence of causal effects. One reason was that it was credible to assume that the “treatment”

under study was exogenous just like in an experiment where treatment status is randomly assigned.⁷ A second was that the studies effectively made the case that there was no other credible source of confounding beyond the failure of exogeneity.

One of the main reasons that confounding can be ruled out is that the response to treatment was immediate. All of the examples described earlier are forms of before-and-after-type analysis.⁸ An immediate “after” response helps to rule out other possible confounders because their timing must closely correspond to the event under study. The longer the time gap between the intervention and the response, the greater the chance that some other events may alter the path of the types of crime under study. This observation points to an important weakness of analyses of nonexperimental data compared with experimental data. In the case of an experiment, any difference in outcome between treated and control must somehow be attributable to treatment status regardless of time since treatment.⁹

We think it is important to note still another strength of these types of quasi-experiments. Because they often were carried out across large administrative areas, the findings can be broadly generalized. Evaluation studies may lack external validity for a variety of reasons. One is that they were applied to the specific contexts and circumstances in which the studies were carried out. As we have already noted, one strength of the cumulative evidence on hot spots policing is that it suggests that this is an effective policing strategy in a wide variety of circumstances. There have now been 21 tests of hot spots policing in cities of varying sizes, across different types of problems, and using different types of interventions (Braga et al., 2012). Another reason that an evaluation study may lack external validity is that it was not carried out at a high enough scale of aggregation to capture community- or societal-level effects such as general deterrence. A strength of several of the natural quasi-experiments that we have described is that they measured effects across large administrative jurisdictions (e.g., see Ayres and Levitt, 1998; Heaton, 2010; Ross, 1973).

Making the Scene: Opportunities for Criminologists

At the outset, we noted that we were not simply interested in identifying strong quasi-experiments in natural settings, but also in suggesting ways that criminologists could be more proactive in identifying opportunities for research and in encouraging policy makers to implement programs and policies in ways that create strong research opportunities. In this sense, we think that criminologists should “make the scene” of quasi-experiments in natural settings.

7. Note, however, that exogeneity is guaranteed in an experiment, not assumed.

8. Note that the Klick and Taborrok (2005) and Ayres and Levitt (1998) studies analyzed more than one event—four changes in the terror alert level for the former study and the introduction of Lojack across multiple cities in the later study. This is an additional strength of both studies.

9. We note, however, that experimental data have no special status for inferring the mechanism underlying any measured difference.

What this means is that criminologists must become involved in the policy or program process early on.¹⁰ On the one hand, such involvement means simply that criminologists will follow the policy and program processes as they begin to evolve, so that they can identify strong opportunities for evaluation. The study by Klick and Tabarrok (2005) reviewed earlier relied on the fact that police practices were implemented in a specific place (the National Mall) in reaction to a specific problem (a terrorism alert). This provided two of the criteria for a strong evaluation that we noted earlier: immediacy because of the specific and immediate response to terrorism and exogeneity because of the fact that the focus was terrorism (but crime also could be expected to be influenced). Similarly, Ayres and Levitt (1998) took advantage of the fact that Lojack was implemented in only some jurisdictions, and that the implementation was unrelated to the levels of car theft in those areas. We suspect that there are many such opportunities for strong natural quasi-experiments in the everyday world of policy and practice. Examples include Berk and MacDonald (2010) and Cohen and Ludwig (2003).

It is perhaps no accident that three of the four studies we have identified were carried out by economists. Throughout this policy essay, we have emphasized the exogeneity of the intervention. This is an issue that economists are historically well aware of but that has not been a widely discussed question in criminology. As early as the 1950s, economists have been concerned with the problem of identification, for example, in modeling supply-and-demand relationships. Again, in the early 1980s, this issue was made central by James Heckman's seminal work on selection bias, for which he was awarded the Nobel Award in Economics. More generally, the work of leading econometricians, such as Angrist and colleagues (Angrist, Imbens, and Rubin, 1996; Angrist and Pischke, 2009) and Manski (1995), are reflective of the attention given to questions of identification in economics. This has led them to think carefully and critically about the proper interpretation of causal effects estimates based on so-called "natural experiments" (see Rosenzweig and Wolpin [2000] for an insightful discussion of these issues). For our purposes here, one of the most important implications of this work from econometrics is that except in very limited circumstances, the identification of causal relationships cannot be resolved by simply adding more variables to a traditional regression equation. This led economists to seek other methods for establishing exogeneity, which have included the types of natural quasi-experiments we have focused on here.

But we also recommend that criminologists go beyond simply identifying such opportunities. They also should engage practitioners and policy makers and make clear what is required to allow for strong evidence of program effectiveness. For example, we must make it clear that there is much to be gained by programs that are implemented in a specific

10. For a useful discussion of how researchers can successfully embed themselves in a police department and simultaneously meet the demands of producing high-quality evaluative research and provide valuable guidance to the department, see Braga (2013).

and tightly defined period of time. Such programs are likely to provide stronger statements about program outcomes. A new policing initiative, or law related to police powers, that is carried out quickly and within a specific time range will allow for stronger evidence of program outcomes. This will of course affect the nature of implementation of practices and programs, and this may require some change in the everyday practices of police agencies. But criminologists must make clear that the benefits of bringing a program to the field within a clear and limited time frame outweigh the likely costs of such an approach.

Perhaps the hardest criterion for police or other criminal justice agencies to meet is what we have defined as a requirement that policies or practices be exogenous. Economists tend to scan the environment for circumstances where they can argue that the application of the intervention was not dependent on the level of the dependent variable. However, it also is possible to create strong natural quasi-experiments by working in conjunction with policy makers and practitioners. Indeed, this model is likely to provide stronger evidence of exogeneity because it has been explicitly managed.

We recognize, however, that there is a natural tendency to provide treatment to the people or places most in need. How can we encourage policy makers or practitioners to instead provide treatment independently of the intensity of the outcomes examined? First, it is important to note that we are not arguing that treatment should be applied without any consideration of the status of people or places. Going back to the hot spots experiments, police did not apply treatment to “non-hot spots.” Additionally, not just the “hottest” of the hot spots were selected by the police for treatment. Instead, hot spots were randomly assigned to treatment (meaning that all eligible places had an equal probability of being selected into treatment or control conditions). Similarly, in natural quasi-experiments, we would expect that the treated areas would evidence the problem that an intervention seeks to address. However, there is no reason why policy makers and practitioners cannot be encouraged to select a small number of potential units that fit these criteria from a larger group. For example, one of the authors (with Charlotte Gill) has implemented a community policing program with the Seattle Police Department to address hot spots of juvenile crime. It was not possible to conduct a true randomized experiment. However, the city was willing to implement the program in three sites that were chosen from a larger group of juvenile crime hot spots without regard to the level of crime. In this sense, the intervention avoided the problem of intervening at the hottest spots.

Criminologists can become better at both identifying opportunities for strong natural quasi-experiments and encouraging policy makers and practitioners to implement programs in ways that will make evaluations stronger. It is important to add a criterion for successful natural quasi-experiments. And that is that rich data be collected that allow researchers not only to assess the direct impacts on outcomes but also to provide a convincing argument for why other factors may not be influencing observations of outcomes. In the Ross (1973) study, he was able to examine not only whether the number of fatalities declined but

also whether the effects were mediated by time-of-day or day-of-week impacts. These data allowed Ross to rule out key competing hypotheses regarding the mechanism for changes in road fatalities. In contrast, the Klick and Tabarrok study did not have good information on the level of police activities on the Mall, which would have added relevant information regarding the possible impacts of police staffing on the effects observed. Although we still think that the Klick and Tabarrok study is a strong one, it illustrates the importance of having rich data for assessing possible impacts of confounding on the outcomes observed. In this regard, criminologists must begin the process of identifying what data are needed early on. They must make the scene and encourage policy makers and practitioners to collect information that will aid the evaluation later.

We close this section with several observations about the connection of arguments that we advance in this essay and those advanced by Sampson, Winship, and Knight (SWK, 2013, this issue). The theme of this special issue of CPP is the implications of causal uncertainty in research findings for the formulation of public policy. As SWK point out at the outset of their article, it is important to distinguish two distinct sources of causal uncertainty—internal versus external validity. Internal validity pertains to the degree to which a study that aims to identify the causal effect of a policy change or intervention in a particular setting is successful in validly estimating that effect. A properly conducted randomized experiment, for example, has high internal validity. In this essay, we described various forms of quasi-experimental designs that also can have high internal validity.

In the introductory essay for this issue, Blomberg, Mestre, and Mann (2013) discuss at length whether and how uncertainty about internal validity of evidence should affect its use in policy settings. Our focus has been on identifying quasi-experimental designs that can be useful in policy deliberation without sacrificing internal validity. SWK (2013) do not discuss the problem of using evidence with uncertain internal validity in the policy process. Instead, they focus on how effect heterogeneity raises questions about the policy applicability of even those studies with impeccable claims to high internal validity regardless of whether its source is an experiment or not.

Effect heterogeneity originates because in most circumstances, treatment effects are not homogenous. Treatment effects will vary across the unit of analysis—whether a person, location, or community—for a host of reasons. SWK (2013) put special emphasis on mechanisms and context as sources of effect heterogeneity. They go on to discuss why it is important that public policy be designed in recognition of the reality that there is no “one effect.” We concur with the importance of such recognition. A meta-analysis of the effectiveness of hot spots policing, for example, found that effectiveness depended on the type of strategy used, for example, with problem-oriented policing having stronger impacts than simple additions in preventive patrol (Braga et al., 2012). Policy on the use of hot spots policing should be designed with recognition that effects sizes depend on circumstances.

For innumerable reasons, causal estimates with high internal validity may have limited utility in predicting treatment effects outside the circumstances in which the treatment effect was estimated. An experiment on the effectiveness of hot spots policing in Jersey City, NJ, may be limited in what it can tell us about its effectiveness in Kinshasa, Democratic Republic of the Congo. What it tells about its effectiveness in Pittsburgh, PA, is still uncertain, but far less problematic than Kinshasa. It is important to recognize, however, that the possibility of divergence between the external and the internal validity of a causal estimate is not limited to causal estimates that are the product of randomized experiments. All causal estimates whether from a randomized experiment or not are vulnerable to this problem. The problem of extrapolating causal estimates outside the specific circumstance on which they were estimated is a universal problem that is independent of the method of estimation. For an extended discussion of the policy implications of this universal problem, see Manski (2013).

We emphasize the universality of the external versus internal validity problem because analyses of nonexperimental data often uncritically assume that the findings of the studies generalize beyond the particular historical and social setting of the study. Here we are referring to studies of how demography (e.g., age or racial composition), socioeconomic status (e.g., poverty or single-headed household), context (e.g., community income or community ethnic composition), and person status (e.g., employed or married) affect crime rates or individual-level criminal involvement. Discussions of the effects of such variables on crime commonly fall into the trap of implicitly assuming that the effects estimated by the study are homogenous and therefore generalizable to other social and historical settings. All of SWK's (2013) admonitions about the generalizability of experimental results apply equally to studies based on nonexperimental data.

There is no panacea for the problem of disjuncture between internal and external validity, but one practical way to ameliorate the problem is to evaluate replications of a generic policy in different settings for the purpose of understanding the degree and sources of effect heterogeneity. Thus, our recommendation that criminologists be proactive in identifying and designing quasi-experimental evaluations of policy changes and interventions also will serve to provide data for analyzing potential sources of effect heterogeneity. In addition, we think that criminologists must recognize the importance of heterogeneity in treatments and treatment impacts in experimental studies. This problem has been recognized for many years (e.g., see Farrington, Ohlin, and Wilson, 1986; Weisburd, 1993), but it continues to be ignored in many studies. Using block randomized designs, for example, that would recognize heterogeneity of treatment outcomes and design them into the study is one way of dealing with this problem in experimental studies (Weisburd and Gill, 2013).

We began our policy essay by recognizing the distinct benefits of making causal inferences using experimental studies. We have argued that quasi-experiments under specific conditions can offer similar benefits. However, randomized experiments remain the most believable method of drawing causal inferences when the conditions of experimental studies are met.

Conclusions

As noted in the opening paragraph of this policy essay, the online dictionary, Dictionary.com, defines evidence as “that which tends to prove or disprove something.” In this essay, we argued that both science and the policy process are well served by research with high evidentiary value. We also argued that experimental research is valuable in the policy domain not only because of its high evidentiary content but also because of its transparency. To exemplify this point, we described how a decade-long research program on hot spots policing overturned the conventional wisdom that police could not affect crime and, in so doing, has profoundly altered police practice.

Still, we recognize that for a variety of practical and ethical reasons, policy-relevant research on policing and criminal justice policy more generally cannot be based entirely on experiments. We discuss two key features of randomization—balance in expectation between the treated and control group on all potential confounders and exogeneity of treatment status—that are the source of the high evidentiary status of randomized experiments. We go on to describe how these same two features can be credibly replicated in quasi-experimental studies that are not subject to the ethical and practical obstacles that may make experiments impractical in some circumstances. We urge criminologists to follow the lead of economists in taking greater advantage of quasi-experimental research opportunities.

References

- Andenaes, Johannes. 1974. *Punishment and Deterrence*. Ann Arbor, MI: University of Michigan Press.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91: 444–455.
- Angrist, Joshua and Jön-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Ayres, Ian and Steven D. Levitt. 1998. Measuring positive externalities from unobservable victim precaution: An empirical analysis of Lojack. *The Quarterly Journal of Economics*, 113: 43–77.
- Bayley, David H. 1994. *Police for the Future*. New York: Oxford University Press.
- Berk, Richard and John M. MacDonald. 2010. Policing the homeless: An evaluation of efforts to reduce homeless-related crime. *Criminology & Public Policy*, 9: 813–840.
- Blomberg, Thomas G., Julie Mestre, and Karen Mann. 2013. Seeking causality in a world of contingencies: Criminology, research, and public policy. *Criminology & Public Policy*, 12: 571–584.
- Braga, Anthony A. 2001. The effects of hot spots policing on crime. *The ANNALS of the American Academy of Political and Social Science*, 578: 104–125.
- Braga, Anthony A. 2013. Quasi-experimentation when random assignment is not possible: Observations from practical experiences in the field. In (Brandon C. Welsh, Anthony A. Braga, and Gerben J. N. Bruinsma, eds.), *Experimental Criminology*:

- Prospects for Improving Science and Public Policy*. New York: Cambridge University Press.
- Braga, Anthony A., Andrew V. Papachristos, and David Hureau. 2012. Hot spots policing effects on crime. *Campbell Systematic Reviews*, 8.
- Braga, Anthony A., David L. Weisburd, Elin J. Waring, Lorraine Green Mazerolle, William Spelman, and Francis Gajewski. 1999. Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology*, 37: 541–580.
- Campbell, Donald T. 1969. Reforms as experiments. *American Psychologist*, 24: 409–429.
- Clarke, Ronald V. and David L. Weisburd. 1994. Diffusion of crime control benefits: Observations on the reverse of displacement. In (Ronald V. Clarke, ed.), *Crime Prevention Studies*. Monsey, NY: Criminal Justice Press.
- Cohen, Jacqueline and Jens Ludwig. 2003. Policing crime guns. In (Jens Ludwig and Philip J. Cook, eds.), *Evaluating Gun Policy*. Washington, DC: Brookings Institution Press.
- Di Tella, Rafael and Ernesto Schargrofsky. 2004. Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, 94: 115–133.
- Draca, Mirko, Stephen Machin, and Robert Witt. 2008. *Panic on the Streets of London: Police, Crime and the July 2005 Terror Attacks*. Bonn, Germany: Institute for the Study of Labor.
- Eck, John E. and Julie Wartell. 1998. Improving the management of rental properties with drug problems: A randomized experiment. In (Lorraine Green Mazerolle and Jan Roehl, eds.), *Civil Remedies and Crime Prevention*. *Crime Prevention Studies*. Monsey, NY: Criminal Justice Press.
- Farrington, David P., Lloyd E. Ohlin, and James Q. Wilson. 1986. *Understanding and Controlling Crime: Toward a New Research Strategy*. New York: Springer.
- Fielding, Henry. [1751] 1977. *An Enquiry into the Causes of the Late Increase of Robbers*. Montclair, NJ: Patterson-Smith.
- Gay, William G., Theodore H. Schell, and Stephen Schack. 1977. *Prescriptive Package: Improving Patrol Productivity, Volume I Routine Patrol*. Washington, DC: Office of Technology Transfer, Law Enforcement Assistance Administration.
- Gottfredson, Michael R. and Travis Hirschi. 1990. *A General Theory of Crime*. Stanford, CA: Stanford University Press.
- Green Mazerolle, Lorraine, Jan Roehl, and Colleen Kadleck. 1998. Controlling social disorder using civil remedies: Results from a randomized field trial in Oakland, California. In (Lorraine Green Mazerolle and Jan Roehl, eds.), *Civil Remedies and Crime Prevention*. *Crime Prevention Studies*. Monsey, NY: Criminal Justice Press.
- Haviland, Amelia, Daniel S. Nagin, and Paul R. Rosenbaum. 2007. Combining propensity score matching and group-based trajectory modeling in an observational study. *Psychological Methods*, 12: 247–267.
- Heaton, Paul. 2010. Understanding the effects of anti-profiling policies. *Journal of Law & Economics*, 53: 29–64.

- Hope, Timothy. 1994. Problem-oriented policing and drug market locations: Three case studies. In (Ronald V. Clarke, ed.), *Crime Prevention Studies*. Monsey, NY: Criminal Justice Press.
- Klick, Jonathan and Alexander Tabarrok. 2005. Using terror alert levels to estimate the effect of police on crime. *Journal of Law & Economics*, 46: 267–279.
- Leary, Mary Lou. 2013. Acting Assistant Attorney General for the Office of Justice Programs Mary Lou Leary speaks at the 8th Annual Harry Frank Guggenheim Symposium on Crime in America. Retrieved April 30, 2013 from justice.gov/iso/opa/ojp/speeches/2013/ojp-speech-130205.html.
- Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.
- Manski, Charles F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.
- National Research Council. 2004. Effectiveness of police activity in reducing crime, disorder and fear. In (Wesley G. Skogan and Kathleen Frydl, eds.), *Fairness and Effectiveness in Policing: The Evidence*. Committee to Review Research on Police Policy and Practices. Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Poutvaara, Panu and Mikael Priks. 2006. *Hooliganism in the Shadow of a Terrorist Attack and the Tsunami: Do Police Reduce Group Violence?* Finland: University of Helsinki.
- Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag.
- Rosenzweig, Mark R. and Kenneth I. Wolpin. 2000. Natural “natural experiments” in economics. *Journal of Economic Literature*, 38: 827–874.
- Ross, H. Laurence. 1973. Law, science, and accidents: The British Road Safety Act of 1967. *Journal of Legal Studies*, 2: 1–78.
- Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. Translating causal claims: Principles and strategies for policy-relevant criminology. *Criminology & Public Policy*, 12: 587–616.
- Sherman, Lawrence W. and John E. Eck. 2002. Policing for crime prevention. In (Lawrence W. Sherman, David P. Farrington, Brandon C. Welsh, and Doris L. MacKenzie, eds.), *Evidence-Based Crime Prevention*. New York: Routledge.
- Sherman, Lawrence W., Patrick Gartin, and Michael Buerger. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27: 27–55.
- Sherman, Lawrence W., Denise C. Gottfredson, Doris L. MacKenzie, John E. Eck, Peter Reuter, and Shawn D. Bushway. 1997. *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: National Institute of Justice, U.S. Department of Justice.
- Sherman, Lawrence W. and Dennis P. Rogan. 1995a. Deterrent effects of police raids on crack houses: A randomized controlled experiment. *Justice Quarterly*, 12: 755–782.
- Sherman, Lawrence W. and Dennis P. Rogan. 1995b. Effects of gun seizures on gun violence: “Hot spots” patrol in Kansas City. *Justice Quarterly*, 12: 673–694.

- Sherman, Lawrence W. and David L. Weisburd. 1995. General deterrent effects of police patrol in crime hot spots: A randomized controlled trial. *Justice Quarterly*, 12: 625–648.
- Valentine, Jeffrey C., Anthony Biglan, Robert F. Boruch, Felipe Gonzalez Castro, Linda M. Collins, Brian R. Flay, Sheppard Kellam, Eva K. Moscicki, and Steven P. Schinke. 2011. Replication in prevention science. *Prevention Science*, 12: 103–117.
- Weisburd, David L. 1993. Design sensitivity in criminal justice experiments. In (Michael Tonry ed.), *Crime and Justice*. Chicago, IL: University of Chicago Press.
- Weisburd, David L. and Anthony A. Braga. 2006. Hot spots policing as a model for police innovation. In (David L. Weisburd and Anthony A. Braga, eds.), *Police Innovation: Contrasting Perspectives*. New York: Cambridge University Press.
- Weisburd, David L. and John E. Eck. 2004. What can the police do to reduce crime, disorder and fear? *The ANNALS of the American Academy of Social and Political Sciences*, 593: 42–65.
- Weisburd, David L. and Charlotte Gill. 2013. Block randomized trials at places: Rethinking the limitations of small N experiments. *Journal of Quantitative Criminology*. Epub ahead of print. Retrieved August 6, 2013 from link.springer.com/article/10.1007%2Fs10940-013-9196-z.
- Weisburd, David L. and Lorraine Green Mazerolle. 1995a. Measuring immediate spatial displacement: Methodological issues and problems. In (John E. Eck and David L. Weisburd, eds.), *Crime and Place. Crime Prevention Studies*. Monsey, NY: Willow Tree Press.
- Weisburd, David L. and Lorraine Green Mazerolle. 1995b. Policing drug hotspots: The Jersey City drug market analysis experiment. *Justice Quarterly*, 12: 711–735.
- Weisburd, David L. and Jerome McElroy. 1988. Enacting the CPO role: Findings from the New York City pilot program in community policing. In (Jack Greene and Stephen Mastrofski, eds.), *Community Policing: Rhetoric or Reality*. New York: Praeger.
- Weisburd, David L., Laura A. Wyckoff, Justin Ready, John E. Eck, Joshua C. Hinkle, and Frank Gajewski. 2006. Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. *Criminology*, 44: 549–592.
- Welsh, Brandon C., Anthony A. Braga, and Gerben J. N. Bruinsma, eds. 2013. *Experimental Criminology: Prospects for Improving Science and Public Policy*. New York: Cambridge University Press.
- Welsh, Brandon C. and David P. Farrington. 2009. *Making Public Places Safer: Surveillance and Crime Prevention*. Oxford, UK.: Oxford University Press.
- Wilson, Orlando W. 1967. *Crime Prevention—Whose Responsibility?* Washington, DC: Thompson Book.

Daniel S. Nagin is Teresa and H. John Heinz III University Professor of Public Policy and Statistics in the Heinz College, Carnegie Mellon University. He is an elected Fellow of the American Society of Criminology and of the American Society for the Advancement of Science and is the 2006 recipient of the American Society of Criminology's Edwin H. Sutherland Award. His research focuses on the evolution of criminal and antisocial behaviors

over the life course, the deterrent effect of criminal and non-criminal penalties on illegal behaviors, and the development of statistical methods for analyzing longitudinal data.

David Weisburd is Distinguished Professor of Criminology, Law and Society at George Mason University and Director of its Center for Evidence Based Crime Policy. He also holds a joint tenured appointment as the Walter E. Meyer Professor of Law and Criminal Justice at the Hebrew University Faculty of Law in Jerusalem. Professor Weisburd is an elected Fellow of the American Society of Criminology and of the Academy of Experimental Criminology. He is the 2010 winner of the Stockholm Prize in Criminology and the 2011 winner of the Klachky Prize for the Advancement of the Frontiers of Science. His main research interests are in studies of crime at place, policing, and research methods.

Supermax Prisons

The Policy and the Evidence

Daniel P. Mears

Florida State University

Research Summary

Supermaximum-security prisons—or “supermaxes”—symbolize the “get tough” criminal justice policies that have developed over the past three decades in the United States and in other countries. Proponents believe that they effectively address critical prison system problems; opponents believe that they do not and that they create substantial harm. This essay examines the available evidence about supermaxes.

Policy Implications

A range of considerations are relevant to determining whether supermaxes constitute effective policy. These include (a) definitional problems in discussing supermax incarceration; (b) five critical dimensions along which evidence for policies is desirable and along which supermaxes fall short, including demonstration of policy need, credible policy theory, high-quality implementation, impact, and benefits that exceed costs and do so more than other policies; (c) the challenge of assessing causal claims related to supermaxes; (d) legal and ethical issues; (e) policy and political challenges confronting states; (f) policy options other than supermaxes; and (g) research gaps that remain to be addressed. The essay argues that causal uncertainty about supermax incarceration makes it difficult at present to claim credibly that it achieves intended goals. Policy implications and recommendations are discussed.

Keywords

supermax prisons, effectiveness, evidence, causation

The author thanks the Editor and anonymous reviewers for thoughtful and helpful suggestions for improving this essay. Direct correspondence to Daniel P. Mears, Mark C. Stafford Professor of Criminology, College of Criminology and Criminal Justice, Florida State University, 145 Convocation Way, Tallahassee, FL 32306 (e-mail: dmears@fsu.edu).

Supermaximum-security prisons—informally referred to as “supermaxes”—ascended into prominence during the 1980s and have since become a central feature of American corrections and of prison systems in other countries. They did not previously exist, although close approximations, such as Eastern State Penitentiary and Alcatraz, paved the way for them (Mears, 2008b; J. I. Ross, 2013; Ward and Werlich, 2003). They did not emerge in a policy vacuum. Rather, throughout the past three decades, states began investing in supermaxes at the same time that they were investing in a wide range of other “get tough” policies, such as three-strikes-and-you’re-out laws. Supermaxes thus symbolize the modern era emphasis on more severe punishment and control-oriented efforts aimed at reducing crime. They are of interest for this reason, but they warrant attention for additional reasons. For example, supermaxes arose against a background of calls for government accountability, evidence-based policy, and cost-effectiveness, but, as this essay argues, significant questions exist about how well they respond to these calls. Not least, supermaxes represent a type of policy that does not neatly lend itself to “gold standard” evaluations—that is, experiments. Careful consideration of nonexperimental approaches to assessing impact thus are needed, which in turn raises questions about what standards should be used for assessing causality and “evidence-based” policy.

A focus on supermax prisons is relevant for understanding criminal justice policy and identifying “evidence-based” approaches to punishment. And it is timely. Supermaxes emerged in part because of a belief that they cause improved outcomes, and yet considerable causal uncertainty exists regarding their impacts. In addition, supermaxes continue to be lightning rods for litigation (Mears and Watson, 2006; Schlanger, 2013). Some states have closed or are contemplating closing their supermaxes (Goode, 2012; U.S. Government Accountability Office, 2013). And nationally, some legislators, such as Senator Dick Durbin (2012), have called for reform of solitary confinement, including the use of supermax housing. The motivation for such calls stems in part from concerns about public safety and financial costs and perhaps from the prominent concerns that arose about conditions at the Guantanamo and Abu Ghraib prisons (Mears, 2004). For example, in his testimony, Senator Durbin emphasized human rights concerns, which have been prominent in many accounts of supermaxes (Jeffreys, 2013; J. I. Ross, 2013) and in condemnations of them by organizations such as Amnesty International, Human Rights Watch, and the United Nations, noting: “Our colleague and former POW John McCain said, ‘It’s an awful thing, solitary. It crushes your spirit and weakens your resistance more effectively than any other form of mistreatment.’” Subsequent to Senator Durbin’s testimony, the Federal Bureau of Prisons reportedly cut by 25% the number of inmates in segregation, closed two “Special Management Units,” and agreed to obtain an independent assessment, undertaken by the National Institute of Corrections, of federal supermax housing (Durbin, 2013). What changes, at the federal and state levels, will occur over the next decade remains, of course,

unknown, but the potential stakes for correctional systems, taxpayers, and inmates and their families are great.

With that context in mind, the goal of this essay is to examine the available evidence about supermaxes and, in so doing, examine a range of considerations relevant to determining whether supermaxes constitute effective policy. At the same time, this essay touches directly and indirectly on, and illustrates, themes emphasized in the main article to this issue by Sampson, Winship, and Knight (2013). To achieve this goal, the following topics are discussed: (a) definitional problems in discussing supermax housing and in distinguishing it from other types of solitary confinement or extended isolation; (b) five critical dimensions along which evidence for policies is desirable and along which supermaxes fall short, including demonstration of policy need, credible policy theory, high-quality implementation, impact, and benefits that exceed costs and do so more than other policies; (c) the challenge of assessing causal claims related to supermax prisons; (d) legal and ethical issues; (e) policy and political challenges confronting states; (f) policy options other than supermaxes; (g) research gaps that remain to be addressed; and (h) links between Sampson et al.'s (2013) arguments about policy making and causal uncertainty, on the one hand, and supermaxes, on the other. The essay concludes by discussing policy implications and recommendations.

Definitional Problems

Descriptive and historical accounts of supermaxes have proliferated over the past decade (see, e.g., Browne, Cambier, and Agha, 2011; Bruton, 2004; King, 2005; Kurki and Morris, 2001; Mears, 2008b; Neal, 2003; Rhodes, 2004; J. I. Ross, 2013; Shalev, 2009; Smith, 2006; U.S. Government Accountability Office, 2013; Ward and Werlich, 2003). These accounts typically have highlighted that the idea, or the design, of supermax incarceration originates with the Walnut Street Jail in the late 1700s; Eastern State Penitentiary and Auburn Prison in the early to mid-1800s; Alcatraz in the mid-1900s; and, most recently, the Marion, Illinois, federal facility. In each instance, there is an emphasis on extended solitary confinement.

What makes modern supermax incarceration different is the large-scale use of extended isolation as a “concentration” strategy (Riveland, 1999). That is, rather than disperse troublesome inmates to different facilities, states “concentrate” the putatively “worst of the worst” inmates in one place. Prison systems have always had recourse to brief lockdowns, in which inmates are confined to their cells with little to no opportunity to leave, and to maximum-security prisons. But “seldom have those prisons operated on a total lockdown basis as normal routine. Even prisons designated as maximum security have generally allowed movement, inmate interaction, congregate programs, and work opportunities” (Riveland, 1999: 5).

Supermax incarceration is not unique to the United States. Smith (2008), for example, has described the use of supermax-like confinement in Denmark from 1870 to 1920. (Interestingly, and echoing themes that emerge from contemporary critiques of supermax

prisons, he found that during this period, prison authorities acknowledged what they saw to be damaging mental health effects of supermax incarceration.) A recent edited volume by J. I. Ross (2013) has documented the use of supermax incarceration in many other countries, including Australia, Brazil, Britain, Canada, Denmark, France, Mexico, and South Africa. Other studies have pointed to the use of supermax incarceration in the Netherlands and Wales (King and Resodihardjo, 2010); Russia and Turkey (King, 2007); and Iceland, Norway, and Sweden (Smith, 2006). Although not exclusively found in countries guided by democratic forms of government, supermax incarceration seems to be largely “a product of democracies” (J. I. Ross, 2013: 2). Any such observation, however, must be tempered by the fact that no extant study comprehensively documents—through the use of a consistent definition of what constitutes “supermax” housing—the historical or contemporary use of supermax prisons globally.

Before discussing this definitional issue further, there is, first, the question of why supermax housing has become widespread. Many accounts exist, some more theoretically elaborate than others, but none that are easily susceptible to empirical assessment. Prison systems may have felt that the rapid increase in prison populations required more aggressive approaches to managing inmates. Inmates in fact may have become more difficult to manage. Outbreaks of prison violence, including murders and riots, may have played a key role. Prison system expansion led to increased hiring of less experienced officers in an occupation that can be highly stressful and combustible; the expansion and hiring of such officers may have contributed to the felt need to take new, dramatic steps to control inmates. States may have sought to find new ways of demonstrating their tough-on-crime credentials. Correctional systems may have sought to appear or to be at the “vanguard” of new ways to operate prisons. Conservative ideologies, emphasizing punishment and control, may have been a factor. Broader, societal shifts toward locating the cause of social problems in individuals, too, may have been a factor. Dissatisfaction with “failed” rehabilitative efforts may have dovetailed with such possibilities. These and other explanations have been proffered (see, e.g., King, 1999; Mears, 2008a, 2008b; Smith, 2006; Toch, 2001, 2003; Ward and Werlich, 2003). There remains, however, little empirical assessment of them, and many examples exist where the explanations do not neatly fit (see, generally, J. I. Ross, 2013).

How many supermax prisoners exist, and what exactly is a supermax? Thirty years ago, the only supermax facility in the United States was the Marion, Illinois, supermax. As of 2005, there were an estimated 44 states with supermax housing that held approximately 25,000 inmates (Mears, 2006, 2008b); Rhodes (2004) estimated that 60 supermax facilities existed. Other estimates exist (see, e.g., King, 1999; Naday, Freilich, and Mellow, 2008; National Institute of Corrections, 1997; Smith, 2006). But arriving at a precise current estimate has proven challenging. Why?

One reason is that there remains no unequivocal, agreed-upon census of supermax inmates (Butler, Griffin, and Johnson, in press; Naday et al., 2008). States use different

names for supermax housing, and some embrace the “supermax” terminology while others eschew it. Architectural issues come into play: In some cases, supermax prisons are stand-alone facilities, and in others a wing or tier of a facility may be designated for supermax “housing.” Some entail a heavy reliance on technology, others less so. Most states seem to hew to a focus on single-cell confinement for supermax housing, but some accounts have indicated that supermax housing can be used to hold two inmates to a cell. Supermax confinement entails 23-hour-per-day confinement, but some accounts have indicated that it may be shorter and that some minimal movement outside of the cell is allowed. States vary in the number of months that inmates may serve in supermax. But states do not provide annual, comprehensive empirical descriptions of actual inmate stays and variation in such stays over the life course of an inmate’s term of incarceration, or if they do, such descriptions are not readily available; a study in Florida stands as one of the few exceptions (Lowen and Isaacs, 2012; see also Mears and Bales, 2010). States vary, too, in whom they view as appropriate for supermax confinement. The phrase “worst of the worst” frequently is used to typify inmates who “belong” in supermax incarceration, but no consistent, operationalized definition of such inmates exists. In some cases, inmates must commit certain acts, such as violence, while in prison to be sent to supermax; in others, belonging to a prison gang, or “security threat group,” is sufficient (Mears and Watson, 2006).

Some accounts have distinguished between types of solitary confinement or segregation, with supermax housing constituting one type. For example, Shalev (2009: 2) described three types: (a) punitive segregation, which involves temporary isolation as punishment for misconduct; (b) protective segregation, which involves temporary isolation as a means of protecting vulnerable inmates; and (c) administrative segregation, which involves long-term isolation as a means of managing certain types of inmates (e.g., security threat group members and prisoners who have committed extreme violence, are at risk of violence, or incite others to act out). The latter is typically what captures the “spirit” of what is meant by supermax housing, although even then ambiguity emerges from the fact that prison systems sometimes house inmates in long-term isolation for what seem to be punitive, and not only managerial, purposes (Kurki and Morris, 2001; Mears and Watson, 2006; Riveland, 1999; U.S. Government Accountability, Office, 2013).

If all of these different groups are combined, and if we include inmates in isolation who are on death row, then the total number of inmates in some form of isolated confinement—both shorter term isolation and longer term “supermax” isolation—necessarily exceeds the 25,000 estimate stated earlier. For example, according to an analysis of the Census of State and Federal Correctional Facilities from the Bureau of Justice Statistics, most recently conducted in 2005 (Stephan, 2008), an estimated 81,622 inmates reside in some form of what the Bureau referred to as “restricted housing,” including individuals who are in protective custody, disciplinary segregation, administrative segregation, and death row (see Browne et al., 2011: 46; see also Jacobson, 2012).

This latter estimate is more relevant to discussions about whether *any* use of isolation—whether for short-term stays or for long-term stays—for any correctional system goals is appropriate (see, e.g., Browne et al., 2011; Shalev, 2009). Even then, definitional ambiguity undermines the validity of this estimate (see Jacobson, 2012); for example, some inmates might be placed in supermax housing for shorter periods of time than an inmate in protective custody (see Mears and Bales, 2010). The estimate of 25,000 supermax inmates likely is preferable if we wish to focus primarily on the approximate number of inmates placed in longer term segregation that is termed “supermax” or that is used primarily for managing certain inmates to achieve goals other than protecting such inmates or temporarily punishing them. Even then, it stands as but a rough approximation. It does not, for example, take into account the extent to which some inmates are repeatedly placed in such housing. And it encounters the same challenges that have arisen with other attempts to estimate the supermax population (e.g., Butler et al., in press; King, 1999; Naday et al., 2008; Riveland, 1999).

In the end, the elements of supermax housing that are common to most accounts, and that constitute the core aspects of the definition provided by the National Institute of Corrections (1997), include single-cell confinement (i.e., one person to a cell) for extended periods of time—beyond what would be associated with temporary stays for punitive or protective segregation—with little to no opportunities for programming, services, and visitation. Indeed, in a national survey of prison wardens, which used a similar definition, greater than 90% of respondents agreed with this definition (Mears and Castro, 2006). That does not mean that the definition is “correct” (Butler et al., in press; Naday et al., 2008). But it does suggest that it captures several critical dimensions of what conventionally is meant by “supermax” incarceration, including the emphasis on extended isolation with a priority given to achieving management-oriented goals (Shalev, 2009: 2).

Supermaxes are funded by taxpayers, but decisions to enact the policy derive in large part from legislatures, which may call for the use of supermax housing, or from corrections officials, who also may call for them. However, they do not reflect a type of sentencing law. Rather, they constitute a type of administrative tool that corrections officials can use to help manage the prison system (Shalev, 2009: 2). Accordingly, the placement of inmates in supermax housing is an administrative decision. The courts do not sentence individuals to supermax confinement. Rather, prison officials determine whether to have supermax housing and how and for whom it shall be used. Here, then, we have the basis not only for considerable variation in what is meant by “supermax” housing but also for variation in the goals and uses to which any type of extended isolation, whether termed “supermax” or not, is put.

Ambiguity concerning the definition and goals of supermaxes is paralleled by ambiguity concerning the relevant counterfactual condition—that is, what would have been implemented instead of supermax housing. For example, supermax housing has been described as arising in lieu of interventions that would have been used to achieve similar goals

(Mears, 2006). That is, such housing is not always created as an alternative to building some type of prison; instead, it represents an intervention aimed at achieving particular goals. Accordingly, any account of supermax effectiveness must address the precise purposes for supermax housing in a particular context, how and for whom such housing is employed, and what alternatives to supermaxes, or what other uses of the funds used for them, would have been adopted.

Five Key Dimensions Along Which Evidence Is Needed

Policies ideally are evidence based. Indeed, that ideal has been emphasized explicitly by policy makers in recent decades (Welsh and Harris, 2008). But what do we mean by “evidence?” Some accounts have defined evidence as the existence of results from experimental studies. Others have defined it as resulting from using experiments or quasi-experimental designs to evaluate a policy. Still other accounts have defined evidence-based policy as arising when we have “scientific evidence,” such as empirical research of some unspecified type. Similar inconsistency arises in descriptions of what is meant by government “accountability” (Mears, 2010).

Without a doubt, a central focus when discussing “evidence-based” policy should presumably center on the question of whether the policy produces intended outcomes. Specifically, does it cause beneficial outcomes to arise that otherwise, absent the policy, would not? However, this approach is arguably too narrow. It ignores, for example, the basic question of whether an effective policy is needed, and it ignores the salience of existing theory and research for a priori assessing the credibility of causal impact claims and evaluations.

A broader, more systematic approach to defining evidence—and to considering causal claims about specific policies (or programs, interventions, rules, practices, etc.)—is to use the evaluation hierarchy. Rossi, Freeman, and Lipsey (2004) described the hierarchy as consisting of five levels, including assessment of (a) the need for a policy; (b) the theory, or causal logic, of a policy; (c) policy implementation; (d) impact; and (e) cost-efficiency. Accordingly, we can define an “evidence-based” policy as one where we can successfully conclude that each level of the hierarchy has been examined and resulted in an affirmative assessment of the policy (Mears, 2010). For example, an evidence-based policy would be one in which there is an empirically demonstrated need for it, credible theoretical or causal logic for expecting impacts, high-quality implementation, demonstrated impact(s), and empirical documentation that the benefits exceed the costs and do so to a greater extent than potential alternatives.

The type of “evidence” needed for each level of the hierarchy may vary, but theory, logic, and empirical research typically apply in each instance. In addition, although causal claims are clearly prominent when discussing impact, they are relevant as well to the other dimensions of the hierarchy, but in ways that are not always or even frequently resolvable through recourse to experiments or quasi-experimental research. Use of evaluation hierarchy is helpful for many reasons, including identifying what has not been addressed in prior

research and in understanding better the policy dimensions that a given study or set of studies addresses. Among other things, it can provide a corrective against overstating the relevance of any one study to debates about the merits of a given policy (Mears, 2010; Rossi et al., 2004).

If we apply the evaluation hierarchy to supermax prisons, what do we learn and what lessons can we draw? In the subsequent discussion, I briefly summarize the state of research with respect to each of the dimensions discussed earlier and, in so doing, draw on and extend a prior assessment (Mears, 2008b). Although meta-analyses are ideal for arriving at systematic assessments of research on a given topic, there remain too few studies of supermaxes, with respect to the five dimensions, to undertake such assessments. In the subsequent sections, I discuss the implications of research on supermaxes and related issues that bear on discussions more broadly of causality and policy.

Need

Were supermaxes needed in the first place? Are they needed now? The short answer—nobody knows, at least not if empirical research is the basis for answering the question. The longer answer entails consideration of several related questions. For example, what problems were supermaxes designed to address? How large were the problems? What produced them? Were supermaxes the most logical solution? For example, did they best address the cause of the problem and do so at the least financial or human cost?

In many accounts, policy makers and prison officials justify supermaxes on the grounds that they serve to “control” the “worst of the worst” (King, 2005; Riveland, 1999; Shalev, 2009). But placing inmates in extended solitary confinement is not a goal. Rather, it constitutes a means to an end. What, then, is that end, or goal? It varies across states and depends on the specific officials who are asked; frequently, there is no explicitly stated goal or the language opaquely refers to a focus on “controlling” certain inmates, without clarifying what broader goals this focus addresses (Mears and Watson, 2006). In one national survey of wardens, respondents said that the main goal of having supermaxes was to increase safety and order throughout the prison system, but half or more said that supermaxes served other goals, including incapacitation of violent inmates, decreasing riots, reducing the influence of gangs, preventing escapes, punishing inmates, and reducing recidivism (Mears, 2005). One third of respondents said a goal of supermaxes is to improve public safety through general deterrence.

In discussions with policy makers and officials in different states, I have been told in no uncertain terms what “the” goal of supermaxes is, only then to be told by other policy makers and officials—again, in no uncertain terms—that “the” true goal differs. Frequently, no government or correctional system documents are available to document a given goal, or the documents, as noted, describe a strategy of housing violent and disruptive inmates in isolation without clarifying precisely the goals that this strategy achieves (Mears and Watson, 2006).

In the end, there is no “right” or “wrong” goal, only the goal that was envisioned by those responsible for creating supermaxes, that is, policy makers and corrections officials. To date, the precise and varied goals associated with specific supermaxes have not been systematically assessed within or across states, save for one study that documented a wide range of competing views about the putative goals that supermaxes are supposed to achieve (Mears, 2005; Mears and Castro, 2006; see also Mears and Watson, 2006). Regardless, to the extent that a given state has several goals for its supermax housing, then any evaluation of impact should assess the degree to which the housing achieves these specific goals.

Focusing on the goals associated with supermax is important because it draws our attention to the problem that they are intended to solve. In turn, we can turn to the questions outlined earlier. For example, how large is the problem and what caused it? Take the system-wide order and safety goal as an example. How much disorder and violence existed in a given prison system, and to what extent was it caused by a handful of the “worst of the worst” inmates (putting aside the question of how such inmates are defined and identified)? Answering that question, of course, requires having reliable and accurate information about the prevalence of disorder and violence and various factors that might have contributed to it. Few, if any, states, however, have database systems that accurately collect such information (Bottoms, 1999; Logan, 1993; Mears, 2008a; Reisig, 1998). Perhaps not surprisingly, then, it is difficult to identify any comprehensive, credible, empirical assessments of system-wide disorder and violence, the causes of that violence, and evidence that a select group of “worst of the worst” inmates contributed to it. Perhaps they did, or perhaps the levels of or increase in disorder and violence stemmed primarily from other factors, such as administrative management approaches; staff professionalism; availability and quality of programs, treatment, and services; and staff and inmate culture (Adams, 1992; Gendreau and Keyes, 2001; Irwin, 2005; Mears and Reisig, 2006; Sparks, Bottoms, and Hay, 1996).

By extension, it also is perhaps not surprising that, at least in some cases, states have begun to rethink their use of supermax prisons (U.S. Government Accountability Office, 2013: 34). For example, Walter Dickey, who was the director of the Wisconsin Division of Corrections from 1983 to 1987 and subsequently the Federal Monitor for the Supermax Prison at Boscobel, Wisconsin, noted on National Public Radio (2012): “I think one of the things that’s happened, at least in a lot of states, Wisconsin’s one of them, is I think we grossly exaggerated the need for the supermax prison and overbuilt it, and I think, not surprisingly, when you’ve got empty cells in a crowded prison system, you tend to fill them up.”

A final observation—one of the more compelling arguments for how supermaxes emerged—is that corrections officials and policy makers felt that an emergency, such as a riot or spate of murders, existed (Mears and Watson, 2006). They then felt that dramatic steps needed to be taken to take control of prison systems and return to some level of order and safety (J. I. Ross, 2013: 7). However, prison officials do not all agree that supermax incarceration is appropriate or necessary (Goode, 2012; Mears and Castro, 2006; Mears and Watson, 2006).

In addition, it is not clear that supermaxes have been built in response to riots (J. I. Ross, 2013: 180). Even if they were, the need in such cases is highly specific—an emergency exists that needs to be resolved in the immediate term until a longer term solution can be found. Supermax housing might help correctional systems to regain order, if only temporarily, in situations in which riots or widespread violence occur, although that is open to debate (Bottoms, 1999; Ward and Werlich, 2003). For example, a supermax cannot be built instantaneously to resolve a riot. Regardless, even if supermaxes were useful in emergency situations, using such an approach on a long-term basis would be akin to using emergency rooms on a widespread basis to address basic health-care problems. It simply would not make sense. For everyday health-care problems, non-emergency-room approaches constitute a more effective approach to creating health and treating sickness. Similarly, any correctional system response to emergency situations does not, on the face of it, logically indicate that the response is appropriate to everyday management of the system.

Recently, a U.S. Government Accountability Office (2013: 33–34) report documented several federal Bureau of Prison officials, who stated that supermax-like housing has enabled the prison system to reduce the use of lockdowns. Here, again, the question arises as to whether supermax housing stands as an obvious solution to some unspecified number of avoidable lockdowns. An answer necessarily involves understanding how specific inmates contribute to the lockdowns and what provoked their behavior or allowed it to occur. It also necessarily involves understanding what approaches other than supermax housing might work potentially better and at less cost.

Theory

A credible theory, or causal logic, typically is the bedrock of any policy. Without it, the policy is unlikely to produce the desired outcomes (Rossi et al., 2004). A credible policy theory might be one that draws on different theories that themselves have been well tested. It also might make assumptions that are empirically justified. For example, a program aimed at reducing recidivism might focus on reducing an ex-prisoner's level of strain. Doing so would fit well with a general strain theory perspective. However, it would not make sense if the ex-prisoner experienced relatively little strain and if, instead, his or her likelihood of offending stemmed largely from low self-control. A policy also is more likely to produce benefits to society if no sound theoretical argument can be presented for the policy causing harm. So, supermax incarceration may make sense if it rests on credible theories and if the assumptions on which it rests are empirically supported. It also may make sense if no credible theoretical argument exists that such incarceration may worsen inmate or prison outcomes.

Several accounts have highlighted that these conditions are not met. Certainly, theoretical arguments can be presented that seemingly make a credible case for supermaxes. For example, drawing on general deterrence theory, one might argue that supermax incarceration is so severe as to inspire fear among general population inmates and so induce order.

Similarly, one could make the case for an incapacitation or specific deterrent effect; housing especially unruly and violent inmates in isolation should prevent them from committing violent acts or from inciting others to commit them, and it may inspire fear and lead them to refrain from such acts after release.

Upon closer inspection, such accounts can be found to rest on assumptions that seem to be unrealistic or speculative. For example, few inmates are sent to supermax housing and those who are, in theory, will have committed extreme acts of violence. General population inmates thus could reasonably infer that there is little certainty of supermax confinement (Pizarro and Stenius, 2004). And supermax inmates, if they truly constitute the “worst of the worst,” would seem to be unlikely to be deterred by any special type of incarceration; indeed, some inmates may prefer supermax housing or may seek it out as a badge of honor (Mears and Watson, 2006; Singer, 2003). Some accounts have suggested that removing disruptive inmates “normalizes” the prison environment in the rest of the prison system. However, this argument rests on the tenuous notion that removing a handful of inmates from a given facility indeed frees up officers to invest more time watching other inmates, allows typical operations to unfold “normally,” and enables inmates to conform with prison rules. Perhaps in some instances these claims hold true, but it is not clear that they would be typical (Mears and Reisig, 2006).

At the same time, many accounts have suggested that supermaxes, by design, may anger inmates and lead them to act out more violently (Mears and Watson, 2006; Rhodes, 2004; Ward and Werlich, 2003). For example, in her study, Rhodes (2004) described inmates who, after long-term solitary confinement, could no longer function in general population facilities. One officer noted: “[Inmates] spend so much time in single cells they get very paranoid. We have an inmate [who] went to a regular unit, but he only lasted a day. . . . He said, I can’t stand it, people come up and talk to me” (p. 34; emphasis in original). The lack of services and programming in supermax housing, too, could be argued to contribute to a greater, rather than to a lesser, likelihood of engaging in violence after release (Haney, 2003; King, 2005).

Numerous accounts point to the seeming inhumanity of supermax incarceration and how it seems designed less to improve inmate behavior and more to degrade it (see, generally, Kurki and Morris, 2001; Rhodes, 2004; J. I. Ross, 2013; Shalev, 2009; Tietz, 2012). These accounts may be correct. Indeed, a large body of studies, examined in several reviews (e.g., Haney, 2003; Schlanger, 2013; Shalev, 2009) and highlighted in popular accounts (e.g., Gawande, 2009; Tietz, 2012), has suggested that supermax confinement adversely affects mental health. These studies have provided compelling, empirically based arguments that the very conditions of such confinement contribute to mental illness, suicide, and impaired psychological functioning. At the same time, as other reviews have emphasized (e.g., Mears, 2008b; O’Keefe et al., 2013), few of them employ strong research designs with appropriate comparison groups. Also, some studies have found no adverse effects of supermax housing on mental health (see, generally, O’Keefe et al., 2013; Smith, 2006). There is, for example,

the question of how much worse supermax incarceration affects mental health as compared with what would happen in maximum-security facilities. Considerable challenges arise in deciphering extant research and, not surprisingly, considerable debates exist about which approaches and studies should be given greater weight (Smith, 2006: 450). Even so, the body of work to date has provided reasonable theoretical counter-arguments that undermine the notion that supermaxes produce substantial benefits. In addition, historical accounts have indicated that prison systems themselves frequently have condemned the use of supermax housing because of its perceived-to-be adverse effects on inmate mental health (Smith, 2006, *passim*).

If we focus on system-wide order and safety as a goal, the logic of supermaxes rests on the assumption that a large amount of disorder and violence result from the actions of a small subset of the inmate population. This theoretical claim runs counter to the vast bulk of criminological theories on prison order. As Sparks et al. (1996: 313) have emphasized: “What [special prison units] cannot do is magically to unlock the problem of order for a prison system as a whole.” For example, the administrative philosophies of prison systems may have a considerably larger aggregate effect on system-wide order than that of individual inmates (DiIulio, 1987) and so, too, can the culture among administrators, wardens, officers, and inmates (Bottoms, 1999). Prison officials can hope, of course, that supermaxes do constitute an effective silver bullet solution. As Rhodes (2004: 36) has noted, “The dream of the perfect prison has deep historical roots.”

In the end, though, can we, on theoretical grounds, anticipate that supermaxes improve system-wide order and safety? Yes, but the assumptions for this effect are difficult to sustain. In addition, theoretical arguments can as credibly be made that they worsen outcomes, not only for inmates who experience supermax incarceration but also for the entire system. If we examine the theoretical logic for other outcomes, such as reduced riots, escapes, or recidivism, similar problems arise (see, generally, Briggs, Sundt, and Castellano, 2003; Gendreau and Keyes, 2001; Irwin, 2005; King and Resodihardjo, 2010; Mears and Watson, 2006; Shalev, 2009).

Implementation

Are supermaxes well implemented? For example, are they operated in ways that accord with their design? Are the “correct” types of inmates placed in supermax housing? Are they provided access to some programs or services? Are protocols followed for ensuring that the constitutional rights of supermax inmates are met and for releasing inmates from supermax housing? To date, few studies have directly addressed these or related questions (Mears, 2008b). Indeed, not unlike many aspects of prison systems (Mears, 2008a), supermaxes operate largely in the equivalent of a “black box” (Katel, 2012; Kurki and Morris, 2001; Mushlin, 2012; see, however, Reiter, 2012a; U.S. Government Accountability Office, 2013). With that said, several concerns can be gleaned from prior research.

First, absent a clearly defined supermax “design”—one that precisely defines the goal of supermax housing in a particular system, who should be in supermax, how long an inmate should be in supermax, and so on—it is impossible to assess how well supermaxes are implemented. For example, perhaps fewer or more inmates warrant placement in supermaxes. Opponents may well feel that no one should be in supermax confinement, but proponents presumably would want information on how many inmates are not placed in supermax incarceration that should be. Such information does not at present exist or has not been made public by correctional systems. Some accounts have provided empirical descriptions of supermax inmates (see, e.g., Cloyes, Lovell, Allen, and Rhodes, 2006; Lovell, Cloyes, Allen, and Rhodes, 2000; Mears and Bales, 2010; Reiter, 2012a). However, despite the frequent reference to the notion that supermaxes are supposed to house the “worst of the worst” inmates, it remains unclear what this expression means both conceptually and empirically (King, 1999: 164). It is unclear, by extension, what specific behaviors land inmates in supermax confinement. Here, the focus is not just on those behaviors listed in administrative codes or written policies but those that *actually* result in inmates being sent to supermax housing. Many accounts exist that have documented seemingly arbitrary application of policies to individual inmates (see, e.g., Bauer, 2012; Cohen, 2013; Goode, 2012; Kurki and Morris, 2001; Rhodes, 2004).

Second, many different dimensions of supermax incarceration bear measurement and monitoring and yet remain largely unexamined. To illustrate, few studies exist that have document how long inmates serve in supermax custody (U.S. Government Accountability Office, 2013). Media accounts typically have profiled those inmates who serve many years in supermax housing (see, e.g., Goode, 2012). It is, however, unclear how representative such inmates are, and even presenting averages can be misleading if a sufficient number of extreme cases exist. In Texas, for example, it was reported that the average duration of supermax confinement was 4 years, with the longest period of time served in such confinement being 24 years (Jennings, 2009). By contrast, in Virginia, it was reported that inmates in supermax confinement averaged 2.7 years there, with some inmates serving up to 7 years (Kumar, 2012). In a study of Florida supermax inmates, a colleague and I found the following: Many inmates spent “only” a few months in supermax confinement, while others spent years in it; over half were placed in supermax confinement three or more times; some spent less than 15% of their total term of incarceration in supermax confinement, while 14% spent more than half of their total term of incarceration in such confinement; and 28% of supermax inmates were released from supermax housing within 3 months of release to society (Mears and Bales, 2010). Perhaps Florida is an exception; perhaps it is typical. At present, nobody knows.

The duration of supermax housing constitutes, however, but one dimension for which information about implementation is needed. Other dimensions about which little is known empirically include the variation in discretion exercised by wardens in determining who is sent to supermaxes, training and experience of supermax officers, and the extent to which

intermediate sanctions were used prior to using supermax housing. Little is known, too, about dimensions that have been expressly identified in various court decisions, such as provision of medical care, appropriate use of force, screening for mental illness, due process when admitting and releasing inmates from supermax housing, allowing access to the courts, and more (Collins, 2004).

Third, many studies have highlighted that states house a seemingly non-trivial number of inmates in supermax housing who do not belong in them. For example, Haney (2003) and others (e.g., Austin and McGinnis, 2004; Browne, Agha, and Austin, 2012; Browne et al., 2011; DeMaio, 2001; Kurki and Morris, 2001; Lovell, 2008; Lovell et al., 2000; Mears and Watson, 2006; Rhodes, 2004; Riveland, 1999; Shalev, 2009; Sullivan, 2006) have highlighted that nuisance inmates and mentally ill inmates are found in supermax housing and do not in any obvious way fit the profile of who, by design, should reside in such housing and who, under some court decisions, should not be allowed in it. Given the diversity of views that wardens have about who should be placed in supermax housing, such occurrences should not be surprising (Mears, 2005; Mears and Castro, 2006; Wells, Johnson, and Henningsen, 2002). Even so, it raises concerns about the appropriate use of supermaxes, recognizing here that to some opponents of supermaxes no use of supermax housing is appropriate.

Fourth, given concerns about differential treatment of minorities in the criminal justice system (Schlanger, 2013), attendant concerns arise about the potential for supermax incarceration to be used disproportionately for minorities. The previously mentioned Florida study, for example, identified that Blacks were more likely to be placed in supermaxes, although this difference seemed to be explained by greater involvement in misconduct (Mears and Bales, 2010). In the latter instance, though, there is the question of whether racial or ethnic differences in misconduct stem from ways in which prisons are structured and operated. If, for example, prisons operate in ways that contribute to racial and ethnic differences in violence, then supermaxes potentially serve to amplify such differences. Few studies have investigated these possibilities, although two recent accounts have suggested that supermax incarceration is used more for minorities than for Whites. Hispanic inmates have been reported to be overrepresented in supermax housing in California (Janquart, 2013; Reiter, 2012a) and in Arizona (Lowen and Isaacs, 2012; J. Ross, 2013). Schlanger (2013) has reported that Black inmates constitute a disproportionate percentage of the New York supermax population and found that racial and ethnic overrepresentation in supermax facilities may exist in other states, including Arkansas, Colorado, and Connecticut. Whether such differences arise from differences in misconduct or how prison systems manage or treat minority inmates remains unknown (Mears and Bales, 2010).

Ultimately, these and similar “black holes” in our understanding of supermax prisons raise fundamental concerns. They signal, for example, the risk that supermaxes are not implemented as designed and that abuses may exist. More generally, they indicate that any potential effectiveness of supermaxes may be undermined by poor implementation. Not

least, without information about these and other dimensions of implementation, it would be nearly impossible to identify how exactly any identified beneficial effect of supermax prisons arose.

Impacts

Despite the aggressive expansion in the use of supermaxes over the past three decades, there remain virtually no comprehensive assessments of their impacts along a range of dimensions (e.g., serious or minor misconduct among supermax inmates and non-supermax inmates, system-wide order and safety, escapes, riots, murders, and programming). Indeed, almost no impact evaluations have been conducted, and those that have been undertaken examine a narrow range of relevant outcomes and have methodological shortcomings that limit their generalizability. Elsewhere, I have discussed several studies that have examined different outcomes and issues (Mears, 2008b) and so here highlight several of the more prominent considerations.

First, the wide range of goals associated with supermaxes means that any balanced assessment of impact should consider all of them (Clare and Bottomley, 2001; Mears and Castro, 2006; Mears and Watson, 2006; J. I. Ross, 2013; Shalev, 2009). For example, if the goals are to make entire prison systems safer and more orderly, as well as to reduce recidivism among the most difficult inmates, then these dimensions should be included in an impact assessment. No such studies exist to date.

Second, a credible assessment of impact requires an understanding of the counterfactual—that is, what would have happened had a given state not created and used supermax housing? The answer to that question seems to vary by state and, as discussed, to be complicated. For example, in some cases, supermaxes were conceptualized as the equivalent of interventions (King, 1999; Mears and Watson, 2006). Here, the counterfactual is uncertain. Absent a supermax, would the state have done nothing? Enhanced officer training? Increased dispersion strategies? Built a maximum-security facility? Invested in rehabilitative programming for violent prisoners? Each question creates a different point of reference for determining whether supermaxes achieve their goals and for the types of research designs needed for answering them. Such contextual factors likely limit the external validity of state-specific studies of supermax prison impacts (see, generally, Mears, 2008b; Mears, Cochran, Greenman, Bhati, and Greenwald, 2011; see also Sampson et al., 2013).

Third, an evaluation of supermax housing impacts cannot, for ethical reasons, rely on an experimental design. Accordingly, we will be unlikely to ever have scientific “evidence” of impacts if the standard is the use of experiments. If, however, other definitions of “evidence” are used, then, as we will discuss, we may be able to accumulate evidence about supermax impacts.

Fourth, a quasi-experimental approach may be feasible in some instances, although it has not been employed except in a small handful of cases (U.S. Government Accountability

Office, 2013). For example, Briggs et al. (2003) compared inmate assaults before and after supermaxes were opened in several states (Arizona, Illinois, and Minnesota) and used Utah, which did not open a supermax, as a point of comparison; the study found little evidence of supermax effects on system-wide assaults. Using quasi-experimental designs, Lovell, Johnson, and Cain (2007) and Mears and Bales (2009) examined supermax inmate recidivism and identified small increases in postrelease offending. No studies to date have employed strong research designs for examining the effects of supermax housing on inmate misconduct during and after release from such housing (see, however, Ward and Werlich, 2003). Qualitative accounts exist as well and have shed some light on possible impacts of supermax incarceration (see, e.g., Bidna, 1975; Crouch and Marquart, 1989; Rhodes, 2004); yet they did not employ the types of evaluation designs that would allow for greater confidence in causal claims about the effects of supermaxes on beneficial outcomes (e.g., reduced violence) or adverse outcomes (e.g., increased violence, mental illness, or suicide). Systematic assessment of a wider range of outcomes, using a wider range of measures, remains to be undertaken. A central challenge for any credible impact evaluation, especially one focused on systems-level effects, is that of taking into account the many other changes that typically occur when supermaxes are opened that may account for any observed changes in order or safety.

Fifth, an increasingly large literature has highlighted the importance of identifying and quantifying the potential harms, noted earlier, that may arise from supermaxes (see, generally, Mears, 2005; Mears and Watson, 2006). These harms may occur among all inmates in supermaxes or among some groups of them, such as the young (Human Rights Watch, 2012), or among officers (Mears and Watson, 2006; Shalev, 2009) or inmates' families or among the prison system as a whole (Briggs et al., 2003). For example, any benefits of supermax incarceration may be offset, if only partially, by any effect that it has in causing or aggravating mental illness among supermax inmates. Indeed, a large literature—including research studies and inmate and warden accounts—has argued that extended isolation causes serious mental illness and is psychologically debilitating in ways that endure long after inmates are released (e.g., Bulman, Garcia, and Herson, 2012; Cloyes et al., 2006; Haney, 2003; Kupers, 2008; Lovell, 2008; Lowen and Isaacs, 2012; O'Keefe et al., 2013; J. I. Ross, 2013; cf. Shalev, 2009; Tietz, 2012). In a related vein, it is conceivable, too, that the use of supermax prisons may worsen the behavior of inmates placed in them and of general population inmates (Briggs et al., 2003; Lovell and Johnson, 2004; Mears and Reisig, 2006). That possibility has been raised not only by opponents of supermax prisons but also by prison officials and wardens who have direct experience with supermaxes (see, e.g., Mears, 2005; Goode, 2012).

Cost-Efficiency

Is supermax housing cost-efficient? No one knows. The costs alone may vary from state to state or even within a state. In some cases, states retrofit a wing of an existing facility,

which entails different costs than building a brand new facility to house only supermax inmates. Estimates typically suggest that the costs to build supermaxes are substantially greater than those to build other types of facilities because of the greater investment in technology and single-cell housing. Operational costs tend to be substantially greater because, in contrast to traditional prison facilities, inmates do none of the work; all services, treatment, food delivery, and the like must be undertaken by officers. In a study that a colleague and I conducted, we heard estimates from corrections officials that indicated that supermax housing typically was two to three times more expensive to build and operate than would be the case with traditional maximum- or medium-security facilities (Mears and Watson, 2006). News accounts and some scholarly accounts tend to echo this estimate, but a systematic cost assessment, along the lines of what would be ideal (see, e.g., Gaes, Camp, Nelson, and Saylor, 2004; Lawrence and Mears, 2004), remains to be undertaken. Without credible estimates of costs, it is impossible to obtain valid estimates of cost-efficiency. It also is impossible to obtain them without valid estimates of positive and negative impacts.

Uncertainty about the impacts of supermax housing is critically relevant for generating accurate cost-benefit estimates. And so, here, again, we encounter the question of what counterfactual to use (Lawrence and Mears, 2004). For example, a state may build a supermax facility and the counterfactual may be that the state would have invested in a maximum-security prison. In this case, the cost-efficiency estimate should center on the estimated impact of the supermax, as compared with use of a maximum-security prison, and the estimated costs of the supermax housing and the maximum-security prison, respectively. If, however, the state would have invested in officer training or strategic dispersion of inmates throughout the prison system, then the estimated supermax impacts are relative to these comparisons and to the costs associated with each them.

Other challenges, too, exist in arriving at credible estimates of cost-efficiency. For example, estimated impacts of supermaxes should consider all relevant outcomes, including potential harms. We have abundant evidence that supermaxes may have a wide range of effects, including potentially increasing the risk of suicide (Smith, 2006: 499), and so this step is essential. In addition, there is the importance and difficulty of monetizing the diverse outcomes (e.g., increased or decreased assaults, murders, mental illness, rule compliance, and recidivism). Not least, given the wide range of differences in how supermaxes are designed and operated and, again, the purposes for which they were built, any one cost-efficiency evaluation likely would not generalize to other states (Browne et al., 2011; Jacobson, 2012; King, 1999; Lawrence and Mears, 2004; Mears, 2008b). Accordingly, until a body of studies emerges that arrives at credible estimates of impact and involves comparisons with similar counterfactual scenarios, and until such work emerges that also involves credible efforts to monetize various impacts, it will not be possible to estimate the cost-efficiency of supermaxes.

Causation and Assessing Effectiveness

Causation, the focus of this special issue, is a complicated topic (see, generally, Marini and Singer, 1988). From a policy perspective, the “simple” approach to the topic entails using some number of experimental designs as evidence that a given program, policy, intervention, or the like can produce a beneficial outcome. As Sampson et al. (2013) have highlighted, even this approach is not so simple. The results of a given experiment, for example, may not generalize to other contexts or populations, and in many instances, such as various state laws or U.S. Supreme Court decisions, experiments that have real-world relevance cannot be undertaken (Heckman and Smith, 1995; Mears, 2010; Mears et al., 2011; Rossi et al., 2004; Sherman, 2003). Such issues aside, discussions about “evidence-based” policy frequently have treated experimental or quasi-experimental assessments of impact as the primary foundation on which to justify policy and where, by extension, causal questions are most relevant (Welsh, Braga, and Bruinsma, 2013).

Here, I briefly revisit the five dimensions outlined earlier—need, theory, implementation, impact, and cost-efficiency—and highlight how causation is relevant to each and why discussions about policy effectiveness should focus on each. In so doing, the observations take heed of Sampson et al.’s (2013) argument that “descriptive data, noncausal analysis, and criminological theory remain essential to improving the link from causal claims (even if unassailable) to policy.”

Need

An assessment of need ideally not only documents the problem that should be addressed but also identifies the policy that can best address it. Policies often are put forward as solutions to vaguely specified problems (Rossi et al., 2004). In such cases, greater clarity about the problem can help. Clarity about why a particular policy or intervention would be effective can be helpful as well. To arrive at such clarity, it helps to know what caused the problem, whether it is feasible to target those causes, and whether doing so would likely produce, or cause, desired outcomes. Here, then, we arrive squarely at questions of causality, ones that typically may not be readily answered through experiments or even quasi-experimental designs. If we consider prison systems, for example, what exactly produces order and safety? What role do certain subsets of inmates have in producing disorder and violence? What factors moderate the effects that such inmates have? What mechanisms could be leveraged to affect these inmates or these moderating factors? These constitute questions of direct relevance to establishing the need for a given policy, and they necessarily give rise to a host of questions about causal relationships and what solutions to a given problem would be most indicated or effective.

Theory

Social policies—including, again, programs, interventions, rules, protocols, and the like—typically entail relatively complicated combinations of services, treatments, dosages, and

more, and these unfold over time. Assessing the likelihood that a given policy design will contribute to a desired outcome is therefore challenging and, as emphasized earlier, it may not be possible to employ rigorous research methodologies to evaluate policy impact. Even so, the many assumptions that undergird a policy can be subjected to evaluation (Mears, 2010; Rossi et al., 2004; Silver, 2012). If no existing theories—ideally ones that have been well tested—can be identified to support some, many, or even all of the claims on which a policy is built, then as a logical matter, we have grounds to question whether the policy can or will be effective. Similarly, if empirical research has cast doubt on key claims that undergird the policy, we again should have doubts about whether causal claims about the policy are accurate.

When we turn to supermax prisons, for example, questions about whether they can or do achieve desired outcomes arise. The theory underlying them is frequently unspecified by states, the application of theories to their operations suggests that they may create as much harm as benefit, and the empirical claims on which they rest are generally untested or implausible (see, e.g., King, 2007; Mears and Reisig, 2006; Mears and Watson, 2006; Pizarro, Stenius, and Pratt, 2006; J. I. Ross, 2013). For example, one assumption is that supermax inmates will be deterred into conforming behavior; yet such inmates have not, on the face of it, been deterred by any other approach, and some studies have suggested that, in some instances, inmates may want to be placed in supermax housing, whether for protection or as a symbol of toughness (Mears and Watson, 2006). Such assessments do not mean that supermaxes do not or cannot create the outcomes for which they were intended, but they do raise red flags about the likelihood of such occurring.

Implementation

The quality of policy implementation is central to assessing causal effects (Rossi et al., 2004). For example, an impact evaluation might identify the positive effects of a program, and in this case we might reasonably conclude that the program caused the effects. Yet, programs frequently are not implemented as designed. In such cases, we should question whether an identified impact results from a program or from some other factor. Process, or implementation, evaluations can assess the quality of policy implementation and, in turn, support causal claims arising from impact evaluations. With complicated policies or programs, they also can be used to help identify what parts of the policies or programs may be most responsible for producing particular outcomes. Consider drug courts—positive impacts may arise from treatment, closer supervision, more frequent court contact, or some combination of these factors (Mears, 2010). An implementation evaluation can be used to assess the degree and quality of the activities in a given drug court and so provide some foothold for determining whether one activity may be more important than another. For example, perhaps the implementation evaluation identifies that treatment services in fact were minimal and yet a large program effect emerged. Here, one then has grounds to question the causal claim that treatment produced the effects.

In the end, no amount of research concerning the causal effect of a given policy can assure that the policy will be implemented in a high-quality manner in a given context. Implementation evaluations thus are essential to increasing the likelihood that the expected benefits of a given policy, when adopted in a different setting, occur. For example, extant studies have suggested that states not infrequently place the “wrong” types of inmates, such as those who engage in nuisance infractions, in supermax housing (Browne et al., 2011, 2012; Haney, 2003; Lovell et al., 2000; Mears, 2008b; National Institute of Corrections, 1997). More broadly, studies and media accounts have indicated that the disjuncture between ideal and actual practice in supermaxes can be considerable (Bauer, 2012; Filho, 2013; Mears and Watson, 2006). In such instances, not only do inmates not receive the intervention that was intended, but also they potentially are abused or perceive their treatment as unfair, which can result in worse rather than better outcomes (Carlton, 2007; Mears, 2008b; Shalev, 2009). Here, again, causation is implicated—failure to implement a policy in the manner in which it was designed should, as a logical matter, reduce the likelihood of achieving intended outcomes.

Impacts

As noted, few studies have examined the effects of supermax prisons on intended goals. The challenges involved in assessing their effects include the fact that many different goals must be considered (and the goals may vary by state); the counterfactual condition frequently is not clear or obvious; experiments are not possible or feasible; and data for quantifying outcomes, intended and unintended, typically are not available. Here, supermaxes are illustrative of causal analysis challenges for many other criminal justice policies. For example, if experimental designs are the standard, we will not accumulate evidence about the causal impact of supermax prisons and thus “evidence” of their effectiveness. The same holds true for many large-scale policy efforts, such as sentencing laws, community policing, or the like (Mears, 2010; Sampson et al., 2013). Fortunately, many different types of quasi-experimental approaches can be undertaken that, under certain conditions, can provide highly credible estimates of policy impact (Heckman and Smith, 1995; Welsh et al., 2013).

Even if experiments were possible with supermax housing, the resulting evidence would not magically lead to an understanding of how supermax housing causes certain outcomes. The same is true of many quasi-experimental designs that one might undertake in assessing supermaxes. For example, if an experiment found that placement in supermax housing reduced misconduct among those placed in it, there would remain the question of *how* the effect was achieved. It might arise from specific deterrence borne of mental suffering or from fear of further social isolation, reduced exposure to criminogenic peer networks, or a moral awakening in line with what the Quakers anticipated would happen to inmates at the Walnut Street jail, or other such possibilities (Mears and Watson, 2006). The causal mechanisms matter because they provide a foundation on which to target changes that might increase beneficial outcomes. For example, supermaxes might increase system-wide prison

order by creating a general deterrent effect among general population inmates, but this effect might be offset by perceptions among such inmates that supermaxes are used inappropriately. In such a context, greater effects might be generated by operating supermaxes appropriately. The relevance here simply is that experimental and quasi-experimental research designs provide only one basis for establishing the causal relationship between a policy and intended impacts and for guiding policy modifications aimed at increasing these impacts and limiting unintended harms.

Cost-Efficiency

Causation is important for understanding whether a given policy or intervention can produce a particular outcome, but in and of itself, it does not constitute “evidence” that it is needed or implemented well, or that it could be implemented well. In addition, it does not provide evidence about whether the given policy or intervention produces the most impact for the least cost. For that, we need causal analyses of other approaches to address the identified problem. It may be, for example, that improved professionalism among administrators and staff might produce larger effects on system-wide order and violence than would supermax prisons, and do so at far lower cost. A focus on cost-efficiency allows us to select among interventions that are evidence based with respect both to research on causation and to research on the financial and human costs associated with them.

Cost-efficiency studies speak in part to causal claims in another way. Specifically, they encourage us to be clear about a range of assumptions and uncertainties that underlie claims about impacts and benefits. Outcomes that cannot be quantified or monetized are so noted. Estimated cost–benefit ratios that are highly vulnerable to assumed or estimated impacts can be highlighted, too. In this way, one can highlight the vulnerability or resilience of a given estimate to the uncertainty of different assumptions, including assumed impacts for a group or area. In this way, the benefits of cost-efficiency evaluations are similar to what Sampson et al. (2013) have described for directed acyclic graphs (DAGs) and what Rossi et al. (2004) have described for theory evaluations. In each instance, a larger context is provided by which to situate specific assumptions and claims. In the best-case scenario, estimated benefits are large and impervious to a range of assumptions and claims; in the worst-case scenario, small changes in assumptions can dramatically change estimates (see, generally, Silver, 2012; Taleb, 2012).

Legal and Ethical Issues

Evidence-based policy ideally includes the selection of interventions where research has documented a causal relationship. Such studies may include experiments or other types of research, such as needs and theory evaluations. Regardless, research on causal impacts cannot directly resolve legal and ethical issues that arise with many criminal justice policies. That insight is not novel but nonetheless too often is omitted from discussions about what it means to have an “evidence-based” intervention. Presumably, we want

policies that not only address critical needs, build on credible theory, are implemented well, achieve intended impacts, and do so at minimal cost, but that also pass legal and ethical muster.

If we turn our attention to supermaxes, many legal and ethical considerations arise. No single one necessarily indicates that we should or should not use supermax housing. Collectively, though, they raise questions about whether, or when, such housing is warranted. For example, the constitutionality of supermaxes has been repeatedly litigated; the main arguments have been that they are cruel and unusual punishment and so violate the Eighth Amendment and that they violate due process rights in violation of the Fourteenth Amendment (Reiter, 2012b). The courts typically have upheld the constitutionality of supermax housing, but their decisions have identified problems. As a review of this issue by Collins (2004) identified, existing case law and litigation point to many “constitutional violations and operational problems” (p. xvi), “suggest that mental health issues will pose the greatest legal challenges” to supermax-like housing (p. xvi), and highlight continued uncertainty about whether such housing “imposes an atypical deprivation on an inmate and therefore requires due process protections” (p. xviii). In addition, “services that are especially critical from a legal perspective—e.g., health care and access to the courts—are difficult to deliver in [extended security units], and use of force is an ever-present issue” (p. xx).

From an ethical perspective, related questions exist. There have been, for example, concerns expressed by scholars and advocacy groups about reliance on a type of incarceration that some allege is inhumane and violates human rights (Bauer, 2012; Browne et al., 2011; Cohen, 2013; Durbin, 2013; Jeffreys, 2013; Katel, 2012; Kurki and Morris, 2001; Lowen and Isaacs, 2012; Rhodes, 2004; Shalev, 2009; Tietz, 2012). Some evidence exists that supermax housing may be used more for minorities than for Whites (Mears and Bales, 2010; Schlanger, 2013). Research has suggested that it may induce mental illness or aggravate existing mental health problems (Haney, 2003; Rhodes, 2004; Shalev, 2009) and that these effects may be greater among the youngest inmates (Human Rights Watch, 2012). It also may cause a range of other harms, including worsening system-wide order and safety (Mears and Watson, 2006). Reliance on supermax incarceration may reduce or preclude investment of scarce resources in a range of other approaches that might more effectively achieve the same goals at less cost (Bottoms, 1999; Mears and Reisig, 2006). And these and related problems or harms all arise in a context in which little oversight or documentation exists about what happens inside supermax facilities (Mears, 2008a, 2008b; Mushlin, 2012).

How exactly each of these possibilities should be weighed and balanced remains unclear. And it is unclear, as well, how causal analyses of supermax impacts would address them. They can, of course, identify whether specific outcomes are achieved. But they cannot resolve debates or concerns about, say, racial disparities in the use of supermaxes. By contrast, empirical research more broadly—including descriptive accounts of the types of inmates housed in supermax facilities, the duration of supermax incarceration,

adherence to protocol, etc.—can be used to inform discussions about the types of concerns mentioned.

Policy and Political Challenges

Evidence-based policy arguably should include consideration of contextual factors, such as the extent of public support, available policy options, and political dynamics. Consideration of such factors does not constitute a substitute for relying on research that documents policy need, impact, or cost-efficiency. Yet, these factors undoubtedly provide a central platform for why certain policies, including supermax prisons, are pursued. For that reason, they should, on the face of it, feature in discussions about the “evidence” base for any given policy. This idea is illustrated here through reference to supermax prisons.

First, policy makers have freely referenced the “public will” when advocating for particular interventions or approaches (Burstein, 1998, 2003) including supermaxes, and so it would make sense to consider the extent of public support for supermax housing. Notably, however, only one study has examined public views about supermax housing even though it is second only to the death penalty in severity. The study found that more than 80% of Florida citizens supported the use of supermax prisons but that this support dropped to 60% if no public safety benefit was anticipated (Mears, Mancini, Beaver, and Gertz, 2013). It also found that 70% of Floridians disagreed that supermaxes are inhumane. The study is notable primarily for highlighting how variable public support can be depending on the characterization of the policy option. In these cases, support dropped 20 percentage points if respondents were asked about a situation wherein supermax prisons had no effect on public safety, a percentage drop similar to what holds in studies of support for the death penalty when respondents have been given a scenario in which life without parole is an option (Cullen, Fisher, and Applegate, 2000). Ultimately, public opinion should not likely constitute the sole basis for public policy. At the same time, it is referenced by policy makers when they advocate for particular policy positions. Accordingly, accurate information about public views toward policies is critical for informing discussions about the need and support for a given policy.

Second, some of the motivation for creating supermax housing stems from the belief that it is a necessary last-resort option when all else has failed. As discussed, making sense of any such claim requires clarity about what problems supermaxes are supposed to solve. If, however, we take as our point of departure the idea that a select set of inmates directly or indirectly contributed disproportionately to prison disorder and violence, a question arises: For any given state that has a supermax, were in fact a range of alternative options attempted prior to turning to supermaxes as a solution? There is little empirical evidence that states systematically pursued any of a wide range of alternatives to supermaxes. Empirical evidence that such avenues were exhaustively pursued would help bolster claims that a last-resort option of some sort was needed.

Third, political considerations seem to be part of the explanation for why supermaxes have been adopted and so would seem to be important to include in an assessment of why they may be needed or appropriate (Sundt, Castellano, and Briggs, 2008). Mears and Watson (2006), for example, found that corrections officials in one state advocated for a new maximum-security prison and were told by legislators that a supermax was a better idea even though the officials did not feel that one was needed. The legislators wanted an approach that would symbolize their tough-on-crime stance. Briggs et al. (2003: 1342) have argued that a related motivation has driven some states to invest in supermaxes: “For many within the prisons industry, the establishment of the supermax is viewed as the sine qua non of a progressive prison regime that is concerned with the safety needs of its inmates and staff.” Being tough on crime or seeking to appear progressive may be important political goals, but they do not reflect a concern with adopting policies that are, on some objective, empirical basis, needed (Browne et al., 2011, 2012; King, 1999; Mears, 2008b).

The “political” dimension of supermax prisons does not necessarily reflect “conservative” versus “liberal” divides. For example, policy makers of either persuasion can and do advocate for prison construction in their home districts (Mears and Watson, 2006). Community leaders and labor unions may weigh in on such decisions as well. In 2012, for example, debate arose in Illinois over the governor’s proposal to close the Tamms supermax facility. The sustained downturn in the economy and resulting adverse effects on the state budget, more so than evidence of a need for a supermax, seemed to be central to the debate. For a similar reason, many states, including Colorado, Illinois, Kansas, Maine, Maryland, Mississippi, New Mexico, Ohio, Texas, and Washington State, have begun to rethink their use of supermax housing (Browne et al., 2012; Chammah, 2012; Garcia and Geurrero, 2013; Goode, 2012; Katel, 2012; Kupers et al., 2009; U.S. Government Accountability Office, 2013). Whether closing facilities makes sense, however, depends on what the levels of need and impact are, and on what benefit they provide as against their costs.

Juxtaposed against these considerations—public views, available policy options, political dynamics—is the fact that states operate with limited funding and have prison systems that confront numerous challenges and problems. Accordingly, the stakes are high. In such a context, enacting policies that may not be needed, designed well, or implemented appropriately, or that achieve intended effects at levels sufficient to offset their costs is problematic. Empirical research on these dimensions thus is critical. The argument here, though, is that an understanding of political context and dynamics arguably should constitute an additional part of a comprehensive approach to describing the “evidence” for a policy.

Policy Options

Evidence for a particular policy—such as a series of experiments that show that the policy produces a given level of impact—does not constitute evidence about whether that policy is

needed, possible to implement, or the most cost-efficient investment in a given context. As noted, proponents of supermax housing have argued that such housing constitutes a “last resort” option, implying that other options were exhausted and failed to work. Setting aside the lack of empirical evidence that such options indeed were tried or that they failed, there is the question of whether effective alternatives to supermax prisons exist.

Answering the question is difficult without knowing the precise problem a given state wanted addressed and, by extension, what exactly is meant by “effective.” Here, let us assume that two problems exist: (a) an increase in system-wide prison disorder and violence and (b) an increase in the number of extremely difficult or violent inmates. Our goals here might be to return to previous levels of disorder and violence and to provide the level of control over the difficult and violent inmates that is typical, as defined by past practice in a given prison system. What approaches might we consider at the outset before investing in supermaxes? One is to examine the many potential factors that contribute to the system-wide disorder and violence. Perhaps greater disorder and violence is occurring at facilities that have hired relatively “green” wardens or officers. If so, then we might want to target our energy in that direction. Perhaps it turns out that a small number of inmates, spread throughout the prison system, is simultaneously causing disorder and violence to escalate. If so, we might target our energy there instead. Here, of course, there remains the question of what to do.

What possibilities exist for reducing inmate disorder and violence? At the most general level, one possibility is to improve administrative practice through organizational efficiency and greater professionalism (DiIulio, 1987; Mears, 2008a; Reisig, 1998; Sparks et al., 1996). Another is to create an administrative and front-line staff culture that leads inmates to perceive prison authority as legitimate; consistent, fair, and reasonable implementation of rules is one way to achieve this perception (Bottoms, 1999). A range of programming measures might be effective as well, including drug and mental health treatment, educational and vocational programs, and, in particular, behavioral programs (French and Gendreau, 1996; Gendreau and Keyes, 2001; Mears and Watson, 2006). There is also the practice, widely used prior to the 1980s, of strategically dispersing certain disruptive or violent inmates throughout the prison system (Briggs et al., 2003; Riveland, 1999).

These ideas do not derive only from research; rather, they also come from practitioners (Mears, 2005). For example, in a national survey of prisons, wardens were asked to identify effective alternatives to supermaxes (Mears and Castro, 2006). Eighty-eight percent said that at least one such alternative—from a list of possibilities—existed and 76% said that at least two did. Half or more of the wardens identified a range of approaches as effective alternatives, including staff training, provision of some type of rehabilitative programming, and using specialized housing other than supermaxes. Many wardens viewed still other approaches as effective alternatives to supermaxes. These included strict enforcement of rules, relying on incentives-based strategies for gaining inmate compliance, and using maximum-security prisons.

There are, in short, a wide range of possible approaches to preventing and reducing disorder and violence in prisons. Indeed, the range suggests that no one approach likely would be sufficient alone and that, rather, the most effective strategy would be to pursue several at once. Have states systematically exhausted the approaches mentioned, or others? That remains unknown. All that we do know is that states invested substantially in what, on the face of it, amounts to a silver bullet solution to many different problems that themselves have a variety of causes.

Research Gaps

Unanswered questions always exist about any given policy, and that is certainly true of supermax prisons and, more broadly, of extended isolation, including housing used for protective custody and punitive segregation (Shalev, 2009). Indeed, scholarly accounts have enumerated a litany of them (e.g., Clare and Bottomley, 2001; King, 1999, 2005; Kurki and Morris, 2001; Mears, 2008b; Mears and Watson, 2006; Naday et al., 2008; J. I. Ross, 2013; Shalev, 2009). At the broadest level, we lack information along the five evaluation dimensions discussed: need, theory, implementation, impact across a range of outcomes, and cost-efficiency. Until these gaps are addressed, it will be difficult to justify opening, continuing, or closing supermax prisons on evidence-based grounds.

A central barrier to creating research that addresses these different dimensions is the lack of research infrastructure in state correctional systems. Research staff typically have too many competing demands on their time, and databases can be unwieldy to use. Even so, many examples of productive research efforts exist that have attempted to exploit available data in a creative way to address key research questions. For example, the groundwork for examining supermax implementation questions has been laid by the Vera Institute's Segregation Reduction Project (SRP, <http://www.vera.org/project/segregation-reduction-project>; see also Browne et al., 2011, 2012; Clare and Bottomley, 2001; Lovell, 2008; Lovell et al., 2000; Mears and Bales, 2010; Ward and Werlich, 2003). This work has identified some of the ways that states can better describe their supermax inmate population, including how appropriate the fit is between the types of inmates placed in supermax housing and who is "ideally" supposed to be placed in it, how long inmates serve in supermax confinement, and what their behavior is like after confinement in supermax housing. This work can help to document appropriate implementation, but it also can be used to assess the effects of supermax housing on misconduct and recidivism and on general population facilities (see, e.g., Briggs et al., 2003; Lovell et al., 2007; Mears and Bales, 2009).

A stronger research infrastructure also can create the groundwork for quasi-experimental evaluations of supermax prison impacts (Mears, 2008a). For example, some wardens may be more aggressive than others in sending inmates to supermaxes (Mears and Castro, 2006). One could compare facility-level rates of infractions at these wardens' facilities with those at comparable facilities. The critical but tractable challenge would be to identify similar facilities that house similar inmates. Other types of studies can be undertaken as well,

including those that employ matching designs and that investigate the effects of varying lengths of time served, or number of placements, in supermax confinement (Lovell et al., 2007; Mears and Bales, 2009).

Sampson et al. (2013) and Causal Uncertainty

Sampson et al. (2013) have identified three challenges related to efforts to improve policy making despite causal uncertainty: (a) effect heterogeneity, which makes it difficult to arrive at generalizable causal claims; (b) ambiguity regarding the causal mechanisms or pathways that contribute to improved outcomes; and (c) contextual factors that may condition whether a policy effect occurs or the magnitude of the effect. What are the implications of these challenges for understanding supermax housing and its impacts?

First, the emphasis on effect heterogeneity, at a general level, underscores the need to think carefully about how a policy creates an impact and, in turn, why it may have different effects for different groups of individuals or areas. As the previous discussion highlights, the theory underlying supermax prisons is poorly articulated, inconsistent, and largely untested. Even so, the very conditions of supermax housing—extended isolation, in particular—suggest that we might well anticipate different effects for different groups. Indeed, a large literature on the potentially adverse effects of such housing on mentally ill inmates reinforces this idea.

That, though, only scratches the surface. What about inmates who experience imprisonment for the first time and who, within weeks of incarceration, are placed in such housing? Other inmates, especially those with a history of incarceration, may have familiarity—whether through first-hand knowledge or through exposure to inmates who have been in supermax confinement—with what extended isolation is like. That, in turn, may help them to cope with such confinement. Similarly, what about inmates with minimal reading ability or low self-control? For such inmates, supermax housing may constitute an especially frustrating experience, which may result in more misconduct and potentially recidivism (Mears and Bales, 2009). What about inmates placed in supermax housing repeatedly, for longer periods of time, or just prior to release back into society (Mears and Bales, 2010)? In each instance, we might well anticipate that supermax effects on behavior would vary. Similarly, at the systems level, supermax housing may produce different effects depending on staffing levels throughout the prison system or on inmates' perceptions of the legitimacy of prison officials and officers.

The very act of focusing on the idea of effect heterogeneity highlights these and many other possibilities. In so doing, it contributes to greater conceptual clarity about policies, including supermax prisons, how exactly they might be effective, and how their effectiveness might vary for certain groups, across different conditions, or in certain contexts. Sampson et al. (2013) have highlighted how statistical approaches to causal modeling allow for investigation of effect heterogeneity. These show promise for providing more robust and accurate estimates of causal effects. Yet, the more fundamental insight they have

offered is that causal uncertainty can be reduced in part by systematically conceptualizing the precise nature of a given policy and, in turn, how and why its effects might vary. Given such conceptualization, it then becomes possible to contemplate suitable research designs and to apply methodologies appropriate to the specific research question at hand. From the previous discussion, it is evident that many different effects of supermax housing can be anticipated and that a wide range of different research studies are needed to identify the diverse effects, intended and unintended and beneficial or harmful, that supermaxes may have. Until such work is undertaken, the effects of supermax prisons—in general, for specific groups or prison systems, or under varying contexts—will remain unknown.

Second, in advocating the use of directed acyclic graphs to assist with identifying causal pathways and factors, Sampson et al. (2013) have echoed the call of researchers to develop causal logic models when evaluating programs or policies (Rossi et al., 2004). The benefits are considerable, and include, not least, the ability to zero in on important policy considerations. For example, in mapping out different ways in which supermax housing might improve system-wide prison order, we can see that many causal mechanisms exist (Mears and Reisig, 2006). Such housing might provide a specific deterrent effect for those placed in the housing, or it might provide a general deterrent effect for general population inmates. It also might “normalize” the prison environment by freeing up officers in general population facilities to monitor inmates better. Each causal pathway ideally should be examined. By doing so, policy makers might be able to target the use of supermax housing in a more effective manner. For example, if supermax housing does contribute to system-wide order and if the effect arises primarily through general deterrence, prison officials might want to employ such housing more frequently but perhaps for less serious violations and for shorter durations. By so doing, there would be the potential for increasing general deterrence among general population inmates.

A recurring theme throughout this essay has been a focus on the lack of clarity about the theoretical mechanisms through which supermax housing might contribute to any of a range of outcomes (e.g., supermax inmate misconduct, general population inmate misconduct, escapes, system-wide prison order, and recidivism) (Mears and Watson, 2006). From a causal claims perspective, then, the situation is far from ideal—few credible empirical studies on the impacts of supermax prisons exist *and* the theoretical or causal logic is poorly developed. By embracing Sampson et al.’s (2013) call for carefully delineating and testing different causal mechanisms, there is a greater possibility to arrive at more definitive claims concerning policy and program impacts and how intended effects might be increased and unintended harms decreased.

Third, Sampson et al. (2013) have argued that causal claims can be difficult to obtain in part because context may matter (see also Welsh et al., 2013). That is, contextual factors may influence the outcomes that arise in one setting versus another. In a similar vein, it may be that a policy can only be implemented in a certain way, or with a certain population, in

a given setting, and this difference in turn may affect the policy's impact. Bardach (2004) has referred to this issue as an extrapolation problem, one that is captured in part by the distinction between efficacy studies (i.e., research on impacts of programs under laboratory-like settings) and effectiveness studies (i.e., research on impacts of programs as applied in real-world settings) (Mears, 2010). In short, what parts of a program or policy can be carried over to a new setting and be expected to produce similar effects to those identified in other settings? And what parts must be adapted, in what ways, to achieve such effects? Context matters for these reasons and still another. As Sampson et al. (2013) have emphasized, a policy may produce an outcome but also result in unintended effects that ripple throughout a given system and in turn create more harm than good.

Such insights are relevant for developing a better understanding of the impacts of supermax housing. Some states may have more capacity to house inmates in supermax housing for longer periods of time, which might affect intended outcomes. Some state prison systems may be managed less professionally than others, and they may engender greater ill will among inmates. In such contexts, inmates may be more likely to view the use of supermax housing as abusive and an unfair demonstration of authority. Accordingly, in such contexts, it may be that supermax housing might produce a specific deterrent effect for inmates placed in supermax housing, but this benefit might be offset by an increase in misconduct among general population inmates. As discussed, it remains unknown what impacts supermax housing has on any of a range of outcomes. By extension, at present we have little basis for knowing what aspects of supermax housing can be adopted in various settings and still achieve a given type and magnitude of effect while avoiding unintended effects on other aspects of prison operations.

Policy Implications and Recommendations

One reason for examining supermax housing—and solitary confinement more broadly—is that it fits broadly within the paradigm of the last 30 years, one that has emphasized “get tough” approaches to crime prevention and, at the same time, government accountability, evidence-based practice, and cost-efficiency. As this policy essay has highlighted, substantial concerns exist about how well supermaxes and other types of extended isolation (e.g., protective custody and punitive segregation)—and many criminal justice policies—fare when held to these standards. One implication, then, is that there should be a better alignment between theory and practice.

How do we improve this alignment? Many possibilities exist and have been enumerated at length over recent decades (see, e.g., Cullen, 2005; Mears, 2010; Petersilia, 1991; Sampson et al., 2013; Sherman, 2004). Improved funding and research infrastructure are critical (Mears, 2008a). Resources alone cannot in and of themselves result in studies that systematically address critical policy questions, but they are essential for such efforts. Stronger ties among universities, research organizations, and criminal justice agencies is essential, especially for investigating the more complicated types of questions that require

understanding of the intricacies of large-scale database systems. Requiring states to undertake needs evaluations and break-even “sensitivity” analyses of high-cost policies, too, would help considerably. Here, again, supermax housing is instructive. In the end, it may well create improvements, but its costs alone raise the bar substantially for the magnitude of benefit required simply to “break even” (Lawrence and Mears, 2004). In some cases, sensitivity analyses can alert us to the possibility that unrealistically large benefits must be assumed to produce benefits that offset policy costs. These and related steps could help improve the evidence base for criminal justice policies (Mears, 2010). At the same time, they provide the foundation for more and better assessments of the causal relationships that advocates hope underlie these policies.

For the time being, states that have supermax housing and other types of extended isolation or solitary confinement face a quandary. Should they keep, expand, or eliminate it? For those who view such confinement as fundamentally inhumane, the only option is to stop employing it. That argument is bolstered in part by some research on supermax incarceration, but it rests ultimately on differences of opinion among various policy makers and members of the public. Even so, it seems difficult to contest that extended solitary confinement constitutes the most extreme form of deprivation currently in use in America and in many other countries. A large body of scholarship, however weak many of the methodological studies may be, attests to the potential for supermax incarceration to create potential harms to inmates and to prison systems as a whole. Trenchant critiques and strongly worded international condemnation of supermaxes also supports the view that extended isolation can reasonably be viewed as inhumane. To the extent that such an assessment holds true, the implication is clear—either eliminate supermax incarceration or raise the bar for the conditions under which it can be used. In legal terms, that would mean subjecting any use of extended isolation to strict scrutiny, requiring strong empirical evidence of a compelling state objective, and ensuring that inmate stays in such isolation occur only for as long as is needed to achieve that objective.

Not everyone agrees with this assessment. Indeed, robust defenses of supermax incarceration abound. In my view, empirical research can provide one critical platform for resolving debates. Here, I see no substitute for the systematic application of the evaluation hierarchy to supermaxes and the types of careful development of causal logic and of causal analysis advocated by Sampson et al. (2013). Doing so will not fully solve philosophical or political differences of opinion, but it can contribute to efforts to place policy debates on a more solid, evidence-based foundation. In short, assess the *need* for current or proposed supermax housing and other forms of solitary confinement; such an assessment requires clarifying the goals of such housing and how each goal should be weighted. Develop the *theoretical or causal logic* through which supermax housing contributes to specific outcomes, and identify the conditions under which such outcomes might reasonably be anticipated. Ensure that *implementation* is sound—that is, take steps to document that inmates in supermax housing reflect the profile of who should be in such housing. Conduct *impact* evaluations

that examine the different outcomes that supermaxes are designed to achieve. Not least, conduct *cost-efficiency* studies that determine whether impacts, or anticipated impacts, offset estimated costs.

What policy guidance might be offered given the state of research on supermax housing? At present, there remains virtually no research that has identified effective “doses” of supermax incarceration—or other forms of extended isolation used to manage inmates—for achieving reductions in misconduct or system-wide order or violence (Mears and Bales, 2010; Mears and Reisig, 2006). For that reason, it would be reasonable to consider a more delimited use of supermax housing; this strategy would accord with a medical approach of relying on the least amount of treatment necessary to achieve some level of health. Here, of course, we do not know the dose necessary to achieve improved outcomes, and most prison systems have long since moved on from “emergency” situations (e.g., riots or sudden increases in inmate murders) that, in some cases, justified a “last resort” option like supermaxes. Accordingly, a small-dose-first approach would seem the more reasonable strategy. A stronger version of this recommendation is to require that, as discussed, states provide solid empirical evidence of a compelling state objective and the precise length of stay necessary to achieve that objective. Given the state of research to date, such a requirement likely would eliminate the use of supermax prisons because of the challenges in undertaking the necessary research to meet this standard.

A strategy of transitioning supermax inmates back into general population facilities prior to release into society and monitoring outcomes during and after supermax incarceration also would constitute a straightforward pragmatic change. Such a strategy would help to address concerns about the potentially adverse effects that supermax housing may have on inmate misconduct (King, Steiner, and Breach, 2008; Sundt et al., 2008; Ward and Werlich, 2003) and successful reentry, especially among inmates released directly from supermax housing back into society (see, e.g., Kupers, 2008; Lovell et al., 2007; Lowen and Isaacs, 2012; Mears and Bales, 2009).

Concerns about the “black box” of supermax housing stem from the fact that little is known about what happens in supermaxes, including the extent to which protocols for cell extractions, services, programming, treatment, and admission and release are appropriately followed (Mears, 2008b; Smith, 2006; U.S. Government Accountability Office, 2013). They stem, too, from the fact that prison settings are dangerous and thus create conditions ripe for abuse (Conover, 2000; Rhodes, 2004; Shalev, 2009). That does not mean that wardens or officers allow or undertake abuse, only that, from an organizational perspective, the conditions are highly conducive to abuse. Thus, improved oversight of supermax housing—and all forms of solitary confinement—coupled with research on its use, would seem to be indicated if such housing is to be used (Ward, 1995).

There also is a need for research that systematically compares the use and effects of various types of segregation. Prison systems use segregation, or isolation, for different periods of time, different reasons, and different inmates. For example, relatively short stints

in segregation may be used to provide punishment or protection to some inmates, while relatively longer stints may be used for managerial purposes (see also Browne et al., 2011: 46; Shalev, 2009: 2). In reality, there may be considerable overlap among these groups of inmates, including the length of time spent in isolation. In addition, inmates placed in “protective custody,” “disciplinary confinement,” or “administrative segregation” in reality may be similar and serve similar periods of time in isolation. The effects may be similar as well. Research that systematically examines this issue could help ensure that the various uses of isolation accord with established protocols. It also could help prison systems better assess the effects of the different types of solitary confinement. And, not least, it could help ground policy debates by providing more credible information about the uses, benefits, and harms of such confinement.

An additional policy implication stands out—a central justification for supermax housing has been the claim that it constituted a “last resort” option. One step for states to take is to explore the veracity of that claim and document whether all other reasonable alternatives—including staff training, strategic dispersion of inmates, behavioral programming—truly have been exhausted. It may not always be clear that these alternatives are or will be effective. However, they typically cost substantially less than the various forms of extended isolation, and in many instances, there in fact may be research to justify them (see, e.g., Gendreau and Keyes, 2001). Even in the absence of research to support them, the argument for employing such alternatives is that they likely would be less financially costly and would raise fewer concerns about harm to inmates or prison systems.

The time is none too soon to place supermax and other types of solitary confinement—and criminal justice policy more generally—on a more evidence-based foundation. As this special issue of *Criminology & Public Policy* highlights, and as Sampson et al.’s (2013) article argues, advances in statistical and causal analysis and database systems provide unique opportunities to improve our understanding of policy impacts. Capitalizing on these opportunities can help contribute evidence-based policy discussions about “what works.” Ultimately, evidence-based policy ideally should include credible empirical research not just on impacts but also on mapping the causal logic that identifies the mechanisms and pathways through which a given policy produces impacts (Sampson et al., 2013). In addition, it should be based on assessment of policy need, implementation, and cost-efficiency. All of these dimensions, as well as vigorous discussion about legal and ethical dimensions of sanctions and correctional system administrative tools, can help contribute to improved policies, programs, and practice.

References

- Adams, Kenneth. 1992. Adjusting to prison life. In (Michael H. Tonry, ed.), *Crime and Justice: A Review of Research*, Vol. 16. Chicago, IL: University of Chicago Press.

- Austin, James and Kenneth McGinnis. 2004. *Classification of High-Risk and Special Management Prisoners: A National Assessment of Current Practices*. Washington, DC: National Institute of Corrections.
- Bardach, Eugene. 2004. Presidential address—The extrapolation problem: How can we learn from the experience of others? *Journal of Policy Analysis and Management*, 23: 205–220.
- Bauer, Shane. 2012. Solitary in Iran nearly broke me. Then I went inside America's prisons. *Mother Jones*. November/December. Retrieved from motherjones.com/politics/2012/10/solitary-confinement-shane-bauer.
- Bidna, Howard. 1975. Effects of increased security on prison violence. *Journal of Criminal Violence*, 3: 33–46.
- Bottoms, Anthony E. 1999. Interpersonal violence and social order in prisons. In (Michael H. Tonry and Joan Petersilia, eds.). *Prisons*. Chicago, IL: University of Chicago Press.
- Briggs, Chad S., Jody L. Sundt, and Thomas C. Castellano. 2003. The effect of supermaximum security prisons on aggregate levels of institutional violence. *Criminology*, 41: 1341–1376.
- Browne, Angela, Susanne Agha, and Jim Austin. 2012. *Vera's Segregation Reduction Project (SRP): An Innovative Approach to Use of Segregation in U.S. Prisons*. Paper presented at the American Society of Criminology annual meeting, Chicago, IL.
- Browne, Angela, Alissa Cambier, and Susanne Agha. 2011. Prisons within prisons: The use of segregation in the United States. *Federal Sentencing Reporter*, 24: 46–49.
- Bruton, James H. 2004. *The Big House: Life Inside a Supermax Security Prison*. Stillwater, MN: Voyageur Press.
- Bulman, Philip, Marie Garcia, and Jolen Hernon. 2012. Study raises questions about psychological effects of solitary confinement. *NIJ Journal*, 269: 4–7.
- Burstein, Paul. 1998. Bringing the public back in: Should sociologists consider the impact of public opinion on public policy? *Social Forces*, 77: 27–62.
- Burstein, Paul. 2003. The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly*, 56: 29–40.
- Butler, H. Daniel, O. Hayden Griffin III, and W. Wesley Johnson. In press. What makes you the “worst of the worst?” An examination of state policies defining supermaximum confinement. *Criminal Justice Policy Review*.
- Carlton, Bree. 2007. *Imprisoning Resistance: Life and Death in an Australian Supermax*. Sydney, Australia: Sydney Institute of Criminology, Federation Press.
- Chammah, Maurice. 2012. Lawmakers revisit approach to solitary confinement. *The Texas Tribune*. September 5. Retrieved from texastribune.org/texas-dept-criminal-justice/texas-department-of-criminal-justice/lawmakers-address-reentry-after-solitary-confinement/.
- Clare, Emma and Keith Bottomley. 2001. *Evaluation of Close Supervision Centres*. London, U.K.: Home Office.
- Cloyes, Kristin G., David Lovell, David G. Allen, and Lorna A. Rhodes. 2006. Assessment of psychosocial impairment in a supermaximum security unit sample. *Criminal Justice and Behavior*, 33: 760–781.

- Cohen, Andrew. 2013. An American gulag: The mentally ill at supermax. *The Atlantic Monthly*. Retrieved from theatlantic.com/personal/archive/2012/06/an-american-gulag-0151-the-mentally-ill-at-supermax/258818.
- Collins, William C. 2004. *Supermax Prisons and the Constitution: Liability Concerns in the Extended Control Unit*. Washington, DC: National Institute of Corrections.
- Conover, Ted. 2000. *Newjack: Guarding Sing Sing*. New York: Random House.
- Crouch, Ben M. and James W. Marquart. 1989. *An Appeal to Justice: Litigated Reform of Texas Prisons*. Austin: University of Texas Press.
- Cullen, Francis T. 2005. The twelve people who saved rehabilitation: How the science of criminology made a difference. *Criminology*, 43: 1–42.
- Cullen, Francis T., Bonnie S. Fisher, and Brandon K. Applegate. 2000. Public opinion about punishment and corrections. *Crime and Justice*, 27: 1–79.
- DeMaio, Jerry R. 2001. If you build it, they will come: The threat of overclassification in Wisconsin's supermax prison. *Wisconsin Law Review*, 2001: 207–248.
- DiIulio, John J., Jr. 1987. *Governing Prisons*. New York: Free Press.
- Durbin, Dick. 2012. *Reassessing Solitary Confinement: The Human Rights, Fiscal and Public Safety Consequences*. Opening Statement of Senator Dick Durbin for the hearing before the United States Senate Committee on the Judiciary Subcommittee on the Constitution, Civil Rights and Human Rights, June 19.
- Durbin, Dick. 2013. Durbin statement on federal bureau of prisons assessment of its solitary confinement practices. Washington, DC: United States Senate. Retrieved from durbin.senate.gov/public/index.cfm/pressreleases?ID=07260483-4972-4720-8d43-8fc82a9909ac.
- Filho, José de Jesus. 2013. Rise of the supermax in Brazil. In (Jeffrey I. Ross, ed.), *The Globalization of Supermax Prisons*. New Brunswick, NJ: Rutgers University Press.
- French, Sheila A. and Paul Gendreau. 2006. Reducing prison misconducts: What works! *Criminal Justice and Behavior* 33: 185–218.
- Gaes, Gerald G., Scott D. Camp, Julianne B. Nelson, and William G. Saylor. 2004. *Measuring Prison Performance: Government Privatization and Accountability*. New York: AtlaMira Press.
- Garcia, Monique and Rafael Guerrero. 2013. Quinn stands by decision to close 2 prisons, despite inmate crowding. *The Chicago Tribune*. February 18, p. 1.
- Gawande, Atul. 2009. Hellhole: The United States holds tens of thousands of inmates in long-term solitary confinement. Is this torture? *The New Yorker*. March 30, pp. 36–45.
- Gendreau, Paul and David Keyes. 2001. Making prisons safer and more humane environments. *Canadian Journal of Criminology*, 43: 123–130.
- Goode, Erica. 2012. Prisons rethink isolation, saving money, lives and sanity. *The New York Times*. March 11, p. A1.
- Haney, Craig. 2003. Mental health issues in long-term solitary and “supermax” confinement. *Crime & Delinquency*, 49: 124–156.
- Heckman, James J. and Jeffrey A. Smith. 1995. Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9: 85–110.

- Human Rights Watch. 2012. *Growing Up Locked Down: Youth in Solitary Confinement in Jails and Prisons across the United States*. Washington, DC: Author.
- Irwin, John. 2005. *The Warehouse Prison: Disposal of the New Dangerous Class*. Los Angeles, CA: Roxbury.
- Jacobson, Michael. 2012. *Senate Testimony, Prepared for the Senate Judiciary Subcommittee on the Constitution, Civil Rights, and Human Rights Hearing on Solitary Confinement*. June 19. New York: Vera Institute.
- Janquart, Philip A. 2013. Pelican Bay slammed for housing groups by race. *Courthouse News Service*, January 29. Retrieved from courthousenews.com/2013/01/29/54368.htm.
- Jeffreys, Derek S. 2013. *Spirituality in Dark Places: The Ethics of Solitary Confinement*. New York: Palgrave Macmillan.
- Jennings, Diane. 2009. Thousands of Texas inmates held in administrative segregation—one for 24 years. *The Dallas News*. July 16. Retrieved from crime-blog.dallasnews.com/2009/07/thousands-of-texas-inmates-hel.html.
- Katel, Peter. 2012. Solitary confinement: Is long-term isolation of prisoners inhumane? *Congressional Quarterly*, 22: 767–788.
- King, Kate, Benjamin Steiner, and Stephanie R. Breach. 2008. Violence in the supermax: A self-fulfilling prophecy. *The Prison Journal*, 88: 144–168.
- King, Roy D. 1999. The rise and rise of supermax: An American solution in search of a problem? *Punishment & Society*, 1: 163–186.
- King, Roy D. 2005. Effects of supermax custody. In (Alison Liebling and Shadd Maruna, eds.), *The Effects of Imprisonment*. Portland, OR: Willan.
- King, Roy D. 2007. Imprisonment: Some international comparisons and the need to revisit panopticism. In (Yvonne Jewkes, ed.), *Handbook of Prisons*. Portland, OR: Willan.
- King, Roy D. and Sandra L. Resodihardjo. 2010. To max or not to max: Dealing with high risk prisoners in the Netherlands and England and Wales. *Punishment & Society*, 12: 65–84.
- Kumar, Anita. 2012. Virginia prisons' use of solitary confinement is scrutinized. *The Washington Post*. January 7. Retrieved from washingtonpost.com/local/dc-politics/va-prisons-use-of-solitary-confinement-is-scrutinized/2011/11/28/gIQakKHuhP_story.html.
- Kupers, Terry A. 2008. What to do with the survivors? Coping with the long-term effects of isolated confinement. *Criminal Justice and Behavior*, 35: 1005–1016.
- Kupers, Terry A., Theresa Dronet, Margaret Winter, James Austin, Lawrence Kelly, William Cartier, Timothy J. Morris, Stephen F. Hanlon, Emmitt L. Sparkman, Parveen Kumar, Leonard C. Vincent, Jim Norris, Kim Nagel, and Jennifer McBride. 2009. Beyond supermax administrative segregation: Mississippi's experience rethinking prison classification and creating alternative mental health programs. *Criminal Justice and Behavior*, 36: 1037–1050.
- Kurki, Leena and Norval Morris. 2001. Supermax prisons. In (Michael H. Tonry, ed.), *Crime and Justice: A Review of Research*. Chicago, IL: University of Chicago Press.

- Lawrence, Sarah and Daniel P. Mears. 2004. *Benefit-Cost Analysis of Supermax Prisons: Critical Steps and Considerations*. Washington, DC: The Urban Institute.
- Logan, Charles. 1993. Criminal justice performance measures for prisons. In (BJS-Prison Project, ed.), *Performance Measures for the Criminal Justice System*. Washington, DC: Bureau of Justice Statistics.
- Lovell, David. 2008. Patterns of disturbed behavior in a supermax population. *Criminal Justice and Behavior*, 35: 985–1004.
- Lovell, David and L. Clark Johnson. 2004. *Felony and Violent Recidivism among Supermax Prison Inmates in Washington State: A Pilot Study*. Seattle: University of Washington Press.
- Lovell, David, Kristin Cloyes, David Allen, and Lorna Rhodes. 2000. Who lives in supermaximum custody? *Federal Probation*, 64: 33–38.
- Lovell, David, L. Clark Johnson, and Kevin C. Cain. 2007. Recidivism of supermax prisoners in Washington State. *Crime & Delinquency*, 53: 633–656.
- Lowen, Matthew and Caroline Isaacs. 2012. *Lifetime Lockdown: How Isolation Conditions Impact Prisoner Reentry*. Tucson, AZ: American Friends Service Committee.
- Marini, Margaret M. and Burton Singer. 1988. Causality in the social sciences. In (Clifford C. Clogg, ed.), *Sociological Methodology*. Washington, DC: American Sociological Association.
- Mears, Daniel P. 2004. Prisoner abuse is avoidable. *The Hill*. June 16.
- Mears, Daniel P. 2005. A critical look at supermax prisons. *Corrections Compendium*, 30: 6–7, 45–49.
- Mears, Daniel P. 2006. *Evaluating the Effectiveness of Supermax Prisons*. Washington, DC: The Urban Institute.
- Mears, Daniel P. 2008a. Accountability, efficiency, and effectiveness in corrections: Shining a light on the black box of prison systems. *Criminology & Public Policy*, 7: 143–152.
- Mears, Daniel P. 2008b. An assessment of supermax prisons using an evaluation research framework. *The Prison Journal*, 88: 43–68.
- Mears, Daniel P. 2010. *American Criminal Justice Policy: An Evaluation Approach to Increasing Accountability and Effectiveness*. New York: Cambridge University Press.
- Mears, Daniel P. and William D. Bales. 2009. Supermax incarceration and recidivism. *Criminology*, 47: 801–836.
- Mears, Daniel P. and William D. Bales. 2010. Supermax housing: Placement, duration, and time to reentry. *Journal of Criminal Justice*, 38: 545–554.
- Mears, Daniel P. and Jennifer L. Castro. 2006. Wardens' views on the wisdom of supermax prisons. *Crime & Delinquency*, 52: 398–431.
- Mears, Daniel P., Joshua C. Cochran, Sarah J. Greenman, Avinash S. Bhati, and Mark A. Greenwald. 2011. Evidence on the effectiveness of juvenile court sanctions. *Journal of Criminal Justice*, 39: 509–520.
- Mears, Daniel P., Christina Mancini, Kevin Beaver, and Marc Gertz. 2013. Housing for the “worst of the worst” inmates: Public support for supermax prisons. *Crime & Delinquency*, 59: 587–615.

- Mears, Daniel P. and Michael D. Reisig. 2006. The theory and practice of supermax prisons. *Punishment & Society*, 8: 33–57.
- Mears, Daniel P. and Jamie Watson. 2006. Towards a fair and balanced assessment of supermax prisons. *Justice Quarterly*, 23: 232–270.
- Mushlin, Michael B. 2012. Unlocking the courthouse door: Removing the barrier of the PLRA's physical injury requirement to permit meaningful judicial oversight of abuses in supermax prisons and isolation units. *Federal Sentencing Reporter*, 24: 268–275.
- Naday, Alexandra, Joshua D. Freilich, and Jeff Mellow. 2008. The elusive data on supermax confinement. *The Prison Journal*, 88: 69–93.
- National Institute of Corrections. 1997. *Supermax Housing: A Survey of Current Practice*. Washington, DC: National Institute of Corrections.
- National Public Radio. 2012. The grim realities of life in supermax prisons. *Talk of the Nation*. Hosted by Neal Conan, June 21. Washington, DC: National Public Radio.
- Neal, Donice, ed. 2003. *Supermax Prisons: Beyond the Rock*. Lanham, MD: American Correctional Association.
- O'Keefe, Maureen L., Kelli J. Klebe, Jeffrey Metzner, Joel Dvoskin, Jamie Fellner, and Alysha Stucker. 2013. A longitudinal study of administrative segregation. *Journal of the American Academy of Psychiatry and the Law*, 41: 49–60.
- Petersilia, Joan. 1991. Policy relevance and the future of criminology. *Criminology*, 29: 1–16.
- Pizarro, Jesenia and Vanja M. K. Stenius. 2004. Supermax prisons: Their rise, current practices, and effect on inmates. *The Prison Journal*, 84: 248–264.
- Pizarro, Jesenia, Vanja M. K. Stenius, and Travis C. Pratt. 2006. Supermax prisons: Myths, realities, and the politics of punishment in American society. *Criminal Justice Policy Review*, 17: 6–21.
- Reisig, Michael D. 1998. Rates of disorder in higher-custody state prisons: A comparative analysis of managerial practices. *Crime & Delinquency*, 44: 229–244.
- Reiter, Keramet A. 2012a. Parole, snitch, or die: California's supermax prisons and prisoners, 1997–2007. *Punishment & Society*, 14: 530–563.
- Reiter, Keramet A. 2012b. The most restrictive alternative: A litigation history of solitary confinement in U.S. prisons, 1960–2006. In (Austin Sarat, ed.), *Studies in Law, Politics, and Society*. London, U.K.: Emerald Group.
- Rhodes, Lorna A. 2004. *Total Confinement: Madness and Reason in the Maximum Security Prison*. Los Angeles: University of California Press.
- Riveland, Chase. 1999. *Supermax Prisons: Overview and General Considerations*. Washington, DC: National Institute of Corrections.
- Ross, Jeffrey I., ed. 2013. *The Globalization of Supermax Prisons*. New Brunswick, NJ: Rutgers University Press.
- Ross, Jamie. 2013. Civil rights groups fight solitary confinement. *Courthouse News Service*. April 4. Retrieved from courthousenews.com/2013/04/04/56357.htm.
- Rossi, Peter H., Howard E. Freeman, and Mark W. Lipsey. 2004. *Evaluation: A Systematic Approach*, 7th Edition. Thousand Oaks, CA: Sage.

- Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. Translating causal claims: Principles and strategies for policy-relevant criminology. *Criminology & Public Policy*, 12: 587–616.
- Schlanger, Margo. 2013. Prison segregation: Symposium introduction and preliminary data on racial disparities. *Michigan Journal of Race and Law*, 18: 1–12.
- Shalev, Sharon. 2009. *Supermax: Controlling Risk through Solitary Confinement*. London, U.K.: Willan.
- Sherman, Lawrence W. 2003. Misleading evidence and evidence-led policy: Making social science more experimental. *The ANNALS of the American Academy of Political and Social Science*, 589: 6–19.
- Sherman, Lawrence W. 2004. Research and policing: The infrastructure and political economy of federal funding. *The ANNALS of the American Academy of Political and Social Science*, 593: 156–178.
- Silver, Nate. 2012. *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*. New York: Penguin.
- Singer, Paul. 2003. Warden: Inmates volunteer to stay at “supermax” prison. *Associated Press*. March 3. Retrieved from wkyc.com/news/story.aspx?storyid=3330.
- Smith, Peter S. 2006. The effects of solitary confinement on prison inmates: A brief history and review of the literature. *Crime and Justice*, 34: 441–528.
- Smith, Peter S. 2008. “Degenerate criminals”: Mental health and psychiatric studies of Danish prisoners in solitary confinement, 1870–1920. *Criminal Justice and Behavior*, 35: 1048–1064.
- Sparks, Richard, Anthony E. Bottoms, and Will Hay. 1996. *Prisons and the Problem of Order*. Oxford, U.K.: Oxford University Press.
- Stephan, James J. 2008. *National Prisoner Statistics Program: Census of State and Federal Correctional Facilities, 2005*. Washington, DC: Bureau of Justice Statistics.
- Sullivan, Laura. 2006. *Life in Solitary Confinement*. Three-part series, July 26–28. Washington, DC: National Public Radio.
- Sundt, Jody L., Thomas C. Castellano, and Chad S. Briggs. 2008. The sociopolitical context of prison violence and its control: A case study of supermax and its effect in Illinois. *The Prison Journal*, 88: 94–122.
- Taleb, Nassim N. 2012. *Antifragile: Things that Gain from Disorder*. New York: Random House.
- Tietz, Jeff. 2012. Slow-motion torture. *Rolling Stone*. December 6, pp. 58–66.
- Toch, Hans. 2001. The future of supermax confinement. *The Prison Journal*, 81: 376–388.
- Toch, Hans. 2003. The contemporary relevance of early experiments with supermax reform. *The Prison Journal*, 83: 221–228.
- U.S. Government Accountability Office. 2013. *Improvements Needed in Bureau of Prisons' Monitoring and Evaluation of Segregated Housing*. Washington, DC: Author.
- Ward, David A. 1995. A corrections dilemma: How to evaluate supermax regimes. *Corrections Today*, 57: 104–108.

- Ward, David A. and Thomas G. Werlich. 2003. Alcatraz and Marion: Evaluating super-maximum custody. *Punishment & Society*, 5: 53–75.
- Wells, Terry L., W. Wesley Johnson, and Rodney J. Henningsen. 2002. Attitudes of prison wardens toward administrative segregation and supermax prisons. In (Leanne F. Alarid and Paul F. Cromwell, eds.), *Correctional Perspectives*. Los Angeles, CA: Roxbury.
- Welsh, Brandon C., Anthony R. Braga, and Gerben J. N. Bruinsma, eds. 2013. *Experimental Criminology: Prospects for Advance Science and Public Policy*. New York: Cambridge University Press.
- Welsh, Wayne N. and Philip W. Harris. 2008. *Criminal Justice Policy and Planning*, 3rd Edition. Dayton, OH: LexisNexis, Anderson.
-

Daniel P. Mears, Ph.D., is the Mark C. Stafford Professor of Criminology at Florida State University's College of Criminology and Criminal Justice. He conducts basic and applied research on crime and justice. His work has appeared in *Criminology*, the *Journal of Research in Crime and Delinquency*, *Justice Quarterly*, and other crime and policy journals and in the book *American Criminal Justice Policy* (Cambridge University Press), which won the Outstanding Book Award from the Academy of Criminal Justice Sciences.

AFTERWORD

CRIMINOLOGY, CAUSALITY, AND PUBLIC POLICY

Linking Evidence and Criminal Justice Policy

Alfred Blumstein

Carnegie Mellon University

Introduction

We have seen widespread acceptance of the concept of “evidence-based policy,” even in the U.S. Department of Justice, with strong reinforcement by the U.S. Office of Management and Budget (2012). Thus, it is most appropriate for *Criminology and Public Policy* (CPP) to have organized this special issue with initiative from and a lead paper by outgoing editor, Tom Blomberg. We have also seen careful analysis by Sampson, Winship, and Knight (2013, this issue) of the many possible subtle interpretations of the simple phrase “X causes Y” and the many ways in which its simplest demonstration could be misleading and the necessity for attention to dealing with the more complex aspects. Then we have seen three papers examining carefully the evidence in three very different domains, both substantively and in terms of the methods of proving the causal links:

- Crime prevention with a focus on various forms of early-age home visitation (Fagan, 2013, this issue) where there has been a considerable amount of careful research consistently showing not only a clear effect, but a cost-beneficial effect, but only with limited implementation, at least in part because expenditures in those areas would lead to results on “someone else’s watch,” thus highlighting the complexity inherent in policy decision making
- Supermax prisons (Mears, 2013, this issue), where there is an inherent tension between some presumptions of greater effectiveness in prison control and the concern over psychological damage to the individuals assigned to such forms of solitary confinement, but there has been very little strong research addressing either of these issues, although a

Direct correspondence to Alfred Blumstein, Heinz College, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: ab0q@andrew.cmu.edu).

careful reading of the very balanced presentation in the paper could well lead the policy maker in the author's intended direction

- Management of police operations (Nagin and Weisburd, 2013, this issue), examining the attention to "hot spots," where the major evidentiary insight was derived from earlier studies highlighting the very small geographic areas in which crime was concentrated and followed by a variety of randomized experiments that exploited that concentration by showing crime reduction in the more intensively patrolled hotspots and with no nearby displacement of the crime. In addition, they present four quasi-experimental studies that showed how very different changes in police numbers or operational policy could affect crime rates. The studies examined: the use of breathalyzers to reduce drunk driving; increase of police presence on the Washington, DC, Mall; restriction of police profiling leading to an increase in auto theft rates; and the presence of an automatic location-detector in cars to reduce auto-theft rates.

All of these studies highlight the fact that linking evidence to policy is a particularly complex task in a criminal justice context. In the medical profession, evidence and its quality are strongly ingrained as the only basis for introducing medical-policy change, and every practitioner and especially the physicians are inculcated with that perspective. In criminal justice, highest professional standard is that of lawyers, and their professional training is much more focused on skills in advocacy than in science; even at the judicial level, the standard for decision making is based on consistency with law and the Constitution rather than on scientific standards. For the other practitioners in the agencies of the criminal justice system, scientific standards are seen as relevant, but by no means dominant.

Many of the policy choices made within the criminal justice system are made in legislatures, where ideology or political appeal is particularly salient. In the field of medicine, the adherence to scientific findings has resulted from centuries of tradition, and so a large knowledge base has been developed to be drawn upon, and much of biological knowledge is reasonably stationary over time and reasonably homogeneous across population groups. Of course, new knowledge emerges, particularly from new findings in genetics and neuroscience, and those could well challenge existing practice. And policy choices such as the identification of new tests and the frequency of testing are always subject to challenge based on new evidence on their effectiveness and costs.

In contrast, political appeal can vary significantly over time and concerns about criminal justice policy will certainly vary with the public's feeling of safety and that can be strongly influenced by a particularly heinous crime, and that can affect people's concerns and their demands on the political system for response. The U.S. Department of Justice is one of the few cabinet departments that has never had an assistant secretary for science and technology, and it has only recently made a step in that direction by creating a Science Advisory Board for its Office of Justice Programs (Office of Justice Programs, 2010). That board at its first

meeting was charged by the Attorney General to “insert science into the DNA of the Justice Department.”

The issue is made even more complex by the astonishing array of factors that could affect crime over which criminal justice practitioners have no control. These factors include demographic shifts, drug markets, social and economic disadvantage, family structure, gun availability, education, cognitive effects of lead, changing legislation and court decisions, and so many more. Economic performance is a much less complex phenomenon with very rich data available, but there is great concern about the inability to make accurate economic forecasts. The factors impinging on crime phenomena make it so much more difficult. One could try analyses of various sorts to control for this array of exotic causal factors that could affect crime, but isolating and measuring the effect of a particular policy change become so much more difficult.

The base of research funding in different areas also provides an indication of the support provided to gain effective measurement of policy effects. For medical matters, the National Institutes of Health has a total annual budget of about \$30 billion and the Bureau of Labor Statistics has a budget of about \$600 million just for measuring labor markets. In sharp contrast, the National Institute of Justice and the Bureau of Justice Statistics have a combined budget of about \$100 million to cover the entire complex area of criminal justice.

Another important feature of virtually all criminal justice policies is their multidimensional criteria. In most studies, crime reduction is an important criterion, but rarely alone. Almost always, there are cost considerations that must be weighed against the crime reduction. Very often, Constitutional issues, particularly involving the Second, Fourth, Fifth, Sixth, Eighth, and Fourteenth Amendments, can intrude and limit potentially important aspects of a particular treatment. These bring in considerations of intrusiveness or privacy reduction that could limit the shaping of a treatment. Also, public acceptability and public support will certainly affect a policy maker's choices.

One of the important features of crime control policies highlighted by Sampson et al. (2013) is the heterogeneity of response to any treatment. That complicates estimation of any effect size, but is particularly troublesome when an effect may be positive for some individuals and negative for others. In trying to estimate the effect of incarceration on incarcerated offenders, most such measurements show small effect sizes. Nieuwbeerta, Nagin, and Blokland (2009), for example, have shown a small positive criminogenic effect of incarceration using Dutch data, but Nagin and Snodgrass (2013) have shown no effect using Pennsylvania data. It is entirely possible that some individuals may benefit from the special-deterrence effect, while others respond to the criminogenic effect of absorbing pro-crime values, skills, and networks. If one could identify independently those who gain and those who lose, one might be able to emerge with much stronger effect sizes and more appropriate assignments to treatment.

Another critical issue raised by Sampson et al. (2013) is that of context. In a randomized design in a particular setting, the treatment and control groups would have similar

context, and that would strengthen the internal validity of any observations. But that still leaves open the concern about external validity, or the applicability of the results in a different setting. In the absence of a randomized design, then the treatment and the control populations could well be different in their amenability to the treatment for reasons of selection by the treater or self-selection by the treated.

Research Approaches

The general formulation of the evidence problem is that a treatment, T , is applied to some population, P , in some context, C , and some effect, E , is observed. The challenge is finding a way to confidently attribute the effect, E , to the treatment. There are many ways in which that might be done, and they will all differ in their costs and in the degree of confidence a policy maker would have in them. The reputed “gold standard” applied to this task is the randomized controlled trial (RCT) where the treatment is administered to one treatment group and not to another control group, where the members of the treatment and control groups are randomly selected from a population of eligibles. Treatments applied to individuals are ideally suited to such an approach, whereas in so many other realms of treatment in criminal justice operations, randomization is not feasible or at least prohibited by the ethics of the organization, and particularly so in courts, where so many policy choices are located. We can examine here the various approaches for measuring treatment effects.

Randomized Control Trials

Initially, in criminal justice, virtually all RCTs were based on some form of rehabilitative treatment of prisoners. Beginning in the mid-1980s, Lawrence Sherman began to introduce some creative randomized control designs targeted at policing operations. A patrol experiment was initially tested with more intensive patrol in some randomly selected police beats and routine patrol in the others. Contaminating the design, however, was the fact that police cars could be directed to calls outside their assigned beat and in doing so could well cross over other beats. There followed a wide variety of creative experimental designs, probably a majority of which were carried out or stimulated by Sherman.

These problems highlight the limitations of RCTs as the gold standard. In medicine, randomly prescribing a pill that is either a drug or a placebo is fairly easily executed and easily measured in a double-blind mode. Even in such a simple experiment, distortions can result from a patient's failure to follow the protocol because side effects from the treatment can cause some to drop out early. When one translates this model to operations in the criminal justice system, a much broader variety of complications arise. Observing something about the context could affect a police officer's choice to violate the random assignment. The treatment could be observed and undertaken by a control-group member. Interactions between one or more of the many contextual variables and the treatment could distort the outcome. Different officers could display very different skills in their execution of the treatments.

One of the fundamental principles of science is the essential need for replication of any findings. Given all the complications highlighted by Sampson et al. (2013), this need is of special concern in linking criminology research to policy. The case of the Minneapolis domestic-assault experiment (Sherman and Berk, 1984) is of particular interest here. The initial findings suggested that arrest was more effective than counseling in reducing recidivism. The results and its policy implications were reported in the *New York Times* science section (Boffey, 1983) a full year before the paper was published. Following that publication, a number of police departments adopted a policy of arrest, and that policy has even been enacted into statute in some places. NIJ, which supported the Minneapolis experiment, had the good sense to replicate the same experiment in multiple sites, and the findings from the replications were split about evenly in supporting or contradicting the original finding. It was later learned that an important covariate affecting the original finding was whether the offending spouse was employed or not. Thus, even though experiments have been touted as a “gold standard” because of their internal validity, it becomes all the more important that any single experiment be replicated before moving directly to policy because of the possibility of flaws in the design or execution of the experiment. Even experiments with pills and placebos require replication with multiple investigators and evaluators, and any experiment involving criminal justice operations has so many more complexities than a pill.

These issues highlight the attractiveness of RCTs, but they also must raise yellow flags in recognition of the distortions that can creep in. A presumption that one has found a true causal effect because of randomization must be dealt with cautiously. And the fact that mounting an RCT is usually much more expensive than analysis of observational data requires that there be sufficient analysis beforehand to warrant confidence that the RCT will be sufficiently close to the hoped-for gold standard.

Quasi-Experimental Trials

In the many situations when randomization is not possible, there can be opportunities for quasi-experimental analysis that can accumulate to provide strong effect estimates. One such approach includes the regression-discontinuity design where one has a stream of outcome observations, both before and after a treatment is introduced, and one could attribute a shift in the pattern of that data stream to the treatment. Of course, other exogenous changes could occur at about the same time, and could well confound any treatment-effect estimate. Also, many treatments are introduced over time rather than at a discreet point in time, and that could distribute the discontinuity over that introduction period.

One can also explore the differences in the outcome measures in places or populations where the treatment is introduced compared to those where it is not. Since there is no randomization in the assignment of the treatment, then any factors associated with the intentional or even unintentional selection of the treatment targets would be confounded with the treatment effect.

Statistical Analysis

Perhaps the most common estimation of treatment effects is through regression analysis using a model that enumerates the possible factors that could contribute to an outcome measure, augmented by a dummy variable indicating whether a particular observation was subject to the treatment. If the treatment could vary in intensity, that dummy variable would be replaced by a numeric variable indicating the magnitude of the treatment. The quality of the treatment effect derived thereby would be sensitive to factors such as the fullness of the covariates incorporated in the model, the possibility of interactions between those covariates and especially with the treatment that were not included in the model, and the degree to which the linear formulation would be corrupted by nonlinear effects.

Another particularly interesting approach for dealing with the selection biases one finds in a treatment/nontreatment setting without randomized assignment is through propensity-score matching of members of the two groups. Generating a propensity score that characterizes the attributes that distinguish the two groups then provides a basis for creating matched pairs from each of the treatment and nontreatment groups.

Strengthening the Link between Evidence and Policy

It is clear that estimating treatment effects in the criminal justice world is a rather complex and difficult process, but an extremely important one in an environment of such salient public concern. While there are many claims of “evidence,” many of those are based simply on some form of bivariate correlation between one’s favorite policy and a presumed effect in reducing crime. With a desire for invoking stronger scientific evidence, this issue of *CPP* has provided some good examples of how that could be done, even when the research may be short of what one would like.

It is clear that the strongest evidence can come from a randomized controlled trial, but it is also clear that there can be many contaminants of the results of such trials. That argues for replication in different settings. It also argues for replication by practitioners, other than the highly committed and probably highly innovative initiators of a policy. It also argues for observers recording the data of both the treatment application and the outcomes who are disinterested in the outcome, and therefore different from the implementers. All of this can be rather expensive, and so one wants to be sure that one has evidence of the successful performance of a policy initiative that may be less carefully controlled than the RCT, but has been carefully reviewed by those knowledgeable in the methodology before embarking on a more elaborate and expensive RCT.

In the more common situation where an RCT is not feasible or in preparation for one, there is a wide variety of opportunities for observational analysis. In those cases, it is particularly important that the models used for the analysis be robust in the sense that many variable combinations are tried and challenged to ensure strong results. Even if experiments are possible, it is particularly important that they be preceded by other analytic approaches

involving observational opportunities to identify the most appropriate structure for the experimental design and to ensure the fidelity of the experiment.

If the most appropriate approach turns out to be statistical analysis, it becomes important to ensure that as many of the contextual variables influencing the measured effect as possible are incorporated into the statistical model. Any such effort is inherently limited by those variables that can be measured, but perhaps reasonable proxies can be found. Again, it would be desirable to carry out the analysis based on observations in a relevant policy-maker's domain (e.g., within a mayor's city, a corrections commissioner's prison), but if the project is being funded with a federal grant, it would also be desirable to test the robustness by collecting similar observations in another setting.

It would be desirable that the replication of a particular treatment not have the benefit of its originator, who is motivated to make it succeed and has the insights to make fixes along the way. As the treatment is moved into a more routine setting with run-of-the-mill practitioners, their implementation will be less driven and less creative, and so the measured effects are likely to be more representative of the implementation in routine practice.

In statistical analysis we have all been taught to look for a "statistically significant difference" before declaring an implementation successful. Far less frequently emphasized is the sensitivity of any such result to the power of the statistical test, which is driven by sample size. Unfortunately, the term "significant" had been captured by the statisticians, and the term "discernible" would be more appropriate. Thus, it is possible that a test with a small sample size (limited by population or opportunities available) could have a p value greater than .05 that is "significant" (even if not statistically so) and vice versa.

All of this argues that when policy innovations are initiated, it would be very desirable to provide an opportunity to collect data measuring the treatment and its multiple effects as well as variables characterizing the context of the treatment environment and the population being affected. Then, with those data, an independent analyst, perhaps from the local university, could perform a basic evaluation, deriving some preliminary estimates of the effect. That would work even better if the analyst were recruited early in the shaping of the project to incorporate the evaluative perspectives into the data-collection design. It might even be desirable to provide funding for the analyst and follow-up data collection for some period after completion of the treatment. Inevitably, after the evaluation is completed, the treater, the analyst, or some independent third parties will identify shortcomings in the evaluation and call for data that should have been collected but weren't. If the initial evaluation appears promising, then a replication could be attempted, either in that same setting or elsewhere, and incorporating the issues identified in the earlier trial. It is only after a reasonable sequence of such treatment evaluations that one should move on to a more elaborate and more expensive and more carefully controlled experiment, preferably an RCT if possible. One can be more comfortable that the right factors are being measured and controlled for when the experiment follows this sequence of careful trials with promising results.

Conclusion

It should be clear that research in criminology is, and should be, an important contributor to the development of policy regarding crime control and the effective operation of criminal justice agencies. The papers in *CPP*—and those in this issue in particular—highlight the many ways in which that can be accomplished.

There are two major streams of criminological research that will affect policy in different ways. The more common one that is designed to affect policy is evaluation, which typically involves subjecting a particular policy change or innovation, usually in an experimental or quasi-experimental mode with particular attention to measuring an effect size. That effect is most often related to crime, recidivism, or other measures of social benefit associated with the innovation. The key issues here involve adequate control for other related factors contributing to the effect, the internal validity of the experiment, and the transferability of the findings beyond the original setting. There are a large number of evaluations of criminal justice treatments and the better ones are collected in places such as the Blueprints program at the University of Colorado, the Campbell Collaboration, and the federal website CrimeSolutions.gov. Each of these aggregates evaluations and filters out the strongest based on consideration such as the quality of the experimental findings, the replications in different settings, and the magnitude of the effect sizes.

The other realm of criminology is the search for new theories and tests of them, new perspectives for thinking about the issues of concern. New methodologies or analytic perspectives are particularly helpful because they provide an opportunity to organize traditional data series in new and different ways to see if they generate new insights. For example the development of trajectory analysis by Nagin (2005) has led to a variety of re-analyses of many longitudinal datasets. The insights documented by Sherman, Gartin, and Buerger (1989) of the concentration of criminal activity in small segments of a city led to a variety of evaluations of “hot spot” policing, which has become a widespread practice. Here, it is most appropriate to highlight the preface to Hamming’s (1962) classic text on numerical methods “the purpose of computing is insight, not numbers.”

The research and the evaluations get done with a wide variety of methodologies, none of which is inherently superior to any of the others. Given the complexity of linking particular research findings to public policy, the approaches of this journal, with comments on particular research papers provided by other criminologists as well as individuals bringing the perspective of a public policy decision maker allows one to address the various facets of the research findings and how they might be translated into policy. The responders may question some aspects of validity of the findings, the generalizability of the findings, complications or cautions in implementing the policy, and highlight concerns about value dimensions that might complicate the implementation.

There exist many opportunities for criminologists to become sensitized to the complexities of the policy process. Many criminal justice agencies seek to recruit criminologists

as short-time fellows or as regular staff members in order to benefit from the insights and methodological skills they bring to the operation of the agency. Criminologists with that experience can become sensitive to the many complex factors that enter a policy implementation decision, and that could enrich their research and enhance its likelihood of effective implementation.

An important institutional arrangement for sorting through conflicting interpretations of research findings and pointing directions for appropriate policy application is the Committee on Law and Justice of the National Research Council. That committee can organize a panel representing diverse perspectives, which would then be assigned to review the research in a particular field, to identify aspects that warrant further investigation, and can provide guidance on the implementation to improve policy. This provides an important vehicle for linking criminology to the policy process.

References

- Boffey, Philip M. 1983. Domestic violence: Study favors arrest. *New York Times*, April 5.
- Fagan, Abigail A. 2013. Family-focused interventions to prevent juvenile delinquency: A case where science and policy can find common ground. *Criminology & Public Policy*, 12: 617–650.
- Hamming, Richard W. 1962. *Numerical Methods for Scientists and Engineers*. New York: McGraw-Hill.
- Mears, Daniel P. 2013. Supermax prisons: The policy and the evidence. *Criminology & Public Policy*, 12: 681–719.
- Nagin, Daniel S. 2005. *Group-based Modeling of Development*. Cambridge, MA: Harvard University Press.
- Nagin, Daniel, and G. Matthew Snodgrass. 2013. The effect of incarceration on re-offending: Evidence from a natural experiment in Pennsylvania. *Journal of Quantitative Criminology*, 29. doi:10.1007/s10940-012-9191-9.
- Nagin, Daniel S. and David Weisburd. 2013. Evidence and public policy: The example of evaluation research in policing. *Criminology & Public Policy*, 12: 651–679.
- Nieuwebeerta, Paul, Daniel S. Nagin, and Arjan Blokland. 2009. The relationship between first imprisonment and criminal career development: A matched samples comparison. *Journal of Quantitative Criminology*, 25: 227–257.
- Office of Management and Budget, Executive Office of the President. 2012. Memorandum to the heads of executive departments and agencies from Zients, Jeffrey D., Acting Director. Use of evidence, and evaluation in the 2014 budget. May 18.
- Office of Justice Programs, U.S. Department of Justice. 2010. Attorney General Holder Names 18 Experts to New Science Advisory Board. November 23.
- Sampson, Robert J., Christopher Winship, and Carly Knight. 2013. Translating causal claims: Principles and strategies for policy-relevant criminology. *Criminology & Public Policy*, 12: 587–616.

Sherman, Lawrence W., and Richard A. Berk. 1984. The specific deterrent effects of arrest for domestic assault. *American Sociological Review*, 49: 261–272.

Sherman, Lawrence W., Patrick Gartin, and Michael Buerger. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27: 27–55.

Alfred Blumstein is a university professor and the J. Erik Jonsson Professor of Urban Systems and Operations Research and former Dean (from 1986 to 1993) at the H. John Heinz III College of Carnegie Mellon University. He has had extensive experience in both research and policy with the criminal justice system since serving the President's Commission on Law Enforcement and Administration of Justice in 1966–67 as Director of its Task Force on Science and Technology. He has chaired NAS panels on research on deterrent and incapacitative effects, on sentencing research, and on research on criminal careers. Professor Blumstein served from 1979 to 1990 as Chairman of the Pennsylvania Commission on Crime and Delinquency, the state's criminal justice planning agency. He served on the Pennsylvania Commission on Sentencing from 1986–96. He was the recipient of the 2007 Stockholm Prize in Criminology. Professor Blumstein is a Fellow of the American Society of Criminology, was the 1987 recipient of the Society's Sutherland Award for "contributions to research," and was the president of the Society in 1991–92. His research over the past 20 years has covered many aspects of criminal-justice phenomena and policy, including crime measurement, criminal careers, sentencing, deterrence and incapacitation, prison populations, demographic trends, juvenile violence, and drug-enforcement policy.

