CONTENTS

VOLUME 10 • ISSUE 4 • NOVEMBER 2011

MONTANA EARLY RELEASE PROGRAM

TRANSITIONAL JOBS PROGRAM

COMMUNITY-DRIVEN VIOLENCE REDUCTION PROGRAMS

RACIAL DISPARITY IN WAKE OF THE *BOOKER/FANFAN* DECISION

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. Please note that Word 2007 is not yet compatible with journal production systems. Unfortunately, the journal cannot accept Microsoft Word 2007 documents until such time as a stable production version is released. Please use Word's "Save As" option, therefore, to save your document as an older (.doc) file type. 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, 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 2011 are: US$291 (U.S.), US$358 (rest of world), €229 (Europe), £183 (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 visit www.wileyonlinelibrary.com/tax-vat. The price includes on-line access to the current and all on-line back issues to January 1, 2007 (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 Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavor 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.

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. Bi Xian Tan, Production Editor (e-mail: bxtan@wiley.com); Kristin McCarthy, Advertising (e-mail: kmccarthy@wiley.com).

Printed in the USA by The Sheridan Group.

MONTANA EARLY RELEASE PROGRAM

Good in theory

The challenges of early release decisions

Beth M. Huebner
University of Missouri, St. Louis

The rise in incarceration during the past two decades has been well documented (Pew Center on the States, 2008). However, serious fiscal constraints in public spending have necessitated change in the status quo of corrections. Many scholars have argued that this is an opportune time to consider how limited funding could be used in creative ways to manage strategically the carceral population while reducing correctional budgets (Mauer, 2011). This topic is particularly poignant as the overcrowded California penal system was recently declared unconstitutional Eighth Amendment's ban on cruel and unusual punishmentunder the Eighth Amendment's ban on cruel and unusual punishment (*Brown v. Plata*, 2011*)*.

Unfortunately, few empirical studies have been conducted that evaluate the efficacy of "back end" programs designed to reduce prison populations. The work of Wright and Rosky (2011, this issue), in their analysis of the Montana early release program, is a welcome addition to the policy literature. In 2002, Montana implemented an early release program in response to prison crowding and fiscal strain. Like most programs of this type, Montana selected low-risk offenders, defined as offenders sentenced to a 5-year maximum sentence, for early release.

Using propensity score matching, Wright and Rosky (2011) compared recidivism outcomes for individuals in a traditional and early release cohort. When reconviction was used as the dependent measure, the failure rate was similar across groups. However, individuals in the early release group were significantly more likely to return to prison for a technical violation, and their time to failure was substantially shorter. Wright and Rosky contend that the early release policy may contradict the original fiscal goals by increasing returns to prison, thereby escalating the costs of institutionalization.

Direct correspondence to Beth M. Huebner, Department of Criminology & Criminal Justice, University of Missouri—St. Louis, 533 Lucas Hall, St. Louis, MO 63121–4499 (e-mail: huebnerb@umsl.edu).

DOI:10.1111/j.1745-9133.2011.00776.x

Wright and Rosky (2011) highlight several possible explanations for the results. Although the authors did not conduct a process evaluation as part of the study, they suggest that the early release program may shift the burden of enforcement to the local community. The judicial community in Montana expressed concerns about the reforms, particularly in light of the Department of Corrections's expansive autonomy and discretion in making early release decisions. The authors also call for enhanced release programming. Individuals in traditional parole group participated in prerelease planning and the release decision was made by the parole board. The early release cohort was not afforded the same preparatory services, and the staff in the Department of Corrections had sole responsibility for identifying the early release cohort.

The policy essays underscore the individual and organizational challenges faced when implementing a program of this type. Taxman (2011, this issue) suggests that scholars and practitioners look beyond traditional risk-based models of corrections. Instead, she makes a compelling case for a client-centered, health services approach. Like most policies of this type, the Montana program is based on traditional risk- and need-based release and treatment models (Andrews, Bonta, and Hoge, 1990; Feeley and Simon, 1992; Lowenkamp and Latessa, 2005). Taxman argues that programmatic challenges, like those faced in Montana, may reflect the failure of agencies to consider the humanistic needs of parolees. Specifically, she recommends that offenders need more information on the purpose of punishment and how they may benefit from participation and compliance. Active client involvement provides motivation for behavioral change, which is an essential element of long-term success. It is also central to maintain perceptions of justice and fairness. Early release decisions in Wyoming were made outside the traditional parole process. Taxman suggests that decisions to release because of cost savings are not tied to the original purpose of the sentence and may undermine the jurisprudence of punishment. Overall, Taxman's proposals are ambitious and would require a substantial change in the culture and norms of the criminal justice system. That noted, we have little evidence that current models are successful, and behavioral-based models of change have shown evidence of promise in the medical field.

Kulychek (2011, in this issue) argues that we should reflect on early correctional policy to help inform future success. Like Taxman (2011), she suggests we look beyond punishment for punishment sake. For early release policies to be successful, we must better prepare the offender, the system, and the community. Early release should be planned and anticipated and part of a larger comprehensive system of incentives and punishments. Ad hoc changes to policy that are not viewed as legitimate can lead to discretionary decisions made by staff (Petersilia, 2003). Evidence of this type of behavior was observed in the high revocation rates. She concludes with a call to revisit the work of the earliest parole system realized by Maconochie on Norfolk Island (Morris, 2003). Inmates under his supervision were given marks for behavioral change and eventually earned a true second chance.

Finally, Turner (2011, in this issue) summarizes her experiences in California to illustrate the challenges of implementing organizational change. Like Taxman (2011) and Kulychek (2011), she highlights the limitations of risk-based correctional models, particularly given staff distrust in actuarial tools. Risk models also do not consider that certain groups of offenders have higher "stakes" for community and organizational safety. Most recidivism models suggest that traditional low-risk groups, like nonviolent drug or property offenders, actually have some of the highest recidivism rates. Understanding the political and social ramifications of certain offender populations, together with probability-based models of risk, can help craft cogent policy.

Turner (2011) also describes the challenges of transferring management of offenders to the community. California is currently undergoing realignment by shifting the responsibility for low-level offenders to local jails. Devolving prison from the state to the local community makes additional barriers to parole release and revocation, and success of this type of program is conditional on local funding and political support.

As the Wright and Rosky (2011) article and subsequent policy essays indicate, a need remains for inmate-centered, intensive prerelease planning and reentry programming. As important is the necessity for policy makers to be prepared to deal with the landmines that they will find after implementation of a policy of this type.

References

Andrews, Don A., James Bonta, and Robert D. Hoge. 1990. Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice & Behavior*, 17: 19–52.

Feeley, Malcolm M. and Jonathan Simon. 1992. The new penology: Notes on an emerging strategy of corrections and its implications. *Criminology*, 30: 449–474.

Kurlychek, Megan. 2011. What is my left hand doing? The need for unifying purpose and policy in the criminal justice system. *Criminology & Public Policy.* This issue.

Lowenkamp, Christopher T. and Edward J. Latessa. 2005. Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology & Public Policy*, 4: 263–290.

Mauer, Marc. 2011. The challenges of implementing research-based policies. *Criminology & Public Policy*, 10: 69–75.

Morris, Norval. 2003. *Maconochie's Gentleman: The Story of Norfolk Island*. New York: Oxford University Press.

Petersilia, Joan. 2003. *When Prisoners Come Home: Parole and Prisoner Reentry*. New York: Oxford University Press.

Pew Center on the States. 2008. *One in 100: Behind Bars in America 2008*. Washington, DC: Pew Charitable Trusts.

Taxman, Faye S. 2011. The cattle call of reentry: Not all processes are equal. *Criminology & Public Policy*. This issue.

Turner, Susan. 2011. More than just early release: Considerations in prison reduction policies. *Criminology & Public Policy.* This issue.

Wright, Kevin A. and Jeffrey W. Rosky. 2011. Too early is too soon: Lessons from the Montana Department of Corrections Early Release Program. *Criminology & Public Policy.* This issue.

Court Case Cited

Brown v. Plata, 131 S. Ct. 1910 (2011).

Beth M. Huebner is an Associate Professor and Director of Graduate Studies in the Department of Criminology and Criminal Justice at the University of Missouri-St. Louis. Her recent research, funded by the National Institute of Justice, considers the efficacy of sex offender residency laws and explores variation in patterns of offending for urban and rural parolees.

Overview of: "Too early is too Soon

Lessons from the Montana Department of Corrections Early Release Program"

Kevin A. Wright
Arizona State University

Jeffrey W. Rosky
University of Central Florida

Research Summary
Early release procedures will likely become increasingly necessary during a time of fiscal uncertainty in corrections. To date, however, few empirical evaluations exist in the literature to guide correctional administrators in making these potentially unpopular decisions. The failure to appreciate fully the consequences of early release for the criminal justice system (as well as for the general public) could lead to unintended consequences in the form of increased costs and a potential decrease in public safety. The current study seeks to build on the limited information available by evaluating the effectiveness of releasing offenders early in Montana in an attempt to mitigate a budget deficit. The results indicate that offenders released early from a prison setting were more likely to recidivate (and to do so more quickly) than a matched group of offenders experiencing a traditional parole release from prison. Offenders released early from a community setting were somewhat less *likely to recidivate than a matched group of offenders experiencing a traditional parole release from the community. Based on these findings, we assess three plausible explanations for our results:*

1. *A Reduced Deterrent Effect. A possible explanation for the relationship between early release and recidivism identifies a reduction in sentence length as leading to a weakened deterrent effect of criminal justice sanctions. Yet a sizeable body of literature questions the empirical support of deterrence theory in general, and this knowledge coincides with research that suggests that longer sentences produce little gain in terms of reduced recidivism from an incapacitation and a deterrence perspective. In our results, the early releasees from a community setting were less*

DOI:10.1111/j.1745-9133.2011.00764.x

likely to recidivate than their traditional release from community counterparts—a finding that also is at odds with a reduced deterrent effect. Based on these considerations, we conclude that a reduced deterrent effect is unlikely to be responsible for the increased likelihood of recidivism for the early release from prison group.

2. *A Shift in Burden Effect. A second explanation is that the early release of inmates can shift the burden of overcrowding unintentionally from an institutional setting to a community supervision setting. The unscheduled early release of offenders likely increases the caseloads of parole officers and may affect their overall job performance. In our results, we cannot determine definitively that adjustments were occurring without qualitative information from parole officers. It is entirely possible, if not likely, that the label of being an early release may have influenced the differences in technical violations across the four groups—specifically for those released early from prison, which may invoke a different response than those released early who were already in the community. Thus, whereas early release procedures may not influence recidivism rates* directly *through the commission of new criminal acts or violations of parole by those released early, it is possible that recidivism rates may increase because of adjustments made by parole officers. Based on these considerations, we conclude that a shift in burden effect remains a plausible explanation for our findings regarding the increased likelihood of recidivism by the early release from prison group.*
3. *A Failure to Prepare for Reentry Effect. A final consideration is that the early release of inmates from an institutional setting leaves them unprepared to successfully reintegrate back into society. Offenders released today are fundamentally different as compared to those in years past in that they have less programming available to them in prison and have fewer connections with community-based structures. The early release of inmates from prison is likely the epitome of instances in which prisoners are unaware of their discharge date and it is also likely that these offenders were thrust back into society with little time to prepare for successful reintegration. Our findings indicate that the traditional parole from prison group—which in Montana requires that the offender demonstrate a detailed parole plan that includes housing and employment expectations—had the lowest overall recidivism rate. In addition, the early release group from the community setting performed similarly to the traditional release group from the community setting (rather than follow the pattern of the early release from prison group). This finding may indicate a smoother transition into society for these early releases because of a better plan for reentry via placement in the community. Based on these considerations, we conclude that a failure to prepare for reentry effect is a likely explanation for the relationships observed between early release and recidivism.*

Policy Implications

The recent Supreme Court decision in Brown v. Plata *requires California to release approximately 37,000 inmates early due to constitutional deficiencies in healthcare delivery as a result of overcrowding. The situation in California is not unlike other state correctional systems, and underbudgeting and overcrowding will create the need for sensible polices to reduce correctional populations quickly without an increased risk to public safety. The lessons learned from Montana begin with the idea that early release from prison specifically produces several unforeseen consequences beyond that of reducing correctional populations. Yet our research should not be taken to indicate a requirement for offenders necessarily to serve the totality of their sentences. Nor does it advance the idea that an increase in incarceration and its severity would increase public safety. What we argue is that short-term and long-term processes designed to alleviate correctional strain need to be viewed from a reintegration perspective. To that end, we offer three broad and interrelated implications based on the findings of the current analyses:*

1. *Early Release as a Short-term Fix. It would be unwise to suggest that early release procedures should be done away with entirely. Immediate pressures on the U.S. correctional system may produce disastrous consequences for the safety of inmates and correctional officers if they are not ameliorated quickly. Furthermore, budget deficits have left administrators with few alternatives. In the most general sense, however, we join previous scholars in asserting that early release mechanisms should be conceived of as only a short-term remedy for a long-term problem. These approaches should be combined with better inmate projections based on demographics in addition to emergency management planning. Perhaps most importantly, administrators should more fully examine the relationships between prison (re)admissions and effective parole and reentry policies. Short-term emergency release procedures can be successful, but our research indicates that more attention needs to be paid to the transition process from both a supervision and a reintegration perspective.*
2. *Reconsider the Nature of Parole. The current analysis has implications for parole as a discretionary release mechanism as well as a form of postrelease supervision for ex-offenders. First, our findings regarding the successfulness of reintegration for the traditional parole from prison group support the contention that success on parole is increasing in some jurisdictions and suggest that an increase in parole grant rates may do little to increase the rates of criminal behavior (provided that a plan of managed reentry is a requirement of the parole application process). Second, nearly all the ex-offenders (early and traditional released) within the current study were returned for technical violations of their parole; yet significant differences existed across the groups that resulted in early release from prison offenders being more likely to return to custody. Although the commission of a new*

crime clearly necessitates a strong formal response, relatively minor violations (e.g., failure to report and failed urinalysis) could be handled less punitively. Instead, a system of graduated sanctions could be created that resorts to reincarceration as a last option for repeat violators. Such an approach is likely to reduce correctional populations as well as the costs associated with reincarcerating ex-offenders.

3. *Prepare for Reentry. Instead of conceptualizing parole as an extended sentence of supervision for offenders, it could be conceived of as a managed reentry mechanism with an explicit focus on successful reintegration. Reentry should begin within prison walls through specific planning for each offender. After discharge, a system of managed reentry could "seize the moment of release" by providing ex-offenders with the support needed to perform simple but necessary tasks such as obtaining an identification card. The system should be front loaded, with the bulk of services concentrated within the first 6 months of release and should provide offenders with the opportunity to accomplish several requirements (e.g., housing, employment) in one location. Additionally, the opportunity for individuals to "graduate" from parole early would assure that resources were reserved for the ex-offenders most vulnerable for a return to crime. Each of these recommendations recognizes that early release policies will likely affect the opportunities of ex-offenders to make good while producing the added benefit of reducing the burden placed on parole officers.*

Early release procedures will undoubtedly become more commonplace in corrections as a means to overcome budget shortfalls and prison crowding. It is therefore necessary that precautions be taken to ensure that offenders are fully prepared to succeed after reentry into society. A full appreciation for the complexities of early release from a reintegration perspective could indeed serve to save money and correctional space while increasing public safety through the reduction of future offenses. It also could lead to additional benefits such as decreases in child abuse, family violence, and community disorganization, and it would create an opportunity to save considerable time and money to treat social ills through the offender population.

Keywords

early release, *offender reentry*, *prison crowding*, *recidivism*, *parole*

Too early is too soon

Lessons from the Montana Department of Corrections Early Release Program

Kevin A. Wright
Arizona State University

Jeffrey W. Rosky
University of Central Florida

In their seminal work, *Reaffirming Rehabilitation*, Cullen and Gilbert (1982: 176) issued a warning that "in the face of teeming penitentiaries, alternative release procedures could and undoubtedly will be evolved. Yet these adaptations are likely to be hastily instituted and to create new inefficiencies and inequalities in the administration of justice." Nearly three decades later, their premonition has proved to be correct as state correctional administrators have struggled in efforts to combat the "incompatible and powerful forces" (Cullen, Wright, and Applegate, 1996: 70) of underfunding and overcrowding (see, e.g., Lane, 1986). Indeed, by year-end 2009, 19 states and the federal government had prison systems operating at more than 100% of their highest inmate capacity with 27 operating at more than 100% of their lowest capacity (West, Sabol, and Greenman, 2010).[1] Additionally, the current economic crisis has led to significant, across-the-board cuts in the seemingly untouchable sphere of state correctional budgets (Engel, Larivee, and Luedeman, 2009). The task, then, is for researchers, policy makers, and practitioners to find ways to alleviate these strains without compromising the goals of corrections or the safety of the general public.

The authors thank Montana Department of Corrections Director Mike Ferriter and Chief Information Officer John Daugherty for access to the data used in this study as well as Dewey Hall and Mark Johnson of the MT DOC Statistics Unit for help with data management issues. The views and opinions expressed by the authors do not necessarily represent those of the Montana Department of Corrections or its employees. Direct correspondence to Kevin A. Wright, School of Criminology and Criminal Justice, Arizona State University, 411 N. Central Avenue, Suite 600, Phoenix, AZ 85004 (e-mail: kevin_wright@asu.edu).

1. The highest capacity is the sum of the maximum number of beds across three capacity measures (rated, operational, and design), whereas the lowest capacity is the minimum of these three capacity measures (see West et al., 2010).

DOI:10.1111/j.1745-9133.2011.00765.x

In general, three approaches have been taken to reduce the spatial and fiscal constraints faced by the U.S. correctional system (Blumstein, 1988). The first, and perhaps most straightforward, is to increase prison capacity (i.e., build our way out of the problem). This approach was favored in the 1990s and is likely less of an option in a time of fiscal uncertainty and increased prison populations. The second approach is to decrease prison admissions through "front-end" solutions where offenders are diverted to sentences other than prison (e.g., probation). The challenge in doing so is to develop a range of sentences that leave judge, victim, and community satisfied while also avoiding "widening the net of social control" (Tonry and Lynch, 1996). Finally, an increasingly used group of strategies is that of "back-door" solutions—including modifications to parole release and good-time policies, and the use of emergency-release mechanisms to reduce current populations. These options can be implemented in a shorter time frame than front-door strategies and have the added benefit to administrators of often occurring outside of public and judicial view.

Back-door strategies essentially reduce the amount of time served by offenders and thus have historically been a controversial approach to reducing overcrowding. Austin (1986) identified a "dark side" of early release in the form of potential financial and nonpecuniary costs for victims of crimes committed by inmates released early. Furthermore, he noted that these strategies provide an excessive amount of discretion to correctional administrators, which subverts the principles of equity and certainty in sentencing by the court and often leads to public outcry over leniency in punishment. Given that early release often occurs behind closed doors, little is known about the extent to which these unintended consequences outweigh the benefits of immediate population reduction. To be sure, despite the increased popularity of these approaches, relatively few empirical assessments of early release procedures have appeared in the criminological literature (cf. Austin, 1986; Joo, Ekland-Olson, and Kelly, 1995).

Taken together, these concerns signal the need for a more rigorous examination of the consequences of early release as a mechanism to reduce correctional populations. As the state of California searches for ways to comply with a federal court order to release more than 40,000 inmates by 2011 (Archibold, 2010), it is imperative that assessments of existing release programs are made available to guide release policies properly. The purpose of the current study is to take a step in this direction by evaluating an early release program in Montana designed to mitigate the effects of a $9 million budget shortfall. Specifically, we use frailty-adjusted Cox proportional hazard models to compare the reoffending rates of an early release subgroup with that of more traditional releases. Although recidivism is admittedly a limited and often debated measure of policy success, we believe that it is critically important to the long-term goals of reducing prison populations and correctional spending.[2] To that end, we conclude the article by providing a detailed discussion of possible explanations for our findings and their policy implications with special attention given to

2. See, for example, the discussions by Maxwell (2005) and Lynch (2006).

an issue that has thus far escaped much of the early release literature: the importance of evidence-based offender reentry.

Early Release as a Correctional Tool

The difficulty in reviewing the existing literature on early release stems from deciding which studies should be included based on definitional considerations. For example, accelerated release via an increased leniency on the behalf of a parole board could qualify as early release; similarly, the accumulation of good-time credits and corresponding shorter time served could be considered an early release.[3] Nevertheless, without specific mention of a concentrated effort to relieve strains on the correctional system, it would be nearly impossible to ascertain whether early release was, in fact, occurring. Given the scope of the current study, this problem is somewhat avoided by only reviewing those works that examine explicitly an early release procedure designed to alleviate spatial or financial strains.[4] This restricts the pool of available studies considerably, but the dynamics present under these conditions are different compared with a standard early release mechanism triggered by good behavior while institutionalized. Only three published studies examining recidivism among early releases were identified and are examined in more detail in this article.

Sims and O'Connell (1985) assessed six cohorts of early released inmates from 1979 to 1984 in Washington State. More than 1,600 inmates were paroled an average of 6 months early to comply with a court decision (*Hoptowit v. Ray,* 1982) designed to reduce prison crowding. Each cohort varied in terms of composition (e.g., the percentage of violent offenders), as well as the legal authority on which early release was granted. The recidivism analyses compared each group with historical recidivism rates in addition to a control group comprising 1,867 traditional releases. Of the four groups for which information was available for at least 3 years after release, three cohorts alleviated overcrowding somewhat with minimal risk to public safety (in terms of percentage reincarcerated, percentage reincarcerated within the early release period, etc.) compared with controls. The other cohort had higher recidivism rates than the traditional releases—a finding that the authors suggested was because of the higher overall number of inmates released at one time, as well as because of the higher percentage of prior recidivists in that group compared with the others. Sims and O'Connell concluded that early release can provide only a temporary relief to overcrowding and that the risk to the general public is contingent on the availability of low-risk inmates for early release.

3. These two examples represent the most general types of early release for states with indeterminate sentencing (accelerated release via parole board) and for those with determinate sentencing (increased application of good-time credits).

4. These criteria preclude the inclusion of analyses that sought to determine the impact of court-ordered early release for reasons other than crowding or budget concerns (see, e.g., Eichman, 1966; Malak, 1984; see Guzman, Krisberg, and Tsukida, 2008, for a broader review of accelerated release programs).

A series of studies by Elkland-Olson, Kelly, and colleagues (Ekland-Olson, Kelly, Joo, Olbrich, and Eisenberg, 1993; Joo et al., 1995; Kelly and Ekland-Olson, 1991) provided some additional insight into early release through their assessment of four parolee cohorts in Texas. In *Ruiz v. Estelle* (1980), a federal court ruled that prison conditions in Texas were unconstitutional because of severe overcrowding. The state legislature passed the Prison Management Act (PMA) in 1983 to comply with the court's ruling. When populations exceeded capacity, the PMA triggered the administration of more liberal good-time credits in addition to the advancement of parole eligibility. The preceding studies therefore represent recidivism analyses during a time of accelerated release in the Texas prison system.[5]

The four cohorts were composed of releases from 1984 to 1987 who were followed during a 36-month period using surviving analyses to document reincarceration. The 1984 and 1985 cohorts paralleled national trends, and the authors concluded that changes in the administration of justice were unimportant for these groups. The 1986 and 1987 cohorts, however, differed substantially from the other cohorts. Each departed from the baseline in unique ways, but in general both cohorts performed worse (e.g., a greater percentage recidivated or did so more quickly) than the 1984 and 1985 cohorts. The authors speculated that a combination of reduced deterrence and increased strain placed on parole officers may have been responsible for the results of the latter cohorts (to be discussed more fully in the Conclusions section).

One additional analysis of the Texas overcrowding response supplemented these initial works. Joo et al. (1995) added a specific early release cohort to determine the impact of the PMA. The 1987 version of the PMA required that the prison director award 30 days of good time to all eligible inmates when the population reached 95% of capacity. If this measure did not reduce the population sufficiently, then the prison director was to award additional time (up to 90 days), and the Board of Pardons and Paroles was to advance the parole eligibility and review dates of eligible inmates by an equal amount of days. A 1987 cohort released under these provisions was compared with a similar traditional release cohort from the same year. The early cohort was more likely to be reincarcerated at 12, 24, and 36 months, with the first 12 months producing the greatest difference in recidivism between the two groups (76% survival of the early cohort vs. 83% of the baseline cohort). It is important to note that there were some compositional differences between the two groups, but taken as a whole, these studies indicated that the early release process in Texas may have actually contributed to an increased incarceration rate (Kelly and Ekland-Olson, 1991).[6]

5. Indeed, Kelly and Ekland-Olson (1991) noted that prison releases increased to 30,102 in 1989, which was up from 7,180 in 1980. Additionally, nearly 80% of inmates were released on parole after their first hearing in 1989, which was up from 40% in 1983.

6. More specifically, the high-risk group of the early release cohort produced a higher risk score than the high-risk group of the baseline cohort (see Joo et al., 1995: 403).

Finally, perhaps the most comprehensive study to date on the impact of an early release procedure was Austin's (1986) evaluation of the Illinois Department of Corrections (IDOC) program that released more than 21,000 inmates early between 1980 and 1983.[7] Extreme overcrowding—leading to unsafe and inhumane facilities and an eventual riot that killed three guards—was cited as the leading concern in the need to reduce populations in a relatively short time period. Early release was then accomplished by two mechanisms. First, a more formal procedure described as "forced release" identified inmates who were nearing sentence completion as candidates to receive good-time credits (awarded by the Director of Corrections) to accelerate their release. This policy occurred on a weekly basis to best accommodate the impact of prison admission fluctuations on the overall prison population. Inmates selected for early release were primarily property offenders who had been within the custody of the IDOC for at least 90 days and had been approved for release by the warden of the institution. A second, less formal mechanism was the awarding of good-time credits to *any* (i.e., not necessarily near release) inmate by the Director based on the recommendation of the wardens. These recommendations were based on satisfactory work and disciplinary records, and the overall time served was effectively shortened by a sizeable proportion of the inmate population. Of the two mechanisms, the informal procedure was responsible for the bulk of good-time credits awarded by the Director and contributed most heavily to the overall early release program. On average, 105 days were deducted from the sentence of early release inmates, which represented a 12% reduction in their expected length of imprisonment (Austin, 1986).

Using a random sample of 1,500 inmates released during the period of 1979–1982, Austin compared early releases and traditional releases on two forms of recidivism (official arrests and parole violations) using multiple types of analyses (e.g., survival analyses and risk model simulations). He observed that early release had no impact on the overall rates of rearrest or parole violations for all released offenders and that offenders released early actually had lower rearrest rates and were arrested for fewer violent offenses than non–early releasees.[8] Additionally, substantial savings were recorded (approximately $1,480 per early release), and the early release program was partially responsible for avoiding additional overcrowding. Despite these positive findings, early release increased the total amount of crime reported to police in Illinois, and the costs incurred by victims offset a substantial portion of the total savings. Nevertheless, Austin (1986: 469) concluded, "Relatively minor

7. The state of Illinois is again at the forefront of early release controversy as two programs supported by Governor Pat Quinn were suspended amid concerns over public safety. Quinn recently signed a bill requiring the Department of Corrections to post photographs online of offenders who were released early.

8. It is important to note that the early releases represented better public safety risks because of the criteria for inclusion in the program (i.e., good conduct within the institution and held at lower security levels before release) (see Austin, 1986: 443–446, for a discussion of the additional differences between the two groups).

adjustments in time served have little influence on the probability of recidivism compared to the more powerful factors predictive of recidivism."

The IDOC early release program therefore had the intended effect for state officials—substantial overcrowding was avoided and early releases were no more likely to recidivate than traditional releases. The program was not, however, well received by a general public concerned over the increasing amount of crime suffered and by criminal justice system actors who felt their work was undermined. It also must be recognized that the program was just one part of a three-pronged approach to reducing overcrowding that included front-end procedures (e.g., diversion to intensive supervision) and prison capacity expansion. Accordingly, Austin could reach no definitive conclusions on whether the program represented good correctional policy. He advised that early release could provide no more than a short-term remedy—not a permanent solution—for prison crowding.

The overall lack of empirical evaluations of early release procedures should not come as a surprise. As stated, the general public (and researchers in particular) often are not privy to the potentially unpopular decisions made to reduce correctional populations. The paucity of studies should, however, come as a disappointment to researchers and correctional officials who have much to gain from an understanding of the outcomes of these decisions. Currently, little agreement exists in the literature on early release procedures and their consequences; yet it can be expected that they will be implemented to an increasing degree in the face of budget crises.[9] Adding to the appeal of early release strategies are recent Bureau of Justice Statistics data that indicate a lack of relationship between length of stay and recidivism rates, which suggests that prison terms could conceivably be reduced without any major spike in reoffending (Langan and Levin, 2002; see also Austin, 2010). Accordingly, the current study evaluating the Montana Department of Corrections (MDOC) early release cohort assumes an importance beyond that of the typical recidivism analysis.

Early Release in Montana

The MDOC is one of the smaller state correctional systems in the United States. At the time of the release program, it was ranked 44th among state prison populations with 3,340 inmates (Harrison and Beck, 2004). Montana is a large, rural state that also ranked 44th in total population with slightly more than 900,000 total residents in 2000 (U.S. Census, 2000).[10] At mid-year 2010, Montana had 12,983 offenders under supervision with 2,570

9. At the time of writing, Michigan, Kentucky, and Ohio were among states considering early release proposals to alleviate financial strains.

10. A legitimate concern is the extent to which a small, relatively homogenous state like Montana could generalize to a larger, more heterogeneous state like California. Although the states clearly differ in terms of size and demographics, they share similar characteristics regarding criminal justice procedures. In particular, both states count more than 50% of their prison admissions as being parole violators—a sizeable portion of which are composed of technical violations specifically (Travis and Lawrence, 2002). Nevertheless, the subsequent findings should be interpreted with this important limitation in mind.

in a prison; 918 in a prerelease center; 338 on intensive supervision; 8,367 on probation or parole; and 790 in treatment programs. The average time served in prison for released inmates in 2010 was 21.2 months for males and 15.5 months for females (MDOC, 2010a).

The parole process is run by the Montana Board of Pardons and Parole (BOPP). All inmates, including those placed by the prison in prelease centers and on intensive supervision, are eligible for parole except those serving life without parole or a death sentence. Parole eligibility dates are calculated by the MDOC using current statutes and court criteria, although inmates are only considered for parole if they have at least 120 days clear conduct prior to their eligibility date (Montana Board of Pardons and Parole, 2010a). The parole process itself begins with an inmate's application and development of parole plan to include housing, employment or education, and treatment programs coupled with a budget schedule to pay fines, fees, and restitution (Montana Board of Pardons and Parole, 2010b). Additionally, only inmates released from a prison facility are given gate money (Montana Department of Corrections, 2010b). Parole revocations are performed through a hearing process initiated by a parole officer through a revocation report. Offenders are allowed to have counsel, use evidence, and call witnesses to contest the revocation. The BOPP can then dismiss the revocation, revoke the offender back to prison, or use other intermediary sanctions at its discretion (Montana Board of Pardons and Parole, 2010c).

In June 2002, the MDOC began releasing several hundred inmates early in the hopes of mitigating a $9 million budget deficit. The intention of this early release program was to reduce the costs associated with high levels of imprisonment in Montana. Offenders deemed to be a "low risk" toward reoffending were selected for this early release—with the main qualification being that their crime was not of a violent or sexual nature. Contrary to standard release procedures (e.g., parole board review for release), these offenders were released from prison or a community program at the discretion of the MDOC and placed under the supervision of its Community Corrections Division. The only individuals who were eligible for this release were part of a sentencing option known as a "DOC Commitment" in which offenders are sentenced to a maximum of 5 years to the MDOC in lieu of a longer sentence in prison.[11] Thus, these offenders could be considered to pose a lesser threat toward members of society as they are mainly composed of nonviolent, drug, and property offenders.

Not surprisingly, a public battle among the MDOC and legislators, judges, and prosecutors ensued over the appropriateness of the early release program in Montana. The Director of the MDOC at the time attempted to justify to the general public the effectiveness and safeness of the early release strategy. He noted that of the 298 offenders released early as DOC Commitments—only 18 of them (6%) had been returned to prison

11. The "DOC Commitment" sentencing option was created in 1993 by the Montana Legislature that allowed a judge to sentence an offender to the MDOC for appropriate placement within its system rather than sentencing an offender directly to prison. Excluding deferred and suspended probationary sentences, approximately 80% of prison and community admissions to the MDOC are DOC Commitments (MDOC, 2010a).

for violating the rules of the release (Slaughter, 2002). He also attempted to put at rest the idea that violent offenders, as well as offenders with substantial time to serve remaining, were being selected for early release. The early release decision process was characterized by a "dynamic system of checks and balances" (Slaughter, 2002, para. 6) in which multiple levels of the correctional system were involved in selecting appropriate individuals for release. At the time, the conditional release program was said to be achieving the dual objectives of saving taxpayer dollars without impacting public safety in a negative manner.

Those outside of the MDOC, however, did not have as much faith in the program. Montana attorneys in particular were incensed as they believed the policy undermined efforts to punish offenders properly for the crimes they committed (Montana County Attorneys Association [MCAA], 2002). They noted that the program was only in existence for 2 months, and thus, evaluations of recidivism rates were unwarranted because of the lack of an opportunity to reoffend. Images of Willie Horton were invoked as the early release program operated "without legislative approval, without administrative rules, and without public input or comment" (MCAA, 2002, para. 2). The safety of the public, therefore, was depicted as being in extreme jeopardy because of the MDOC early release program. Judges in Montana were equally upset with the policy, and many decided to sentence offenders to prison specifically (rather than to the DOC) to avoid them having eligibility for eventual early release (Associated Press, 2002). Ultimately, the early release program was depicted by outsiders as a selfish attempt by the DOC to overcome budget deficits and overcrowding with little regard for the possible unintended consequences on society. At issue, then, is whether the early release program alleviated the budget crisis without compromising public safety via increased recidivism.

Current Focus

Given the presence of overcrowding and underfunding within the U.S. correctional system, and the general lack of consensus regarding the findings of previous studies, the current work seeks to expand the knowledge base of the potential usefulness of early release policies. More specifically, we compare the reincarceration rates of the early release cohort in Montana with that of more traditional releases—both from institutional and community settings. We move beyond previous works by employing frailty-adjusted Cox proportional hazard models to account for the heterogeneity in propensity to fail across individuals. Finally, the current study adds to the literature by providing a more detailed discussion of the findings and their corresponding policy implications, with specific attention given to the problem of offender reentry.

Data and Measures

Data on all releases from both prison and community correctional facilities were obtained from the MDOC for the period between June 1, 2000 and January 1, 2007 (total

$N = 5{,}668$).The information provided included demographics, historical movements within the correctional system (including date of return to custody, if applicable), and sentencing information such as prior convictions. Most importantly, the data indicated whether an offender was part of the conditional release group. Using this information, the offenders were classified as one of four release statuses: (a) traditional parole from prison, (b) conditional release from prison, (c) traditional parole from a community setting, or (d) conditional release from a community setting. Releases from prisons included inmates from Montana State Prison, Montana Women's Prison, two regional prisons, and a private prison. Releases from community settings included inmates from six prerelease centers, multiple drug treatment facilities, and those on intensive supervision.[12] The bulk of offenders were released via traditional parole from community settings ($n = 2{,}365$, 42%), with similar percentages released via both conditional release from community settings ($n = 1{,}589$, 29%) and traditional parole from prison ($n = 1{,}250$, 22%). The smallest group was composed of offenders who were conditionally released directly from prison ($n = 464$, 8%). Collectively, then, 37% of the releases within this study were conditionally released. The final sample included 4,929 offenders accounting for 5,668 releases, with the large majority of offenders (4,245, 86%) having only been released once in the time frame.

Dependent Variable

The outcome of interest for the current analysis was defined as any return to the same level of custody or higher during the study time frame. Thus, the measure parallels the use of reincarceration as a measure of recidivism by previous studies (Joo et al. 1995; Wilson, 2005), and therefore it represents "a complex measure of criminal behavior combined with formal and informal policy and procedure mechanisms" (Wilson, 2005: 494). More specifically, the inclusion of technical violations within the recidivism measure allows for the examination of potential differential revoking practices across the four groups (e.g., the early release group experiencing more technical violations because of an unexpected, increased workload for parole officers).[13] Additionally, as will be shown in the subsequent discussion, technical violations led to the bulk of returns and removing them from the

12. Most offenders within the community were released from prerelease centers (78%), with smaller percentages of offenders being released from treatment programs (16%) and intensive supervision (6%). Importantly, regardless of placement, nearly all community offenders spent a significant amount of time in the community (i.e., rarely does an offender spend all of his or her time institutionalized).

13. It is critically important to identify any possible differences in revoking practices across the four groups. In particular, revocation decisions for inmates released on traditional parole from both prison and community setting were performed by the BOPP, whereas revocation decisions for inmates conditionally released from prison and community settings were performed by the MDOC. Revocation hearings performed by the BOPP followed formal processes as specified by statute, whereas those performed by the MDOC for conditional releases were informal, shifting the decision to revoke solely to the MDOC rather than to the BOPP. This notable difference likely impacts technical violation rates and is discussed in the Criminal Justice Thermodynamics section.

analyses would miss a large component of returning offenders. Perhaps most importantly, the intention of the early release policy was to remove some of the financial burden placed on the MDOC by an increased correctional population. Documenting offenders who return to custody is therefore the most appropriate assessment of the effectiveness of this policy.

Independent Variables

In addition to release status, we include several theoretically and empirically relevant individual-level characteristics of offenders in our analyses. Prior recidivism studies have documented that offenders who are male, younger, of minority status, and have extensive criminal histories are more likely to recidivate (see, e.g., Gainey, Payne, and O'Toole, 2000; Huebner and Berg, 2011; Spohn and Holleran, 2002). The current sample was primarily male (81.7%) with an average age of 35.2 years. Native Americans are the dominant minority in Montana, and the sample reflected this as 17.1% of the sample were identified as Native American. Whites comprised 77.0% of the sample, Hispanics 3.5%, African Americans 1.4%, and 0.9% were other or unknown. The sample also consisted of mostly nonviolent offenders with only 24.7% of the offenders having a conviction for felony assault, robbery, arson, kidnapping, sexual assault, or homicide. The average length of stay for offenders ranged from a low of 24 months for the conditional release from the community group to a high of 69 months for the traditional parole from prison group. On average, then, conditional release offenders were released approximately 22 months early from prison and 10 months early from the community (see Tables 1 and 2 for full sample demographics).

It is important to note that several additional statistically significant differences were found across groups in terms of demographics and prior criminal involvement. Most notably, the traditional parole from prison group differed from the other three on several dimensions, including gender (higher percentage of males), prior drug convictions (lower percentage), prior violent convictions (higher percentage), and age (older). In addition, the conditional release from prison group contained a substantially higher percentage of Native Americans (24%) than the other three groups. The conditional release from community group was also unique in that it contained a higher percentage of drug offenders than the remaining groups.

Analytic Strategy

The primary focus of the current analysis is to determine whether offenders who were released early in Montana were more likely to recidivate (and to do so more quickly) than offenders released in a more traditional manner. Accordingly, the analysis proceeds in several stages. First, we split the sample into releases from the community and releases from prison because of differences in demographics and release policies across these two categories. Next, we employ a one-to-one nearest-neighbor propensity score matching scheme to

TABLE 1

Comparison of Key Demographics for Prison Full Sample and Prison Propensity Score Matched Sample

	Prison Full Sample			Prison Matched Sample		
Variable	**TP ($n = 1{,}250$)**	**CR ($n = 464$)**	**p Value[a]**	**TP ($n = 434$)**	**CR ($n = 434$)**	**p Value[a]**
Offender						
Gender (1 = male)	89.0%	83.0%	<0.001	83.0%	83.1%	0.978
Mean release age	37.6	34.1	<0.001	34.6	34.2	0.373
Native American (1 = yes)	15.9%	23.9%	<0.001	24.4%	22.8%	0.576
Criminal History						
Length of stay (months)	69.4	47.0	<0.001	50.5	47.9	0.530
Drug[b]	31.0%	33.4%	<0.001	31.6%	33.4%	0.562
Theft	23.2%	25.0%	<0.001	24.9%	24.7%	0.937
Other nonviolent[c]	62.3%	67.7%	<0.001	63.6%	68.4%	0.132
Mean number of violent convictions[d]	0.5	0.3	<0.001	0.3	0.3	0.996
Mean number of nonviolent convictions	2.3	2.3	<0.001	2.3	2.3	0.378
Mean total convictions	2.8	2.7	<0.001	2.6	2.7	0.933

CR, conditional release; *TP*, traditional parole.

[a] p values are from chi-square tests for the categorical variables and Wilcoxon ranked sum test for the continuous and count variables.

[b] Drug possession and manufacture, sale, or possession with intent to sell were combined into a single drug offense variable as there was not enough information to discern whether an individual was involved in drug dealing or simple drug usage as plea bargaining may have been involved. Individually, drug possession and manufacture, sale, or possession with intent to sell were 15.3% and 9.4% of total convictions, respectively, with other types of drug offenses included.

[c] Other nonviolent convictions included burglary, check kiting, felony DUI, forgery, criminal endangerment, and other offenses deemed felony by Montana statute.

[d] Violent convictions included felony assault, robbery, arson, kidnapping, sexual assault, sexual abuse, and homicides. The most common violent conviction was felony assault, which consisted of 58.6% of all violent convictions. Robbery was next at 14.6% followed by sexual assault at 13.7%. Homicide and attempted homicide consisted of only 3.2% of the violent convictions.

reduce potential bias from imbalanced covariates and nonrandom group assignment within the community and prison release samples (Rosenbaum and Rubin, 1983; Stuart, 2010). In the one-to-one matching scheme, a logistic regression is fit modeling the likelihood of being a conditional release based on sex, age, gender, length of stay, and criminal history variables. The resulting predicted probability from the model for each conditional release observation is then matched with one observation from the traditional sample that had the nearest corresponding probability. If a match cannot be found, then the observation is omitted from subsequent analysis. Balancing the groups allows us to be more confident that our results are a result of the treatment condition (i.e., early release status) rather than of the sample selection bias (King, Massoglia, and MacMillan, 2010). The matching scheme

TABLE 2

Comparison of Key Demographics for Community Full Sample[a] and Community Propensity Score Matched Sample[a]

	Community Full Sample			Community Matched Sample		
Variable	**TP ($n = 2{,}365$)**	**CR ($n = 1{,}589$)**	***p* Value[b]**	**TP ($n = 1{,}422$)**	**CR ($n = 1{,}422$)**	***p* Value[b]**
Offender						
Gender (1 = male)	82.2%	74.9%	<0.001	76.4%	77.9%	0.326
Mean release age	35.0	34.1	<0.001	34.2	34.5	0.634
Native American (1 = yes)	16.7%	16.7%	0.995	16.7%	16.6%	0.919
Criminal History						
Length of stay (months)	33.3	23.5	<0.001	26.1	23.9	0.054
Drug[c]	37.2%	41.6%	0.005	40.1%	39.9%	0.567
Theft	26.6%	19.1%	<0.001	19.8%	20.5%	0.640
Other nonviolent[d]	64.2%	58.0%	<0.001	58.7%	58.9%	0.879
Mean number of violent convictions[e]	0.3	0.2	<0.001	0.2	0.3	0.727
Mean number of nonviolent convictions	2.5	2.0	<0.001	2.0	2.0	0.398
Mean total convictions	2.9	2.2	<0.001	2.3	2.3	0.586

CR, conditional release; *TP*, traditional parole.

[a] Community releases includes those from prerelease centers, treatment programs, and intensive supervision.

[b] *p* values are from chi-square tests for the categorical variables and Wilcoxon ranked sum test for the continuous and count variables.

[c] Drug possession and manufacture, sale, or possession with intent to sell were combined into a single drug offense variable as there was not enough information to discern whether an individual was involved in drug dealing or simple drug usage as plea bargaining may have been involved. Individually, drug possession and manufacture, sale, or possession with intent to sell were 15.3% and 9.4% of total convictions, respectively, with other types of drug offenses included.

[d] Other nonviolent convictions included burglary, check kiting, felony DUI, forgery, criminal endangerment, and other offenses deemed felony by Montana statute.

[e] Violent convictions included felony assault, robbery, arson, kidnapping, sexual assault, sexual abuse, and homicides. The most common violent conviction was felony assault, which consisted of 58.6% of all violent convictions. Robbery was next at 14.6% followed by sexual assault at 13.7%. Homicide and attempted homicide consisted of only 3.2% of the violent convictions.

reduced the final sample size used in the models from 3,954 releases to 2,844 releases in the community group and from 1,714 releases to 868 releases in the prison group.[14]

We then estimate Cox proportional hazard models to document differences in the overall recidivism rates between conditional release and traditional parole within the prison and community group matched samples separately. In a proportional hazards model, it is assumed that there is a common baseline hazard ratio for all subjects that changes with

14. An analysis of the 30 observations that were omitted from the prison conditional release group revealed that 29 were serving a driving-under-the-influence (DUI) sentence from the latter part of the 1990s and did not have the same criminal history as others that were retained in the sample. The traditional parole group had few of these DUI-only offenders. The other observation had only a deliberate homicide conviction and no other criminal history. These cases were atypical of those eligible for parole and thus had no match in the control group. Similar results were found for the community sample.

the inclusion of covariates in the model. The estimates of the covariate effects then can be reported as risk or hazard ratios proportional to the baseline hazard (Collett, 1994). The model gives estimates of the average time-to-failure results for the various groups in the study (in addition to the overall failure rates) and allows for speculations to be made about the possible reasons for differences in recidivism (Lynch, 2006). Finally, this method also allows for the handling of censored data where failure (in this case, return to custody) has not yet occurred.

Given that 13.9% of the sample consisted of offenders with two or more releases, we also add a frailty component to the models to account for the variance resulting from repeated measures. Frailty-adjusted proportional hazards modeling is a recent survival analysis method rooted in biomedical applications. The technique allows the proportional hazard model to be modified and, hence, the baseline hazard to be modified, by adjusting for the fact that some individuals are potentially more frail than others with respect to some outcome and, thus, more likely to fail (Shoukri and Pause, 1999). Similar to a random-effects general linear model, this frailty component is structured as an unobserved covariate that is incorporated into the survival model as a random effect and modifies the hazard function to allow for differing propensities to fail (Hougaard, 1995). Finally, with regard to our sample, we lack information on offender treatment, education level, family history, and other pertinent characteristics—the frailty component also adjusts the overall model to account for these unobserved covariates and creates better estimates for the main factors in the model. In short, the frailty-adjusted proportional hazard model allows us to determine the importance of release status for recidivism independent of any potential individual characteristic or risk factor not included in the analyses.

Results

Tables 1 and 2 present the results of the propensity score matching technique applied to the full sample for the prison releases and the community releases, respectively. As noted, offenders from the conditional release group were matched to offenders from the traditional parole group in an attempt to isolate the effects of early release on recidivism for the analyses presented in the subsequent discussion. Prior to matching, significant differences ($p < 0.01$) emerged between the conditional and traditional parole releases on nearly all the covariates in both samples. The propensity score matching method balanced these variables effectively to make the two groups more equal prior to the survival analyses. Given that we are most interested in the effect of early release, the final matched sample essentially produced two groups (traditional parole and conditional release) that were comparable with each other from a prison setting and two groups that were comparable with each other from community settings.

Table 3 presents the recidivism statistics from the survival analyses conducted on the separate matched samples. All groups had similar percentages of individuals who returned

TABLE 3

Percent Recidivated, Type of Return, and Time to Recidivism by Release Group, Matched Samples

Variable	TPP ($n = 434$)	CRP ($n = 434$)	TPC ($n = 1,422$)	CRC ($n = 1,422$)
Total returned	30.2%	36.4%	36.2%	34.2%
Returned new conviction	5.1%	5.2%	5.2%	5.1%
Mean days to failure	250.1	227.5	284.1	223.2
Median days to failure	199.0	164.0	201.0	184.0

CRC, conditional release from community; *CRP*, conditional release from prison; *TPC*, traditional parole from community; *TPP*, traditional parole from prison.

to custody with a new conviction (approximately 5%), which indicates that most offenders had their parole revoked for technical violations. The traditional parole from prison release group had the lowest overall recidivism rate (30.2%), and the conditional release from prison group had the highest overall recidivism rate (36.4%). This latter group also had the quickest median time to failure, with the average ex-offender returning to custody a little more than 5 months from his or her release date. In contrast, those who did recidivate from the traditional parole from prison group did so over a longer time period. Overall, then, the releases from the conditional release from prison group failed more often and did so more quickly than the remaining three groups.

Turning to the frailty-adjusted Cox proportional hazard models, it is again apparent that the conditional release from prison group performed poorly relative to traditional releases. Table 4 indicates that the DOC Commitments released from prison were two times more likely to recidivate than their traditional parole from prison counterparts. Offenders who were younger, male, and Native American were more likely to recidivate, yet none of the criminal history variables emerged as significant predictors of recidivism in the prison sample. Turning to the community sample, the conditional release offenders were somewhat *less* likely to recidivate than those released via traditional parole. Younger, male, and Native American offenders were again more likely to recidivate, and those who had a prior theft or other nonviolent conviction were more likely to fail (see Table 5).

A consistent picture emerged throughout the analyses in regard to the four release groups. The conditional release from prison group was more likely to recidivate and to do so in a quicker time period. This finding held even when balancing the groups on theoretically relevant control variables as well as when adjusting for unobserved heterogeneity via the frailty component. On the other end of the spectrum, those released via traditional parole from prisons performed considerably well overall. As noted, the parole process in Montana begins with an inmate demonstrating a parole plan that includes expected housing and employment details. This requirement could potentially create a preparation for reentry

TABLE 4

Frailty-Adjusted Cox Proportional Hazard Model Predicting Time to Failure, Prison-Matched Sample

Variable	*B*	SE	Relative Risk
Release Status			
Traditional parole from prison	—	—	—
Conditional release from prison	0.691***	0.136	2.00
Offender			
Gender (1 = male)	0.508*	0.212	1.65
Mean release age	−0.034***	0.008	0.97
Native American (1 = yes)	0.674***	0.158	1.96
Criminal History			
Length of stay (months)	0.002	0.001	1.00
Total number of convictions	−0.071	0.052	0.93
Drug conviction	0.000	0.176	1.00
Theft conviction	0.135	0.182	1.14
Other nonviolent conviction	0.279	0.185	1.32
Violent conviction	−0.162	0.186	1.08
Frailty Gamma*			

Notes. Likelihood Ratio Test = 312***; $df = 10$. $R^2 = 0.3$; $n = 868$ releases.*$p < 0.05$. ***$p < 0.001$.

that is not available to those released early from prison. Furthermore, the early release group from the community setting performed similarly to the traditional release group from the community setting. This finding may indicate a smoother transition into society for these conditional releases (compared with their conditional release from prison counterparts) because of a better plan for reentry via placement in the community. The early release from community cohort was slightly less likely to fail overall compared with the traditional releasees from community; yet early release failures returned more quickly than failures from the traditional group. Overall, we have clear indications that the early release procedure, particularly for those released directly from prison, may have exacerbated the financial strain problem in Montana as a result of increased recidivism.

Discussion

The primary focus of the current study was to determine the effectiveness of an early release policy in Montana designed to mitigate a budget deficit. The results indicated that early release offenders (particularly those from a prison setting) were more likely to recidivate than matched counterparts. Yet our analysis would be incomplete if we did not offer potential explanations for these specific findings in Montana. To that end, we address two dominant explanations within the literature for the effects of early release on offender recidivism. First, we assess the degree to which changes in one area of the justice system (i.e., release procedures) may have unintended effects on another (i.e., supervision

TABLE 5

Frailty-Adjusted Cox Proportional Hazard Model Predicting Time to Failure, Community-Matched Sample

Variable	*B*	SE	Relative Risk
Release Status			
Traditional parole from community corrections	—	—	—
Conditional release from community corrections	–0.133*	0.063	0.88
Offender			
Gender (1 = male)	0.403***	0.085	1.50
Mean release age	–0.014***	0.003	0.98
Native American (1 = yes)	0.566***	0.078	1.76
Criminal History			
Length of stay (months)	0.002	0.001	1.00
Total number of convictions	0.020	0.022	1.02
Drug conviction	0.070	0.081	1.07
Theft conviction	0.261**	0.086	1.30
Other nonviolent conviction	0.302***	0.082	1.35
Violent conviction	0.07	0.088	1.08
Frailty Gamma*			

Notes. Likelihood Ratio Test = 138**; *df* = 10. $R^2 = 0.24$; *n* = 2,844 releases.*$p < 0.05$.**$p < 0.01$. ***$p < 0.001$.

within the community). More specifically, we identify the concept of "criminal justice thermodynamics" as being a partial explanation for our findings. Second, we discuss the possibility that early release may have created a reduced deterrence effect in which the certainty and severity of serving a full sentence is undermined via early release. In doing so, we introduce an alternative explanation: Offenders are set up to fail through an early release procedure that does little to promote successful reintegration into society.

Criminal Justice Thermodynamics

The manifest function of back-end early release mechanisms is clear: to reduce correctional populations quickly (and quietly) in times of financial or spatial strain. Yet the latent functions of such an approach are numerous and potentially fatal to the overall intentions of administrators. Decisions made within the criminal justice *system* are likely to affect other actors and dynamics within that system. Walker (2006) described this process as the "law of criminal justice thermodynamics," which argues that actions within the justice system produce an equal and opposite reaction. The concept assumes a particular importance for studies of early release as the decision by administrators to release inmates early may be subverted by parole officers who cannot handle an increased caseload. In short, the early release of inmates can shift the burden of overcrowding unintentionally from an institutional setting to a community supervision setting.

Kelly and Ekland-Olson (1991) alluded to the concept of thermodynamics through their discussions of "administrative discretion effects." The authors noted that the increased volume of parole releases potentially could affect the ability of parole officers to monitor releases. This relationship could present itself in either direction. An increased caseload could lead to overburdened officers who are more likely to “return” offenders to lessen their workload (Wilson, 2005). Alternatively, an increased caseload may lead to reduced surveillance of any one particular offender and, thus, to fewer observations of parole violations. Furthermore, it is likely that parole officers face increased scrutiny from individuals who may be apprehensive about the early release process. In any event, it is clear that the unscheduled early release of offenders would affect the ability of parole officers to do their job, and Kelly and Ekland-Olson (1991) speculated that parole officers may have responded to heightened public concern through the increased use of technical violations for at least one cohort studied.

These concerns point to a larger issue in that widespread discretion occurs in the administration of parole supervision (Austin, 2001; Bottomley, 1990). Indeed, when a violation occurs, the officer may return the parolee to prison, make note of the transgression and strengthen supervision, or take no action at all (Travis and Lawrence, 2002). The label of “early release” likely complicates these decisions even more. Are early releasees granted more leniency given the overall system goal of reducing correctional populations, or are they granted *less* leniency given the additional burden they place on a particular officer? Thus, it is important to determine whether early release outcomes are a product of the behavior of releases or by formal and informal system practices (Wilson, 2005).

The question then becomes to what extent are the results from Montana a function of criminal justice thermodynamics? Unfortunately without qualitative information from parole officers, we cannot determine definitively that adjustments were occurring. It is entirely possible, if not likely, that the label of being a conditional release may have influenced the differences in technical violations across the four groups—specifically for those released early from prison, which may invoke a different response than those released early who were already in the community. Thus, whereas early release procedures may not influence recidivism rates *directly* through the commission of new criminal acts or violations of parole by those released early, it is possible that recidivism rates may increase because of adjustments made by parole officers. Future research would do well to explore the impact of policies (such as early release) on the day-to-day operations of key criminal justice officials. Another observation does, however, merit mentioning. As noted, judges in Montana were modifying their sentencing practices as a result of the early release program. The law of criminal justice thermodynamics was thus taking place on the *front end* of the system as judges were reluctant to turn over placement decisions to the MDOC.

Reduced Deterrence

Another possible explanation for the relationship between early release and recidivism identifies a reduction in sentence length as leading to a weakened deterrent effect of criminal justice sanctions. This perspective is not a novel idea as indeterminate sentencing was attacked by the political right in the 1970s for being too lenient on offenders (Cullen and Gilbert, 1982). Specifically, the charge was made that the existing system was "soft on crime" by mitigating the harshness of penal sanctions through early release policies such as parole. The certainty and severity components of deterrence theory were therefore diluted when offenders weighed the costs and benefits of engaging in criminal behavior. Not surprisingly, the same arguments remain persuasive today in evaluating the consequences of even more conspicuous early release practices.

A major component of the explanation of findings by Kelly and Ekland-Olson (1991) involved the possibility of a reduced deterrent effect among the parolees released early. In at least one cohort studied, the authors concluded that repetitious offending (i.e., engaging in the same crime as previously incarcerated for, specifically by property offenders and burglars in particular) was potentially indicative of reduced deterrence in that property offenders were most likely to make a rational calculation before engaging in criminal behavior. A more detailed analysis of the paroled property offenders by Joo et al. (1995) again suggested that the increase of repetitious offenders returning to prison from 1984 to 1987 was likely caused by a reduced deterrence effect. Furthermore, in those subsequent analyses, the specific early released property cohort was more likely to be returned to prison for another property offense within 3 years compared with a traditional release property cohort. The authors suggested that early release may lead to a "reduction in the deterrence influence on parolees," which may have accounted for the discrepancies in recidivism across cohorts and may have contributed to the overall increase in Texas incarceration rates (Joo et al., 1995: 405).

The preceding conclusions were largely speculative, and, appropriately, alternative explanations are equally prone to additional questioning.[15] Still, several observations lend to the credence of other interpretations of these findings. A sizeable body of research questions the empirical support of deterrence theory in general (Pratt, Cullen, Blevins, Daigle, and Madensen, 2006), with property offenders in particular unlikely to be swayed heavily in either direction by penalty severity on its own (Decker, Wright, and Logie, 1993). This knowledge coincides with research that suggests that longer sentences produce little gain in terms of reduced recidivism from both an incapacitation and a deterrence perspective (Austin, 1986, 2010; Blumstein, 1988; Blumstein and Beck, 1999; cf. DeJong, 1997). Thus, it is questionable whether the reduction of a previous sentence by a maximum of 90 days had any effect on future criminal behavior decision making by paroled property offenders. Additionally, although we are mindful of the body of work that suggests offenders do not specialize in types of offending (e.g., Piquero, Farrington, and Blumstein, 2007; cf. DeLisi et al., 2011; McGloin, Sullivan, Piquero, and Pratt, 2007), it is important to note that property offenders often are more likely to recidivate—independent of whether they were released early (Langan and Levin, 2002; Spivak and Damphousse, 2006; Wilson, 2005).

15. Indeed, the authors note that it is possible that shifts in the economy could be responsible for their findings.

Indeed, our findings from the community release sample analysis suggested that a prior theft or "other" nonviolent conviction increased the likelihood of recidivism.

Yet the findings of the study by Joo and colleagues were robust and were replicated here in the current study: Early release offenders from prison were more likely to fail, and to do so more quickly, than non–early release offenders from prison. Therefore, we wish to advance an alternative interpretation for the increased likelihood of reoffending by early releases. It is likely that these offenders were thrust back into society with little time to prepare for successful reintegration. Accordingly, the increased stress and trauma of returning to society unprepared may lead to subsequent criminal behavior (Richards and Jones, 1997; but see Minor and Courlander, 1979).[16] Joo et al. (1995: 407) recognized this plausibility: "A sudden, unexpected release for those in the [Prison Management Act] PMA cohort may have meant fewer prior arrangements for life on the outside and thus reduced support resources." We believe that an attention to the literature on successful reentry will aid both researchers and correctional officials in their efforts to understand better the consequences of early release programs.

Offender Reentry

The shock value of the statistic that more than 700,000 offenders will return to communities each year is beginning to dissipate (West et al., 2010). Yet an overlooked component of this oft-cited fact is the notion that the group of offenders returning is fundamentally different than in years past (Petersilia, 2003). These individuals have less programming available to them in prison, have fewer connections with community-based structures, and are overall less prepared to reintegrate back into society successfully (Petersilia, 2003; Travis and Petersilia, 2001). Furthermore, nearly three of four releasees never see a parole board and, thus, often are never required to demonstrate a plan for successful reintegration (Travis and Lawrence, 2002). Whereas previous studies (e.g., Austin, 1986; Sims and O'Connell, 1985) found early releasees to be no more likely to recidivate than non–early releasees, it is possible that the changing nature of the release cohort is responsible for our contrary findings from Montana.

States vary to a wide degree in terms of prerelease programming and resources provided (e.g., money and transportation) to inmates before and during release (Austin, 2001), but it is not uncommon for an inmate to have an experience similar to that described by Richards and Jones (1997; see also Richards, Austin, and Jones, 2004):

> These men walked out of prison wearing old, worn out prison uniforms, carrying a cardboard box containing their personal belongings, with $5 gate

16. We again make the admission that most reincarcerations in the current analyses were for technical violations. This fact does not, however, preclude the possibility that these technical violations were a result of being unprepared for reentry (e.g., inability to find employment or drug use to cope with stress).

> money in their pockets. . . . Some of these men, particularly those who served a long time in the penitentiary, had not been required to pay rent, purchase food, or look for employment in years. The problem was that they had not been properly prepared for release from prison. (pp. 10, 12)

It should come as no surprise, therefore, that parolees often cannot meet the relatively lofty goals of their parole requirements. Several structural impediments are in place that make it difficult, if not impossible, to maintain a conventional lifestyle after institutionalization. Finding employment is a particularly difficult requirement for parolees to meet (Richards et al., 2004)—especially with the stain of a criminal record warding off potential employers (Pager, 2007). The broader point is that whereas a technical violation could represent a serious form of misconduct, it also could indicate an issue that would do little to jeopardize public safety while thwarting plans for overcrowding reduction or budget savings.

Travis (2005) identified reintegration as the product of both the individual reentering society as well as the social context into which he or she has been released. To be sure, an emerging body of literature is indicating that ecological factors (e.g., concentrated disadvantage) play an important role in reoffending (Gottfredson and Taylor, 1986, 1988; Kubrin and Stewart, 2006; Mears, Wang, Hay, and Bales, 2008; Reisig, Bales, Hay, and Wang, 2007). This body of research also is notable for uncovering a differential impact of ecological factors on minorities (Kubrin, Squires, and Stewart, 2007), which assumes a particular importance given our findings on Native American offenders. Additionally, some support exists for the idea that community treatment programs that are high in therapeutic integrity may ease the transition process and lead to a reduction in recidivism (Lowenkamp, Latessa, and Smith, 2006). Unfortunately, the social services that are so vital to successful reintegration often are unavailable in the disadvantaged areas where they are needed most (Hipp, Jannetta, Shah, and Turner, 2011; Hipp, Petersilia, and Turner, 2010). Our findings confirm the need for a reinforcement of the link between prerelease preparation and postrelease supervision (Travis and Lawrence, 2002). Although we lack information on the exact location offenders returned to, it is plausible that the conditional release from prison group was placed at a disadvantage in linking up with social services because of inadequate preparation for release.

Implications

The lessons learned from Montana begin with the idea that early release from prison specifically produces several unforeseen consequences beyond that of reducing correctional populations.[17] Yet our research should not be taken to indicate a requirement for offenders

17. The MDOC made several adjustments as a result of the preliminary findings of the early release procedures. An overall change in placement philosophy put more offenders under community supervision and less behind bars. Indeed, although the total supervised population increased by

necessarily to serve the totality of their sentences. Nor does it advance the idea that an increase in incarceration and its severity would increase public safety. What we argue is that both short-term and long-term processes designed to alleviate correctional strain need to be viewed from a reintegration perspective. To that end, we offer three broad and interrelated implications based on the findings of the current analyses.

1. ***Early release as a short-term fix.*** It would be unwise to suggest that early release procedures should be done away with entirely. Immediate pressures on the U.S. correctional system may produce disastrous consequences for the safety of inmates and correctional officers if they are not ameliorated quickly (but see Franklin, Franklin, and Pratt, 2006). Furthermore, budget deficits have left administrators with few alternatives. The challenge, then, is to provide correctional officials with the best possible information going forward. In the most general sense, we join previous scholars (Austin, 1986; Sims and O'Connell, 1985) in stressing that early release mechanisms should be conceived of as only a short-term remedy for a long-term problem.

Montana, in particular, experienced a significant increase in the annual 3-year recidivism rate of releases as a result of the early release program. Thus, although the procedure solved the initial dilemma, it created additional correctional quagmires down the road. Guided by the next two recommendations, correctional administrators need to couple emergency short-term relief procedures with more long-term strategies for dealing with fluctuations in prison admissions and budget availabilities (Blumstein, 1988). Although this approach includes the need for better projections based on demographics in addition to emergency management planning, it also requires administrators to examine the relationships between prison (re)admissions and effective parole and reentry policies. Short-term emergency release procedures can be successful, but our research indicates that more attention needs to be paid to the transition process from both a supervision and a reintegration perspective.

2. ***Reconsider the nature of parole.*** The major functions of parole are twofold—it can be conceived of as a discretionary release mechanism or as postrelease supervision for ex-offenders. The current analysis has implications for both functions. First, evidence of success (i.e., higher survival rates) was found for the cohorts that were selected for release from prison via traditional parole. Indeed, only 30.2% of traditional parole releases from a secure facility failed within a 3-year period. Our findings support the contention that

approximately 20.0% since FY 2004, the percentage of offenders in hard cells dropped from 22.0% in FY 2004 to 18.9% in FY 2008. Additionally, DOC Commitments are now eligible to be placed on direct community supervision (similar to probation), whereas prior to 2003, they were only placed in prison, prerelease, or intensive supervision. Finally, conditional release is still employed but only for the DOC Commitments supervised in community programs. During the past 5 fiscal years, the MDOC reported that only 35.0% of these conditional releases have returned to prison—a finding that is consistent with the current analyses in regard to the initial early release cohort from community settings. Overall, the 3-year return rate to prison in Montana dropped from a high of 47.0% in 2003 to 44.0% in 2005 (Montana Department of Corrections, 2009).

success on parole is increasing in some jurisdictions (Austin, 2001), and they suggest that an increase in parole grant rates may do little to increase the rates of criminal behavior (Wilson, 2005).

Second, in Montana, more than 50% of prison admissions are parole violators, which parallels a national trend of an immense growth in parole violations during the past 20 years (Blumstein and Beck, 1999; Travis and Lawrence, 2002). Austin (2001) argued that the large variation in parole violations in the United States is caused by diverse policies among states in the imposition of technical violations specifically. To be sure, some states such as California return a sizeable proportion of parolees to prison for reasons other than a new arrest or conviction (Hughes, Wilson, and Beck, 2001). Nearly all the ex-offenders (early and traditional released) within the current study were returned for technical violations; yet significant differences existed across the groups that resulted in early release offenders being more likely to return to custody. Although the commission of a new crime clearly necessitates a strong formal response, relatively minor violations (e.g., failure to report and failed urinalysis) could be handled less punitively. Instead, a system of graduated sanctions could be created that resorts to reincarceration as a last option for repeat violators (Travis and Petersilia, 2001; see also Makkai and Braithwaite, 1994).[18] Such an approach is likely to reduce correctional populations as well as the costs associated with reincarcerating ex-offenders (Travis and Lawrence, 2002), and potentially it could provide more options and guidelines to parole officers that are charged with managing the extra burden of early release offenders.

3. ***Prepare for reentry.*** The preceding suggestion comes as part of a broader argument to overhaul the parole system (see Travis and Petersilia, 2001). Rather than focus on parole as an extended sentence of supervision for offenders, it could be conceived of as a managed reentry mechanism with an explicit focus on successful reintegration. The number of ex-offenders who are released back into communities without any sort of supervision (and, thus, no support for successful reentry) is increasing. More than 100,000 offenders had unconditional releases in 2000, with a few states releasing more than half of prisoners without supervision requirements—some of which directly from maximum-security institutions (Austin, 2001; Travis and Lawrence, 2002). Although clearly a lack of supervision could potentially miss the opportunity to detect new crimes on the behalf of ex-offenders, it also neglects the possibility of preventing future transgressions through better preparing inmates for a successful transition back into society.

Our research indicated that the early releasees from prison may have not been adequately prepared to reenter society, with the expedited release likely the epitome of instances in which prisoners are unaware of their discharge date (Richards and Jones, 1997). Reentry

18. It is likely that offenders, often upset at having their parole revoked for minor violations after "doing good" otherwise, would be supportive of such an approach as well (see Richards et al., 2004: 253).

should thus begin within prison walls through specific planning for each offender.[19] After discharge, a system of managed reentry could "seize the moment of release" (Travis, 2005) by providing ex-offenders with the support needed to perform simple but necessary tasks such as obtaining an identification card. The system should be front loaded, with the bulk of services concentrated within the first 6 months of release (Petersilia, 2003) and should provide offenders with the opportunity to accomplish several requirements (e.g., housing, employment) in one location (Travis, 2005). Additionally, the opportunity for individuals to "graduate" from parole early would assure that resources were reserved for the ex-offenders most vulnerable for a return to crime (Petersilia, 2007).

Early release procedures will undoubtedly become more commonplace in corrections as a means to overcome budget shortfalls and prison crowding. It is therefore necessary that precautions be taken to ensure that offenders are fully prepared to succeed after reentry into society. A full appreciation for the complexities of early release from a reintegration perspective could indeed serve to save money and correctional space while increasing public safety through the reduction of future offenses. It also could lead to additional benefits such as decreases in child abuse, family violence, and community disorganization (Petersilia, 2001), and it would create an opportunity to save considerable time and money to treat social ills through the offender population (Travis, 2005). Indeed, offenders represent a significant cross section of individuals in need of health care, childcare, and general counseling and treatment. Perhaps most importantly, it could challenge the idea that even three decades later, we still do not know what works in reducing recidivism. All eyes will turn to California and states in similar situations to observe the outcome of various responses to the budget and spatial crises in corrections. In short, the manner in which these states choose to handle early release now could produce consequences that would reverberate off correctional walls for years to come.

References

Archibold, Randal C. 2010. California, in financial crisis, opens prison doors. *New York Times*. March 24, section A, 14.

Associated Press. 2002, November 12. State deficit: Department of Corrections. *Helena Independent Record (Helena, MT)*. Retrieved March 14, 2007 from helenair.com.

Austin, James. 1986. Using early release to relieve prison crowding: A dilemma in public policy. *Crime & Delinquency*, 32: 404–502.

Austin, James. 2001. Prisoner reentry: Current trends, practices, and issues. *Crime & Delinquency*, 47: 314–334.

Austin, James. 2010. Reducing America's correctional populations: A strategic plan. *Justice Research and Policy*, 12: 9–40.

19. See, for example, Wilkinson's (2001) discussion of the "The Ohio Plan for Productive Offender Reentry and Recidivism Reduction."

Blumstein, Alfred. 1988. Prison populations: A system out of control? In (Michael Tonry and Norval Morris, eds.), *Crime and Justice: A Review of Research*. Chicago, IL: University of Chicago Press.

Blumstein, Alfred and Allen J. Beck. 1999. Population growth in U.S. prisons, 1980–1996. In (Michael Tonry and Joan Petersilia, eds.), *Prisons*. Chicago, IL: University of Chicago Press.

Bottomley, Keith. 1990. Parole in transition: A comparative study of origins, developments, and prospects for the 1990s. In (Michael Tonry and Norval Morris, eds.), *Crime and Justice: A Review of Research*. Chicago, IL: University of Chicago Press.

Collett, David. 1994. *Modeling Survival Data in Medical Research*. London: Chapman and Hall.

Cullen, Francis T. and Karen E. Gilbert. 1982. *Reaffirming Rehabilitation*. Cincinnati, OH: Anderson.

Cullen, Francis T., John P. Wright, and Brandon K. Applegate. 1996. Control in the community: The limits of reform? In (Alan T. Harland, ed.), *Choosing Correctional Options that Work: Defining the Demand and Evaluating the Supply*. Thousand Oaks, CA: Sage.

Decker, Scott, Richard Wright, and Robert Logie. 1993. Perceptual deterrence among active residential burglars: A research note. *Criminology*, 31: 135–147.

DeJong, Christina. 1997. Survival analysis and specific deterrence: Integrating theoretical and empirical models of recidivism. *Criminology*, 35: 561–576.

DeLisi, Matt, Kevin M. Beaver, Kevin A. Wright, John P. Wright, Michael G. Vaughn, and Chad R. Trulson. 2011. Criminal specialization revisited: A simultaneous quantile regression approach. *American Journal of Criminal Justice*, 36: 73–92.

Eichman, Charles J. 1966. *The Impact of the Gideon Decision Upon Crime and Sentencing in Florida: A Study of Recidivism and Socio-Cultural Change*. Tallahassee, FL: Division of Correction.

Ekland-Olson, Sheldon, William R. Kelly, Hee-Jong Joo, Jeffrey Olbrich, and Michael Eisenberg. 1993. *Justice Under Pressure: A Comparison of Recidivism Patterns Among Four Successive Parole Cohorts*. New York: Springer-Verlag.

Engel, Len, John Larivee, and Richard Luedeman. 2009. Reentry and the economic crisis: An examination of four states and their budget efforts. *Corrections Today*, 71: 42–45.

Franklin, Travis W., Cortney A. Franklin, and Travis C. Pratt. 2006. Examining the empirical relationship between prison crowding and inmate misconduct: A meta-analysis of conflicting research results. *Journal of Criminal Justice*, 34: 401–412.

Gainey, Randy R., Brian K. Payne, and Mike O'Toole. 2000. The relationships between time in jail, time on electronic monitoring, and recidivism: An event history analysis of a jail-based program. *Justice Quarterly*, 17: 733–752.

Gottfredson, Stephen D. and Ralph B. Taylor. 1986. Person-environment interactions in the prediction of recidivism. In (James M. Byrne and Robert J. Sampson, eds.), *The Social Ecology of Crime*. New York: Springer Verlag.

Gottfredson, Stephen D. and Ralph B. Taylor. 1988. Community contexts and criminal offenders. In (Tim Hope and Margaret Shaw, eds.), *Communities and Crime Reduction*. London, UK: Her Majesty's Stationery Office.

Guzman, Carolina, Barry Krisberg, and Chris Tsukida. 2008. *Accelerated Release: A Literature Review*. Oakland, CA: National Council on Crime and Delinquency.

Harrison, Paige A. and Allen J. Beck. 2004. *Prisoners in 2003*. Washington, D.C.: Bureau of Justice Statistics.

Hipp, John R., Jesse Jannetta, Rita Shah, and Susan Turner. 2011. Parolees' physical closeness to social services: A study of California parolees. *Crime & Delinquency*, 57: 102–129.

Hipp, John R., Joan Petersilia, and Susan Turner. 2010. Parolee recidivism in California: The effect of neighborhood context and social service agency characteristics. *Criminology*, 48: 947–979.

Hougaard, Philip. 1995. Frailty models for survival data. *Lifetime Data Analysis*, 1: 255–273.

Huebner, Beth M. and Mark T. Berg. 2011. Examining the sources of variation in risk for recidivism. *Justice Quarterly*, 28: 146–173.

Hughes, Timothy, Doris James Wilson, and Allen J. Beck. 2001. *Trends in State Parole, 1990–2000*. Washington, D.C.: Bureau of Justice Statistics.

Joo, Hee-Jong, Sheldon Ekland-Olson, and William R. Kelly. 1995. Recidivism among paroled property offenders released during a period of prison reform. *Criminology*, 33: 389–410.

Kelly, William R. and Sheldon Ekland-Olson. 1991. The response of the criminal justice system to prison overcrowding: Recidivism patterns among four successive parolee cohorts. *Law & Society Review*, 25: 601–620.

King, Ryan D., Michael Massoglia, and Ross MacMillan. 2010. The context of marriage and crime: Gender, the propensity to marry, and offending in early adulthood. *Criminology*, 45: 33–65.

Kubrin, Charis E., Gregory D. Squires, and Eric A. Stewart. 2007. Neighborhoods, race, and recidivism: The community-reoffending nexus and its implications for African Americans. *SAGE Race Relations Abstracts*, 32: 7–37.

Kubrin, Charis E. and Eric A. Stewart. 2006. Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology*, 44: 165–197.

Lane, Michael P. 1986. A case for early release. *Crime & Delinquency*, 32: 399–403.

Langan, Patrick A. and David J. Levin. 2002. *Recidivism of Prisoners Released in 1994*. Washington, D.C.: Bureau of Justice Statistics.

Lowenkamp, Christopher T., Edward J. Latessa, and Paula Smith. 2006. Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. *Criminology & Public Policy*, 5: 575.

Lynch, James P. 2006. Prisoner reentry: Beyond program evaluation. *Criminology & Public Policy*, 5: 401–412.

Makkai, Toni and John Braithwaite. 1994. The dialectics of corporate deterrence. *Journal of Research in Crime and Delinquency*, 31: 347–373.

Malak, Patricia A. 1984. *Early Release*. Denver, CO: Department of Local Affairs, Division of Criminal Justice.

Maxwell, Sheila Royo. 2005. Rethinking the broad sweep of recidivism: A task for evaluators. *Criminology & Public Policy*, 4: 519–526.

McGloin, Jean Marie, Christopher J. Sullivan, Alex R. Piquero, and Travis C. Pratt. 2007. Local life circumstances and offending specialization/versatility: Comparing opportunity and propensity models. *Journal of Research in Crime and Delinquency*, 44: 321–346.

Mears, Daniel P., Xia Wang, Carter Hay, and William D. Bales. 2008. Social ecology and recidivism: Implications for prisoner reentry. *Criminology*, 46: 301-340.

Minor, W. William and Michael Courlander. 1979. The postrelease trauma thesis: A reconsideration of the risk of early parole failure. *Journal of Research in Crime and Delinquency*, 16: 273–293.

Montana Board of Pardons and Parole. 2010a. Administrative Rule 20.25.305. *Administrative Rules*. Helena, Montana. Retrieved December 22, 2010 from mt.gov/bopp/adminrules/Admin_rule_20.25.305.asp.

Montana Board of Pardons and Parole. 2010b. Administrative Rule 20.25.306. *Administrative Rules*. Helena, Montana. Retrieved December 22, 2010 from mt.gov/bopp/adminrules/Admin_rule_20.25.306.asp.

Montana Board of Pardons and Parole. 2010c. Administrative Rule 20.25.801. *Administrative Rules*. Helena, Montana. Retrieved December 22, 2010 from mt.gov/bopp/adminrules/Admin_rule_20.25.801.asp.

Montana County Attorneys Association. 2002, December 13. Release policy dangerous. *Helena Independent Record (Helena, MT)*. Retrieved March 13, 2007 from helenair.com.

Montana Department of Corrections. 2009. *2009 Biennial Report*. Helena, MT: Author.

Montana Department of Corrections. 2010a. *2011 Biennial Report*. Helena, MT: Author.

Montana Department of Corrections. 2010b. Policy No. DOC 4.6.2: Release and Transfer Procedures. Helena, Montana. Retrieved December 22, 2010 from cor.mt.gov/content/Resources/Policy/Chapter4/4-6-2.pdf.

Pager, Devah. 2007. *Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration*. Chicago, IL: University of Chicago Press.

Petersilia, Joan. 2001. Prisoner reentry: Public safety and reintegration challenges. *The Prison Journal*, 81: 360–375.

Petersilia, Joan. 2003. *When Prisoners Come Home: Parole and Prisoner Reentry*. New York: Oxford University Press.

Petersilia, Joan. 2007. Employ behavioral contracting for "earned discharge" parole. *Criminology & Public Policy*, 6: 807–814.

Piquero, Alex R., David P. Farrington, and Alfred Blumstein. 2007. *Key Issues in Criminal Career Research: New Analyses of the Cambridge Study in Delinquent Development*. New York: Cambridge University Press.

Pratt, Travis C., Francis T. Cullen, Kristie R. Blevins, Leah Daigle, and Tamara Madensen. 2006. The empirical status of deterrence theory: A meta-analysis. In (Francis T. Cullen, John P. Wright, and Kristin R. Blevins, eds.), *Taking Stock: The Empirical Status of Criminological Theory—Advances in Criminological Theory*. New Brunswick, NJ: Transaction.

Reisig, Michael D., William D. Bales, Carter Hay, and Xia Wang. 2007. The effect of racial inequality on Black male recidivism. *Justice Quarterly*, 24: 408–434.

Richards, Stephen C., James Austin, and Richard S. Jones. 2004. Thinking about prison release and budget crisis in the blue grass state. *Critical Criminology*, 12: 243–263.

Richards, Stephen C. and Richard S. Jones. 1997. Perpetual incarceration machine: Structural impediments to postprison success. *Journal of Contemporary Criminal Justice*, 13: 4–22.

Rosenbaum, Paul and Donald Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70: 41–55.

Shoukri, Mohamed M. and Cheryl A. Pause. 1999. *Statistical Methods for Health Sciences*, 2nd Edition. Boca Raton, FL: CRC Press.

Sims, Brian and Jack O'Connell. 1985. *Early Release: Prison Overcrowding and Public Safety Implications*. Olympia, WA: Office of Financial Management.

Slaughter, B. 2002, December 1. A look at state prison policy. *Helena Independent Record (Helena, MT)*. Retrieved March 13, 2007 from helenair.com.

Spivak, Andrew L. and Kelly R. Damphousse. 2006. Who returns to prison? A survival analysis of recidivism among adult offenders released in Oklahoma, 1985–2004. *Justice Research and Policy*, 8: 57–88.

Spohn, Cassia and David Holleran. 2002. The effect of imprisonment on recidivism rates of felony offenders: A focus on drug offenders. *Criminology*, 40: 329–358.

Stuart, Elizabeth A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25: 1–21.

Tonry, Michael and Mary Lynch. 1996. Intermediate sanctions. In (Michael Tonry, ed.), *Crime and Justice: A Review of Research*. Chicago, IL: University of Chicago Press.

Travis, Jeremy. 2005. *But They All Come Back: Facing the Challenges of Prisoner Reentry*. Washington, D.C.: The Urban Institute Press.

Travis, Jeremy and Sarah Lawrence. 2002. *Beyond the Prison Gates: The State of Parole in America*. Washington, D.C.: The Urban Institute Press.

Travis, Jeremy and Joan Petersilia. 2001. Reentry reconsidered: A new look at an old question. *Crime & Delinquency*, 47: 291–313.

U.S. Census Bureau. 2010. *2009 Population Estimates*. Retrieved December 19, 2010 from factfinder.census.gov.

Walker, Samuel. 2006. *Sense and Nonsense about Crime and Drugs: A Policy Guide*, 6th Edition. Belmont, CA: Thomson Higher Education.

West, Heather C., William J. Sabol, and Sarah J. Greenman. 2010. *Prisoners in 2009*. Washington, D.C.: U.S. Department of Justice.

Wilkinson, Reginald A. 2001. Offender reentry: A storm overdue. *Corrections Management Quarterly*, 5: 46–51.

Wilson, James A. 2005. Bad behavior or bad policy? An examination of Tennessee release cohorts, 1993–2001. *Criminology & Public Policy*, 4: 485–518.

Court Cases Cited

Hoptowit v. Ray, 682 F. 2nd, 1237 (1982).
Ruiz v. Estelle, 503 F. Supp. 1265 (1980).

Statute Cited

Prison Management Act, Texas Rev. Civ. Stat. Ann. art. 6184 (1983).

Kevin A. Wright is an assistant professor in the School of Criminology and Criminal Justice at Arizona State University. His current research focuses on the importance of ecological context for offender rehabilitation and reintegration, and his work has appeared in *Homicide Studies, Journal of Offender Rehabilitation,* and *Justice Quarterly.*

Jeffrey W. Rosky is an assistant professor of Criminal Justice in the College of Health & Public Affairs at the University of Central Florida where his research interests include correctional treatment programs and sex offender supervision and treatment. Prior to his academic career, he worked as researcher in the Montana, Colorado, and Florida correctional systems and as a statistician in public health and infectious disease research.

POLICY ESSAY

MONTANA EARLY RELEASE PROGRAM

What is my left hand doing?

The need for unifying purpose and policy in the criminal justice system

Megan Kurlychek
University at Albany, State University of New York

Sometimes when reading research in my field, I am reminded of an old saying that goes something like this: "Your left hand doesn't know what your right hand is doing." Indeed, I am sure many of us who work within or study the criminal justice "system" can relate to this saying. Although the term "system" (deriving from the Greek root *synthithemi*, meaning "I put together") implies a network of parts put together to work toward a common goal, it often is the case in the criminal justice system that the parts are in conflict rather than in unison—at times working independently from one another and at times working in direct opposition. Indeed, even when the parts do become aware of one another, as Wright and Rosky (2011, this issue) note with the idea of thermodynamics, the right hand may serve only to slap the left hand, thus thwarting its efforts.

Although a formal discussion of the opposing philosophies of justice is beyond the scope of this essay, as a brief introduction to the insanity that surrounds these conflicting missions, I provide a summary of the most commonly cited goals of criminal justice:

1. To provide "just deserts," meaning the appropriate punishment for an offense committed.
2. To incapacitate the most dangerous or "risky" offenders to protect society from their potential future crimes.
3. To rehabilitate offenders into law-abiding citizens, thereby making it safe to return them to society.
4. To restore victims and communities to the greatest extent possible to their precrime status.

Direct correspondence to Megan Kurlychek, School of Criminal Justice, University at Albany, State University of New York, 135 Western Avenue, Albany, NY 12222 (e-mail: mkurlychek@albany.edu).

DOI:10.1111/j.1745-9133.2011.00771.x

Add to this list of overarching goals the immediate day-to-day struggles of system actors such as collecting the accurate information upon which to make decisions, managing risky populations in overcrowded and underfunded conditions, and navigating the bureaucratic red tape stuck to every turn, and one still can only begin to understand the conditions under which correctional, probation, and parole officers attempt to create order and purpose.

And if the system seems confusing to its workers, imagine the conundrum faced by its clients. Are they being punished or rehabilitated? Do they need protection from a system that is out to get them, or do they need to trust system workers who are there to help them? And perhaps, most importantly for the study and discussion at hand, is the parole officer there to help them navigate the difficulties of reentry or merely to catch them for every mistake made?

I begin by highlighting these various and confusing purposes of justice, particularly the function of parole, because I believe this paradox to be at the heart of the problem highlighted in the research by Wright and Rosky (2011). As a reminder of the starting point, I provide a brief summary and my interpretation of the findings of this study. Wright and Rosky examine early release policies in Montana that were developed to reduce prison overcrowding. However, rather than reducing overcrowding, the findings suggest that the policy might only serve to increase recommitments to prison as the early release group was found to be twice as likely to "fail" as the group with an on-time release from prison.

It is important to note that for this particular study by Wright and Rosky (2011), a failure was defined as recommitment to custody and not as a new arrest. Indeed, all four groups in the sample (traditional prison release, early prison release, traditional community release, and early community release) were *equally* as likely to experience a new arrest (approximately 5% of each group). Thus, the difference in "failures" was entirely driven by recommitments to prison for technical violations of parole.

This crucial finding is in my mind the most prominent takeaway message of the Wright and Rosky (2011) study for several reasons. First, it is clearly contrary to the proposed argument that early release leads to an increase in crime because it sends a message to offenders that they will receive an easy sentence (e.g., reduced deterrence). If policies of early release really sent this message to offenders, the study would find those released early to be committing new crimes at a rate higher than the other groups, not merely committing technical parole violations. Second, it shifts the primary question—and thus the potential answer—raised by the study. Specifically, the question at hand becomes not one of why releasees commit new crimes but of why the group released early from a prison setting was more likely than the other three groups to be recommitted for parole violations. In response to this question, I propose three plausible explanations:

1. The early release group was less prepared to succeed on the outside and therefore actually did "fail" on parole more often for technical reasons such as an inability to find

employment, unsatisfactory housing arrangements, and failure to report to the parole officer.

2. The parole officers did not agree with the early release decision and are "correcting" what they consider an error in another part of the system by quickly returning these releasees to prison (e.g., system thermodynamics as posited by Wright and Rosky).
3. The early release group was placed under more scrutiny by parole officers because of their status and perhaps because of public pressure, thus leading to the revocation of parole for even the most minor infractions.

Clearly, the first explanation is the only true measure of offender behavior, whereas the second is a pure measure of "system" behavior. The third then offers a possible combination of offender behavior, system actor behavior, as well as extending to pressures (real or perceived) from outside the system itself. Although it is impossible from the Wright and Rosky (2011) article to disentangle what portion of the failures might be related to any one cause, my hunch is that all contribute to the final outcome. Thus, any policy designed to reduce "failures" must incorporate components of all. I offer some thoughts and suggestions related to each explanation in the subsequent discussion.

Offender Behavior: Preparing for Reentry

If the finding is driven by true behavioral differences—that is, offenders in the prison early release sample were more likely than others to behave in a manner that required the revocation of parole—then why, one might ask, did the early release group from the community setting not suffer the same fate? Here, I offer two thoughts. First, the community group starts off less disconnected than the prison group. That is, they are living in a community setting with at least some connections to society. As a result, there is less of a divide to be crossed to reintegrate into a society.

Second, Wright and Rosky (2011) inform the reader that 78% of the community sample comes from prerelease centers. Although they do not offer a full explanation of a prerelease center within the article, the Montana Department of Corrections describes it as follows:

> "A facility in the community designed to ease the transition of an offender from a correctional institution to living independently in the community while providing treatment, education, counseling, job training and placement, and transitional living opportunities." (Montana Department of Corrections)

Thus, although the community early release sample was indeed released prior to the full expiration of their sentence, according to Wright and Rosky (2011), it seems that for most, their experience immediately preceding release was one directly aimed at helping them succeed on the outside.

Clearly the notion of preparation for reentry is not a novel one. Much research indicates the importance of keeping ties to the community whenever possible and offering prerelease services to assist with reentry (BI Incorporated, 2002; Cook, 2000; Fulp, 2001; Petersilia, 2003; Seiter and Kadela, 2003; Travis and Petersilia, 2006). However, not only are these suggestions often not taken, when they are taken, they often are done so in piecemeal fashion. For example, one facility may experiment with a new aftercare program or one jurisdiction might open a new halfway house targeting a specific type of inmate. It is rarer, however, to see such efforts incorporated into any systemic model of reform. As reported by Petersilia (2003), our nation has indeed taken the opposite approach by continually reducing prison programming. Recent estimates suggest that less than 35% of state prisoners receive educational services, less than 30% receive drug or alcohol treatment, and only approximately 12% receive any prerelease services.

At this point in the Wright and Rosky (2011) study, however, we reach another illogicality. If prerelease preparation and reentry services are important, why do the offenders in either of the community samples (traditional or early release) fair *worse* than the traditional prison releasees in terms of technical parole violations? Could it be that serving a longer prison sentence is truly a deterrent? If so, one would think this group also would commit less crimes than the others, but they do not. Here, then I turn to look for answers in system, rather than in offender, behavior.

Parole Officer Behavior: Helping or Hindering?

The concept of parole has traditionally encompassed both the decision to release a prisoner and the function of supervising the offender for a given period of time after release. The first component—the authority to release—is particularly relevant to the idea of system thermodynamics. This idea proposed by Walker (2006), suggests that when one hand of the system makes a decision that the other does not agree with, ample room exists for correction. For example, if a police officer makes an arrest, the district attorney does not need to bring the case to trial or may alter the initial charges. If the case is brought to trial, a conviction may not occur, plea bargains can further alter the charges applied, and/or a judge may give a sentence that deviates from the average or normative sentence for the given offense (Kramer and Ulmer, 2002). Furthermore, if a sentence is given that is considered too lenient or too harsh, correctional system workers and parole officers have traditionally had mechanisms to "correct" this sentence (Champion, 2005).

The implication here is that if the parole officer feels his or her authority has been circumvented in the release decision by another decision maker, the officer has an opportunity to regain authority through the discretionary use of parole violations. If this is the case, it would mean that the early release sample is not necessarily performing worse than the on-time release sample but that the officers are more likely to recommit the prerelease group for a behavior that they would overlook for an offender that served a greater portion of his or her sentence. I suggest that this issue may indeed be why the on-time prison release group faired the best of all groups in the sample. Moreover, with the increased ability to monitor parolees with advanced technologies such as electronic monitoring and the greater use of drug testing, Kleiman (1999) noted that parole officers now have greater power than ever to detect parole violations and to increase the rates of revocations when desired.

Another consideration of system behavior strikes more at the heart of the mission or philosophy of parole. For a moment, let us consider the origins of modern-day parole most often traced to the work of Alexander Maconochie at the prison on Norfolk Island in the mid-1800s. Maconochie is noted as a humanitarian who believed the purpose of prison should not be just punitive but rehabilitative as well. The system of "marks of commendation" he created allowed prisoners to progress through stages on their way to earning release. For example, in this system, the inmate progressed through stages from complete confinement, to work on a chain gang, to freedom to work in the community during the day. Finally, after earning the ticket of leave, the offender was provided with a small piece of land to farm and a home in which to live (Morris, 2002). Not surprisingly, this model met with great success and the stigma of "ex-convict" was replaced with the historical notion of "Maconochie's Gentlemen."

Although originally popular and copied in both Europe and America, this model of parole was later greatly abandoned as it was viewed as too lenient. Over time emphasis shifted from helping the offender succeed in society to protecting society from the offender. Today parole officers must balance the two potentially conflicting objectives of reentry and law enforcement (Travis and Petersilia, 2006): the first still clinging to its historical roots, and the second stemming from the more recent "get tough" approach to crime. The parole officer is then faced with a paradox of potential actions. Although he or she might want to assist a releasee in finding housing, employment, and other social services to help the releasee succeed as a free citizen, the parole officer maintains the authority to take the same liberties away.

Great differentials exist between jurisdictions regarding the emphasis placed on these conflicting goals and, I would assume, even within a jurisdiction between individual officers. Without clear mandates of which philosophy is to prevail, the practice of parole becomes subject to individual and situational pressures. So, as with the needed connections between institutional and reentry programming, again evidence of a needed connection is evident. Here, it is between the directive of parole and its actual practice. If early release policies are truly to reduce overcrowding, their mission must be clearly communicated to officers, and strategies must be taken to enmesh this policy within the greater mission of corrections—and not *just* parole. In the following section, I further argue that this system integration of mission, purpose, and programming becomes even more important when considering the public eye under which the criminal justice system operates.

Reentry in Context: Preparing the Public

Although the third explanation offered encompassed some aspects of both offender behavior (e.g., minor behaviors that may be subject to parole violation) and system behavior (e.g., readiness of parole officers to revoke parole at the slightest infraction), the impetus, or motivation, for this proposed synergy actually lies in pressures imposed from the outside: the environment to which the offender returns and the social pressures under which the parole officer functions.

For an instant let us put reality aside and assume that an offender has been adequately rehabilitated and prepared for reentry. Now, continuing on our fictional journey, let us assume that the mandate of parole has been redefined to emphasize assistance rather than enforcement and that all parole officers fully abide by this new mission. Even in this alternate reality, now ask yourself what success would an offender meet if returned to a world trained by the media to fear him or her and populated by obstacles to obtaining even the basic needs of survival such as housing and means to earn a living? This issue clearly brings us back to the world in which we live.

As Wright and Rosky (2011) rightly note, early release policies have not been popular with the public. The American public has to a great extent been taught to "fear" crime and the "criminal" (Zimring, 1998; Zimring and Hawkins, 1999). Fear of the offender is transformed from fiction into fact through the enactment of policies that serve to "protect" the public by restricting opportunities for those with a criminal record. Restrictions range from housing and employment to qualifying for student loans and even the right to vote. Such obstacles appear like land mines along the path to reentry. Their detonation leads to loss of opportunity and liberty to those who fall victim.

Interestingly, even amidst the media's campaign of fear, several recent public opinion studies suggest that when educated about policies, the American public is willing to embrace strategies that serve to lessen crime through rehabilitation and reintegration of offenders (Applegate et al., 1996; Doble Research Associates Inc., 1994, 1995; Nagin, Piquero, Scott, and Steinberg, 2006). Education campaigns are frequently used to warn the public about enhanced penalties for certain crimes or about the dangers of drinking and driving, drug, and even tobacco use. Why not employ such strategies to educate the public about the truths of effective crime reduction? Just as worker buy-in is important for a system strategy to be effective, public buy-in is important for achieving policy and funding support from local and state governments. After all, if public fear or crime served as the impetus for much of the get tough movement, perhaps public education and awareness of truly effective crime prevention and control strategies could provide the needed catalyst for a new and more successful societal response to crime.

Maconochie's Gentleman: May I Shake Your Hand?

What would it take then for early release policies to be excepted and effective? What I suggest here is that it would take more than even the "best" early release policy. It would take a system effort to decide that "correcting" and not just "punishing" is a true mission of the system. It would take a realization that, as best put by Jeremy Travis (2002), "they all come back." Therefore, a viable purpose for the system is preparing offenders from the very beginning of their experience with the system for that ultimate day when they are returned to the community. Such a strategy would require buy-in from all hands of the system—from prosecutor and judge to correctional and parole officers to reentry workers. Moreover, it would take educating and preparing the community for their return as well. Early release would then not be so much "early" but anticipated and planned. It would be a system more like that originally proposed and practiced by Maconochie in which felons once disconnected from society earned a true second chance and gained a meaningful role.

Thus, I will end with another popular saying. It goes like this: "Those who do not learn from history are doomed to repeat it." I propose here that sometimes what we learn from history, particularly that experience on Norfolk Island, is that history is occasionally worthy of repeating.

References

Applegate, Brandon K., Francis T. Cullen, Michael G. Turner, and Jody L. Sundt. 1996. Assessing public support for three-strikes-and you're-out laws: Global versus specific attitudes. *Crime & Delinquency*, 42: 517–534.

BI Incorporated. 2002. *Overview of the Illinois DOC High-Risk Parole Re-Entry Program and three-Year Recidivism Outcomes of Program Participants*. Boulder, CO: Author.

Champion Dean John. 2005. *Probation, Parole and Community Corrections*. Fifth Edition. Upper Saddle River, NJ: Pearson Prentice Hall.

Cook, David. 2000. Transition for success. In (Reginald Wilkinson, ed.), *Correctional Best Practices: Directors' Perspective*. Middletown, CT: Association of State Correctional Administrators.

Doble Research Associates Inc. 1994. *Crime and Corrections: The Views of the People of Vermont*. A Report to the Department of Corrections, The State of Vermont, July 1994.

Doble Research Associates Inc. 1995. *Public Opinion About Crime and Corrections in the State of North Carolina*. A Preliminary Report, January 1995.

Fulp, Elmer. 2001. Project Rio: Reintegration of offenders. In *State of Corrections: Proceedings, Annual Conference* 2000. Lanham, MD: American Correctional Association.

Kleiman, Mark A. R. 1999. *Getting Deterrence Right: Applying Tipping Models and Behavioral Economics to the Problems of Crime Control*. Perspectives on Crime and Justice 1998–1999 Lecture Series 3. Washington, D.C.: National Institute of Justice.

Kramer, John and Jeffery T. Ulmer. 2002. Downward departures for serious violent offenders: Local court "corrections" to Pennsylvania's sentencing guidelines. *Criminology*, 40: 601–636.

Montana Department of Corrections. *Definitions*. Retrieved May 12, 2011 from www.cor.mt.gov/Facts/prerelease.mcpx.

Morris, Norval. 2002. *Maconochie's Gentlemen: The Story of Norfolk Island the Roots of Modern Prison Reform*. New York: Oxford University Press.

Nagin, Daniel S., Alex R. Piquero, Elizabeth S. Scott, and Laurence Steinberg. 2006. Public preferences for rehabilitation versus incarceration of juvenile offenders: Evidence from a contingent valuation survey. *Criminology & Public Policy*, 5: 301–326.

Petersilia, Joan. 2003. *When Prisoners Come Home: Parole and Prisoner Reentry*. New York: Oxford University Press.

Seiter, Richard and Karen Kadela. 2003. Prisoner reentry: What works, what doesn't and what's promising. *Crime & Delinquency*, 49: 360–388.

Travis, Jeremy. 2005. *But They All Come Back: Facing the Challenges of Prisoner Reentry*. Washington, D.C.: The Urban Institute Press.

Travis, Jeremy and Joan Petersilia. 2006. Reentry reconsidered: A new look at an old question. In (Edward J. Latessa and Alexander M. Holsinger, eds.), *Correctional Context Contemporary and Classical Readings*, 3rd Edition. Los Angeles, CA: Roxbury.

Walker, Samuel. 2006. *Sense and Nonsense about Crime and Drugs: A Policy Guide*, 6th Edition. Belmont, CA: Thomson Higher Education.

Wright, Kevin A. and Jeffrey W. Rosky. 2011. Too early is too soon: Lessons from the Montana Department of Corrections Early Release Program. *Criminology & Public Policy.* This issue.

Zimring, Franklin. 1998. *Crime Is Not the Problem: Lethal Violence in America*. New York: Oxford University Press.

Zimring, Franklin and Gordon Hawkins. 1999. Public attitudes toward crime: Is American violence a crime problem?" In (Edward L. Rubin, ed.), *Minimizing Harm: A New Crime Policy for Modern America*. Boulder, CO: Westview.

Megan Kurlycheck is an Associate Professor in the School of Criminal Justice at the University at Albany, SUNY. Dr. Kurlychek received her Ph.D. in Crime, Law and Justice from Penn State University in 2004 and in addition to her academic career has worked in several policy arenas including the Pennsylvania State Senate, the Pennsylvania Commission on Sentencing and the National Center for Juvenile Justice. Her primary research interests involve the connections between offenders and the criminal justice system including ways in which the system helps and/or hinders rehabilitation. Her work has been published in *Criminology*, *Crime and Public Policy*, *Crime and Delinquency*, *The Journal of Research in Crime and Delinquency*, *Justice Research Policy*, *Justice Quarterly and the Journal of Criminal Justice*.

More than just early release

Considerations in prison reduction policies

Susan Turner
University of California, Irvine

Current Crisis in Corrections

Correctional expenditures are attractive targets for state belt tightening in today's uncertain fiscal times. Efforts to curtail spending are being contemplated by some states and implemented by others. One set of options focuses on achieving operational efficiencies, for example, by closing prisons, reducing staff, and curtailing services and programming. Other strategies are focused on the "back end" of the system—reduction in sentence lengths through earned credits or good time, as well as changes to reduce revocations for probationers and parolees. Strategies may even be implemented at the "front end" of the system, diverting offenders to county- rather than to state-level institutions, or changing felonies to misdemeanors in an attempt to reduce prison admissions. States are grappling with which options to use but often without enough information to make informed decisions in terms of expected impacts, costs, and benefits. The scale of the problem and the desire for solutions are reflected in the recent solicitation by the National Institute of Justice for a national evaluation study on prison closings and alternative strategies employed by state correctional systems for dealing with massive state budget shortfalls (National Institute of Justice, 2011: 4).

The article by Wright and Rosky (2011, this issue) provides us with a study to help fill the information void. These authors examine an early release program for releasees from prison and community correctional facilities placed on conditional release or on traditional parole. The findings indicate that offenders released early, "particularly for those released directly from prison, may have exacerbated the financial strain problem in Montana as a result of increased recidivism." As Wright and Rosky note in their conclusions, all eyes will

Direct correspondence to Susan Turner, Department of Criminology, Law and Society, 3336 Social Ecology II, University of California, Irvine, Irvine, California 92697 (e-mail: sfturner@uci.edu).

DOI:10.1111/j.1745-9133.2011.00769.x

be turned to California and other states in similar positions to determine how they fare in these desperate financial times.

Californians are living in interesting times, as the expression goes. With the recent U.S. Supreme Court decision requiring the state to reduce the prison population by more than 30,000 (*Brown, Governor of California et al. vs Plata et al.,* 2011), earned release, or "enhanced credit earning" as it is called in California, is one of several options in the state's correctional policy toolkit. This policy essay highlights several key issues faced, not only in earned release but more broadly as states grapple with overburdened state prisons. I consider a focus on "risk" and structured decision making as well as on its relationship to "stakes," the use of "earned" benefits for inmates and moving the problem "down the road"—which is similar to the concept of "criminal justice thermodynamics" discussed by Wright and Rosky (2011).

Risky "Risk"

Risk assessment most often is conducted to determine who poses better or worse risks for recidivism. Actuarial tools have been used since the early decades of the 20th century and have shown to be superior to pure "clinical" predictions for the past 60 years (Ægisdóttir et al., 2006; Meehl, 1954), although as outlined in the subsequent discussion, practitioners often are resistant to the use of actuarial tools (Harcourt, 2007). In the past 20 years, the Risk, Needs, and Responsivity (RNR) approach (Andrews et al., 1990; Andrews, Bonta, and Wormith, 2006) garnered support for the delivery of correctional resources. In this model, offenders who are at the highest risk of recidivism and have the greatest criminogenic needs are targeted for treatment and services, using methods that are tailored to the learning style and ability of the offenders. This approach has a natural appeal in today's environment, where resources are stretched and agencies need to determine how best to allocate scarce resources. The concept is simple; however, the execution can be challenging.

The California Department of Corrections and Rehabilitation (CDCR) embraced the concept of RNR as integral to their California Logic Model, which was developed in 2007 with the assistance of a national panel of experts (California Department of Corrections and Rehabilitation, 2007). This model prescribed the assessment, treatment, and service provision for offenders during their time incarcerated and while on parole. One major component of the model was the use of a risk tool. In late 2007, the CDCR, in collaboration with the University of California, developed the California Static Risk Assessment (CSRA), which was modeled after a tool developed in Washington State that primarily used information on prior convictions and supervision violations recorded in official records—static factors (Turner, Hess, and Jannetta, 2009). For the first time, the CDCR began to contemplate wide-scale policy decisions and practices that were based on actuarial "risk" of reoffending rather than ones that were based solely on offense (either current or prior). Four lessons from the adoption of a risk-based approach are instructive, however, and highlight the "risk" of using risk prediction in reducing prison populations.

By their very nature—predicting human behavior—risk instruments are not perfect. Some offenders may be predicted to recidivate and do not; others are predicted to remain crime free and recidivate. Most tools (including the CSRA) fall into a range of accuracy that borders on moderately predictive. For example, Yang, Wong, and Coid's (2010) meta-analysis of risk assessment tools for violence found tools ranged between area under the curve of 0.65 to 0.71. In other words, the score of a randomly chosen "recidivist" is higher than that of a randomly chosen "nonrecidivist" approximately 70% of the time—better than chance but not 100% accurate. And, as Skeem and Monahan (2011) pointed out, we may be reaching a point of diminishing returns in instrument development (p. 41). With validated tools, we can estimate the numbers and kinds of errors that risk tools make. However, nonperfect tools serve as fuel for politicians, law enforcement, and others who fear that public safety is compromised unduly by anything but a near-perfect tool. However, it is unrealistic to expect this level of accuracy in currently available tools.

Closely related to accuracy is a perception that "low risk" is synonymous with "no risk." California is known for having the highest recidivism rates in the country. Approximately two thirds return to prison within 3 years after release, primarily for violations of parole (California Department of Corrections and Rehabilitation, 2010). The CSRA divides offenders into five risk categories, including low, moderate, and high risk for drugs; high risk for property offenses; and high risk for violent offenses. In the "low-risk" group, approximately 48% are arrested for a felony offense within 3 years of release. This finding stands in contrast to greater than 80% rearrest rates in the high-risk groups (Turner et al., 2009). Although low-risk offenders are clearly less likely to recidivate than high-risk offenders (by definition), prison reduction policies that target "low risk" cannot be a silver bullet for reducing recidivism to zero. This finding was evident also in the more than 30% return-to-prison rates for low-risk offenders in the Wright and Rosky's (2011) article.

Incorporating risk-assessment tools into correctional practices often requires replacing (or at least integrating) a traditional "clinical" approach to offender supervision with a tool that may not be welcomed initially. In California, the CSRA was incorporated into a parole violation decision-making instrument (PVDMI), in which an offender's risk score (low, moderate, or high) and severity of violation (categorized into four levels of seriousness) was used to define a recommended response for parole agents. Thus, low-risk parolees and low-seriousness violations were to receive lower level responses, with revocation to prison reserved for the higher risk, more serious violations. The implementation of the tool resulted in substantial override and underride responses by the parole agents, often because of concerns about availability of treatment resources. However, a recurring theme by agents was their distrust of a risk tool that replaced their field experience for making case decisions (Murphy and Turner, 2009). Difficulties of agent "buy-in" of a parole decision-making tool also was reported by the developers of Ohio's parole decision tool (Martin and Van Dine, 2008).

Perhaps the most bedeviling aspect of incorporating risk into correctional decision making is its relationship with "stakes." In fact, one might argue that the two often are used interchangeably when they mean different things. For example, in the article by Wright and Rosky (2011), they indicate that the early release program selected offenders deemed to be "low risk"—with the "main qualification being that their crime was not of a violent or sexual nature." These authors note also that these offenders could be "considered to pose a lesser threat toward the members of society as they are mainly composed of nonviolent, drug, and property offenders." But are nonviolent drug and property offenders low risk? Or, are they more correctly identified as "low stakes?"

Analyses conducted by the CDCR on a 2005–2006 cohort of released prisoners found that offenders identified as being serious or violent recidivate at a lower rate than nonserious or violent offenders. Within the first year of release, approximately 50% of the nonserious or violent offenders return to prison versus 44% of serious or violent offenders (CDCR, 2010). In addition, sex offenders had slightly lower officially recorded recidivism rates than non–sex offenders. These findings are not unique to California—Wright and Rosky (2011) too found that a prior theft or nonviolent conviction increased the likelihood of recidivism.

In reality, policies that aim to reduce prison populations may pit "stakes" and "risks" against each others. The RNR principle suggests policy decisions that focus resources on the higher risk; yet the public may demand high levels of supervision for certain "types" of offenders (such as sex offenders), regardless of the risk the individual parolee may pose to the community. The "stakes" to the agency for potential crimes committed by certain offender groups may outweigh concerns about effective targeting of resources and money for optimum crime reduction.

California's nonrevocable parole program (NRP) is an example of a policy that, although ostensibly targeting lower risk parolees for placement on a form of summary parole, incorporated "stakes" into the screening criteria for eligibility. Nonrevocable parole was one of several policies designed to reduce California's prison population as part of SBx3–18, which went into effect in California in early 2010. By removing lower risk parolees from routine parole supervision, the Division of Adult Parole Operations hoped to refocus resources on higher risk parolees, providing them with needed services and programs, which is consistent with an RNR approach. Parolees on NRP are unsupervised by parole and subject to law enforcement search and seizure but cannot be returned to prison for parole violations. High-risk offenders, identified by the CSRA, are not eligible for the program. In addition to risk, however, offenders also are not eligible if they have a current or prior violent or serious conviction, are sex offenders, or are members of a prison gang—in other words, high-"stakes" offenders. Although findings on the impact of NRP on recidivism comparing similar parolees who were on routine parole are not yet available, two aspects of the implementation of NRP are noteworthy. Despite the incorporation of both "risk" and "stakes" into the eligibility criteria, NRP was vociferously criticized particularly by law enforcement (Blankenstein, 2010), and fears about increased crime and pressures on local law enforcement were reported by the media (Hernandez, 2010; Johnson, 2010). Second, original estimates of between 20,000 and 25,000 offenders being eligible did not materialize. Instead, about half the original estimate (or approximately 12% of the entire parole population) are currently on NRP (CDCR, 2011), possibly reflecting "stakes" constraints on a risk-based approach.

Earned Early Release

Early release is perhaps the most frequently discussed policy for reducing prison populations quickly and is the policy option evaluated in the article by Wright and Rosky (2011). They discuss briefly the possibility that increased recidivism in their study might be the result of the reduced deterrent effect of a reduction in time served. Despite prior research suggesting this might be the case, Wright and Rosky offer another explanation for their findings—that offenders were "thrust back into society with little time to prepare for reentry." This explanation suggests that "early release" may be a more complicated concept than first meets the policy makers' or the public's eye.

Could early release, if done differently, actually assist offenders in their reentry? In California and elsewhere, "earned early" release may actually hold promise in preparing offenders for return to the community. In California, eligible offenders can earn up to 6 weeks each year off their sentences if they achieve certain educational and programming milestones. The earned credits are viewed as incentivizing inmates to participate in rehabilitative programs that, in turn, should reduce recidivism after release from prison. In California, the earned credit approach is consistent with a recommendation by the California Expert Panel (as discussed previously) and was developed to be in line with evidence-based practices.

However, reasoned discussions about early release often are overshadowed by fierce opposition based on beliefs. As experience in Wright's own state of Montana, California, and elsewhere has shown, early release often is maligned by the media and politicians as irresponsible practice. Malcolm Young (2010) discussed recently the sobering experiences in Illinois with shuttering early release programs as a result of the inaccurate portrayal of the policy and political pressure, which have lead to increases in the state's prison population. As Young (2001) stated, "politicians in Wisconsin and Illinois are hardly alone in their adherence to the notion that public safety depends on the convicted criminal serving every day of a prison sentence." The irony of the situation is that, as Young pointed out, sentence lengths in laws are not based on research showing what is necessary to punish an offender or protect the public (and states are not consistent in their punishments for the same offense) and the sentence a judge imposes is not often based on a "fine line" determination of what will protect the public (Young, 2011). It may seem disingenuous to attack early release without also questioning existing sentencing structures. However, add to the opposition rhetoric the difficulties in rolling out programs, in which calculation of credits can be confusing, as in California (Stanton, 2010), early release programs face enormous challenges.

Moving the Problem on Down the Road

In their discussion of study findings using the concept of thermodynamics, Wright and Rosky (2011) suggest that the early release of inmates can unintentionally shift the burden of overcrowding from an institutional setting to a community supervision setting. They note that it is unclear with their findings to what extent increased recidivism for early releases was a result of offender behavior or attributed to adjustments made by parole officers in their treatment of violations, given increased workloads.

Thermodynamics can take an even more dramatic and direct form in the shifting of responsibility for certain correctional populations from the state to the "locals." California is engaged in this bold policy experiment, which is referred to as "realignment." Realignment is not a new idea in the state; prior policy shifts from the state to counties have occurred in recent history (realigning mental and health and other social services in the 1990s) (Misczynski, 2011). However, given the current fiscal climate in the state, coupled with the recent U.S. Supreme Court decision requiring a decrease of more than 30,000 inmates in state correctional facilities, realignment in corrections is now front and center in the state. California's governor signed into law AB109 requiring, among other changes, to have low-level offenders serve sentences in local jails, rather than in prison, allowing offenders to be closer to needed education and drug treatment services. Also included is the devolving of parole supervision from the state, to the counties, which also would reduce prison populations by preventing revocations to prison. Opponents call for more prison building and shipping more inmates out of state (California currently has 10,000 inmates housed in private prisons out of state) (Dolan, 2011). Local support exists for the move by county sheriffs, who are responsible for the county jails—however, they cannot house state offenders without sufficient funds to do so. Such a major realignment is guaranteed to place new pressures on local counties, in terms of both housing inmates and assisting them with services needed for their successful reentry.

Conclusions

Whatever states do to reduce overcrowding, they must take into account findings from Wright and Rosky (2011) about offenders needing to be prepared for reentry. As part of the solution, as these authors suggest, states may need to think carefully about the responses to minor violations of parole, reserving revocation to prison as a last option. Wright and Rosky also suggest additional prerelease planning and intense services upon initial release as ways to ensure offenders are better prepared for reentry. However, as I have discussed, savvy policy makers must be prepared for the pitfalls that accompany policies to reduce prison population—how to deal with offender risks versus stakes, deterrence versus rehabilitation, and state versus local incentives and investments in the problem. We all live in interesting times.

References

Ægisdóttir, Stefania, Michael J. White, Paul M. Spengler, Alan S. Maugherman, Linda A. Anderson, Robert S. Cook, et al. 2006. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, 34: 341–382.

Andrews, Don A., Ivan Zinger, Robert D. Hoge, James Bonta, Paul Gendreau, and Francis T. Cullen. 1990. Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28: 369–404.

Andrews, Don A., James Bonta, and J. Stephen Wormith. 2006. The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52: 7–27.

Blankenstein, Andrew. 2010. LAPD complains about parole officials classifying suspect in police shooting as "low-level, nonviolent." *Los Angeles Times*. July 14.

California Department of Corrections and Rehabilitation. 2007. *Expert Panel on Adult Offender and Recidivism Reduction Programming: Report to the California State Legislature*. Sacramento: Author.

California Department of Corrections and Rehabilitation. 2010. 2010 *Adult Institutions Outcome Evaluation Report*. Sacramento: Author.

Dolan, Jack. 2011. California risks violating first deadline to cut prison population. *Los Angeles Times*. June 8.

Harcourt, Bernard E. 2007. *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age*. Chicago, IL: University of Chicago Press.

Hernandez, Salvador. 2010. Parolees may get break. *Orange County Register*. February 2.

Johnson, Scott. 2010. Recent high-profile assaults have raised concern about non-revocable parole status. *The Mercury News*. October 18.

Martin, B. and S. Van Dine. 2008. *Examining the Impact of Ohio's Progressive Sanctions Grid: Final Report*. Columbus: Ohio Department of Rehabilitation and Correction.

Meehl, P. 1954. *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence*. Minneapolis: University of Minnesota.

Misczynski, D. 2011. *Rethinking the State-Local Relationship: An Overview*. San Francisco: Public Policy Institute of California.

Murphy, Amy and Susan Turner. 2009. *Parole Violation Decision-Making Instrument (PVDMI) Process Evaluation*. Irvine: University of California–Irvine, Center for Evidence-Based Corrections.

National Institute of Justice. 2011. *Solicitation: Research and Evaluation in Justice Systems*. Washington, D.C.: U.S. Department of Justice.

Skeem, Jennifer. L., & Monahan, John. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, 20(1), 38–42.

Stanton, Sam. 2010. Early release of Sacramento County inmates to resume. *The Sacramento Bee*. February 16.

Turner, Susan, James Hess, and Jesse Jannetta. 2009. *Development of the California Static Risk Assessment (CSRA), Center for Evidence-Based Corrections, University of California, Irvine*.

Wright, Kevin A. and Jeffrey W. Rosky. 2011. Too early is too soon: Lessons from the Montana Department of Corrections Early Release Program. *Criminology & Public Policy*. This issue.

Yang, Min, Steve Wong, and Jeremy Coid. 2010. The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, 136: 740–767.

Young, Malcolm C. 2011. Turning back the clock on early release. *The Crime Report*. June 8. Retrieved from http://www.thecrimereport.org/archive/2011-06-turning-back-the-clock-on-early-release.

Court Case Cited

Brown, Governor of California et al. vs Plata et al., No. 09–1233, 563 U.S.C. (2011).

Susan Turner is a Professor in the Department of Criminology, Law and Society at the University of California's Irvine campus. She also serves as Director of the Center for Evidence-Based Corrections, and is a board member of the California Rehabilitation Oversight Board (C-ROB). She received her Ph.D. in Social Psychology from the University of North Carolina at Chapel Hill. Dr. Turner is a member of the American Society of Criminology, the American Probation and Parole Association, and is a Fellow of the Academy of Experimental Criminology.

The cattle call of reentry

Not all processes are equal

Faye S. Taxman
George Mason University

With budget crunches capturing the attention of state and local governments, the affordability of long prison (jail) sentences is being questioned. States have taken daring steps to use early release tactics, with the expectations that such moves will both save money and reduce recidivism. Kevin A. Wright and Jeffrey W. Rosky (2011, this issue) explored the impact of early release efforts in one state. Not surprisingly, the results are disappointing in that those individuals who were released early were more likely to recidivate than those who served their time. Wright and Rosky point to several explanations, including the potential actions of parole officers and other attributes covered under the umbrella of "criminal justice thermodynamics" where the mechanics of the criminal justice system continue working in such a fashion to "backfire." The findings of this study are predictable—early releases are more likely to recidivate—and those thrust back into society without preparation are doomed to fail. In this essay, I consider the importance of the messages that are attached to different policy initiatives, the messages that basically support the cattle call that "all things *should* work." Unless we focus on the messages and the "punitive culture," most of our efforts will fail to reform the justice system or people involved in justice environments.

Today's Scenario

The state legislature declares that one strategy to reduce the budget deficit is to release imprisoned drug addicts early and send the offenders to a residential treatment program. The released prisoners are freed up to 18 months early on parole (or supervised release) and placed in a treatment program. After the residential treatment, the person is mandated to continue treatment in the community as part of supervised release with drug testing and

Direct correspondence to Faye S. Taxman, Department of Criminology, Law and Society, George Mason University, 4400 University Drive, Fairfax, VA 22030 (e-mail: ftaxman@gmu.edu).

DOI:10.1111/j.1745-9133.2011.00780.x

monitoring. It looks like a win–win with less use of incarceration and better opportunities to reduce recidivism and improve public safety.

What messages are associated with this early release policy?

> *Politician*: We are saving money. Treatment is *supposed* to be more useful for addicts so we might as well try this option.
>
> *Local community*: We are saving money, and addicts should have not been incarcerated since treatment is the better option. We will change lives by offering treatment.
>
> *Local law enforcement*: Offenders are getting off easy, and we need to make sure they are not using drugs and creating public safety problems.
>
> *Local probation office*: Offenders are getting off easy, and we have to do more work to monitor them; we need to make sure the public is protected with a no tolerance approach (one positive drug test leads to some jail time).
>
> *Local treatment providers*: Treatment is now recognized as a valuable commodity. But the politicians have not given us significantly more resources to handle people who have legal problems and are drug addicts. The funds are not sufficient, and they do not include pharmacological therapies (e.g., naltrexone, suboxtene, and methadone) that are known to be effective. We are being asked to do too much with not enough resources.
>
> *Offenders*: We are getting out of prison early because the public does not have the funds to keep us behind bars—they are making us go to treatment, but they do not really believe that we *need* treatment. It is just a money thing. And we are given "nail 'em and tail 'em" supervision, so we will shortly be back in prison.

The same policy can have different messages for various audiences, ranging from money saving to more effective practices to the "right" type of punishment. And each interpretation sets up different expectations depending on which message one believes (or hears). *It is clear that we cannot "win" reentry with such mixed messages and varied expectations.*

Why does early release have null or little impact on reducing recidivism, as shown by Wright and Rosky (2011)? Or for that matter, why does it seem that few correctional policies or programs alter the recidivism trajectory for offenders? During the last 20 years, we experimented with a variety of means to punish offenders, and in most instances, we expect these experiments to "work" (reduce offending)—this applies to early release, increased sentence lengths, decreased sentence lengths, boot camps, intensive supervision, alternatives to incarceration, and so on. Yet seldom have our efforts generated the desired results, and we express great disappointment that another program or idea did not lead to the "promise land." It may seem rhetorical to ask the question about why we expect programs or policies to "work" (reduce recidivism), but raising this question might compel us to look more closely at how and why we expect a policy or program to do the hard work of impacting individual-level offending behaviors. In other words, if we reduce a prison sentence and release someone early, why do we think this will result in reduced offending?

How can we get results? As a first step, we might consider the notion that the punishment system (or the legal system in general) is an authoritarian process where the only "input" that the accused has is through the defense attorney, if there is a defense attorney (a resource that has been diminishing over time). This imperfect scenario means that the person is seldom involved directly in choosing a punishment, treatment program, and setting (incarceration/community) that is appropriate for them or that addresses the factors that contribute to criminal behavior. It reinforces the punitive culture, a culture that makes it difficult for change to occur. In this era where the risk–needs–responsivity (RNR) model is recognized as a tool to address this gap, it is also necessary to recognize that the processes we use to implement RNR in operational settings may be a determining factor in whether a sentence or correctional program has an impact. That is, if we use the same old culture, then we will get the same old results. We need to consider creating environments where individuals involved in the justice system can change and where individuals who work in the justice system can use the evidence-based tools of RNR and correctional programming. Otherwise, we are creating inconsistent messages to the public, to the justice system, and to the people in the justice system—these messages will neutralize the impact of evidence-based techniques creating the thermodynamics that maintains the status quo.

In the following text, I highlight the rationale for involving the person (offender) in the decision-making process to clarify the expectations at the individual level, and the message as to the purpose and intent of the policy or program. Using the concepts of communication and messaging, the failure to achieve results may have less to do with the program's content than with the degree to which the individual offender understands the purpose of the programming, the link to his or her own behavior, and how he or she can benefit from participation (ownership). Second, the manner in which we involve people in the process is likely to have an impact on the lack of motivation to change or the ambivalence attached to modifying one's own behavior. Feeley (1992) illustrated how the "process is the punishment," and if we expect RNR and evidence-based programming to be effective, then we need to override the correctional culture with one that supports individual change. Third, the criminal justice policies and programs need to be based on theoretical interventions that are designed to help the person to understand the steps involved in the personal change process. These factors are directed at increasing the legitimacy of the programs and services offered, while providing the important and critical linkages to helping individuals link their own actions and behavior to the policies and programs to which they are exposed.

Offender's Involvement in the Decision

The rationale for involving offenders in the sentencing, program assignment, or release decisions has the potential to impact outcomes. Three theories—legitimacy and procedural justice, contingency management, and shared decision making—can all contribute to the different outcomes by outlining expectations and providing the person with more ownership to their own behavior and/or outcomes. Each offers a mechanism to address offender behavior that evolves from defiance or cynicism regarding the legitimacy of justice actors such as police, probation/parole officers, judges, or others. Policies and practices emanate from the mass incarceration (criminalization) movement, where violating offenders for parole and probation rules is commonplace. The inclusion of offenders in the decision-making process is premised on addressing the culture and norms of the justice system. Although the current system passively involves offenders (e.g., plea bargaining and choosing whether to participate in a program), the associated process only serves to fuel more cynicism about the system. Conversely, active involvement in the decision-making system is designed to address how best to help the offender assume ownership for the outcomes.

Procedural Justice

In a series of work, Tom Tyler (2000, 2003) tested empirically the importance of procedural justice or processes that promote the fair and equitable application of the law, particularly in the area of law enforcement. A repeated theme is that police can increase their legitimacy through the use of fair procedures, and that fair procedures seem to be essential to achieve impacts on individual-level behaviors, even if the outcomes are arrest or other negative events. That is, procedural processes shape people's perception of whether they were treated in a manner that they can reconcile as being fair, just, or appropriate. This perception is important because it defines the experience and influences how people interpret events. For example, in one experiment on domestic violence, the police used standard protocol to explain the arrest policy for any domestic situations involving violence. The use of the standard police protocol (language), compared with the traditional practice of merely conducting the arrest, had a deterrent effect on future domestic violence behaviors. The standard language served to inform the individual of the purpose of the police action as well as to clarify expectations as to why the arrest is occurring (see Paternoster, Brame, Bachman, and Sherman, 1997). The veracity of the evidence surrounding procedural justice, and the importance of reinforcing the legitimacy of justice actions, has caused it to be an important component of police research over the last decade (Skogan and Frydl, 2004), and there is now a call for more literature to understand how to advance police legitimacy.

Tyler (2010) extended his argument about procedural justice to the field of corrections where he outlines the core components to consist of "voice, neutrality, treatment with respect and dignity and trust in authorities" (p. 129). Building on the premise that the process and contextual environment influence individual-level behavior, Tyler demonstrated how the procedural justice framework is applicable within correctional settings (primarily prisons). This framework is built on the need to construct processes where individuals can participate in, and be a part of, key decisions in a way that promotes a perception of justice and fairness. Such processes serve to foster compliance with the rules and law. Implicit is that the corrections process also needs to promote legitimacy where the actions and decisions of justice institutions are sound, defensible, and clear. Actions like "early release" premised on loss of budget would not necessarily be considered legitimate because the reason for the release is not tied to the offender's crime, punishment, or conduct in prison, which are the usual rationales underscoring both the original sentence and any modifications to the condition of release. Given that the early release policies studied by Wright and Rosky

(2011) did not address the purpose of the original offense and did not prepare the person for release, it is not surprising that illegitimate actions by social institutions do not translate into good individual-level behavior.

Procedurally *just* corrections processes would support the legitimacy of the punishment by recognizing offender behavior that complies with the rules of the correctional agencies and advances the purpose of the sentence. If the sentence was premised on punishment or rehabilitation, then actions taken by the correctional agencies to reinforce the overall goals would support the legitimacy of the sentence. But actions like early release for the convenience of the state (e.g., to save money) merely undermine legitimacy by suggesting that the original punishment scheme was inappropriate, that the sentence length was not warranted, or that the use of incarceration was an unnecessary punishment tool. All send messages that undermine the legitimacy of the original scheme. But a few changes in how the decisions are made could actually support the appropriate actions by the institutions. That is, a procedurally just corrections or judicial system would allow these persons to have input into the decisions affecting them, thus participating in such a manner that they believe that they are an equal partner. This belief, in turn, leads to increased support for the system based on a perception that institutional processes are fair and equitable. Under this premise, the correctional system should rely less on its authoritarian nature, where the state alone is responsible for making decisions and justice or policy actors can "flex their muscles" in their decisions.

Instead, Tyler's four components promote an environment where the emphasis is more on the processes that are built on a creating a just environment (Blader and Tyler, 2003). Tyler raised the issues of "neutrality" as a consistent "application of the rules" instead of the preference for "individual-level decisions" that often seem to be biased or influenced by participating actors (Tyler and Lind, 2002). Two other tenets, respect and dignity, are important to uphold the humane treatment of individuals as a way of signifying their citizenship. Building the system to respect the individual and reinstate citizenship status is a critical in our punishment system given that many policies suggest that the individual is not a vital part of society.

Contingency Management (CM)

Do positive reinforcers or negative reinforcers promote more socially compliant behavior? Our punishment system is built on the deterrence principle that finds compliance to be more of a product of avoiding unpleasant circumstances. This is similar to the utilitarian concept that people will avoid "costs" of punishment through a calculus that the costs are not worth the "benefits" (the fruits of the offending behavior). Therefore, people are more likely to comply with the law or rules to avoid unpleasant punishment. Severe punishments are premised on increasing the stakes associated with the punishment. Whereas punishment dominates criminology, the psychological literature approaches compliance from a slightly different approach, noting that responses are generally a result of operant conditioning.

Within this framework, nearly two decades of research has shown that the human spirit is more inclined to positive reinforces as a motivating factor for improved compliant behavior. That is, people are more likely to comply if they understand the expectations and they are incentivized in this direction.

Contingency management is a procedure that focuses on rewarding people for desired behaviors (e.g., staying drug free, maintaining employment, and not being homeless) that is recognized as an evidence-based treatment (National Institutes on Drug Abuse [NIDA], 2000). In many ways, it is similar to the old token economy systems in prisons that rewarded offenders with early release if they complied with prison rules and worked hard toward correcting their ways. CM interventions have been developed primarily for use in substance abuse treatment, where reinforcement-based interventions have been shown to improve short-term outcomes such as drug-free days (Petry, Alessi, Ledgerwood, and Sierra, 2010; Stitzer, Petry, and Peirce, 2010). Three systematic reviews confirm that CM improves a variety of client-level outcomes, including drug use, treatment attendance, and treatment retention (Griffith, Rowan-Szal, Roark, and Simpson, 2000; Lussier, Heil, Mongeon, Badger, and Higgins, 2006; Prendergast, Podus, Finney, Greenwell, and Roll, 2006). CM involves using operant-based behavioral reinforcement strategies to enhance positive behaviors through the use of either material or social reinforces.

CM has a formula that emphasizes an individual's involvement in decisions, similar to the procedural justice processes outlined by Tyler (above). CM begins with desired target behaviors such as applying for jobs, providing drug-free urines, attending treatment, and going to self-help group meetings. These target behaviors are considered incremental steps to addressing a problem behavior. In essence, the process that the counselor (or probation officer) uses to work with the person is designed to help the person understand the nature of the problematic behavior, help them identify target behaviors that are "within reach" (doable and feasible), and outline the contingencies associated with achieving the target behavior. Along with this process are the contingencies associated with repeated negative behaviors. Clarifying expectations is part of the negotiation process where the individual is empowered to address his or her own choices. In a desirable CM scenario, the defined target behaviors would be agreeable to the offender and the benefits would be clearly laid out. CM uses rewards schedules that may involve "bonus points" (extra rewards) for special efforts or duration of periods where the person maintains the target behaviors. The goal is to put in place a structured pathway of success where individuals know up front what they are likely to gain from participating in the rewarding protocol. The CM framework enhances the core components of procedural justice: outline expectations, outline benefits, and follow through on expected (positive) outcomes.

In a recent implementation study where CM was applied, probation officers were receptive to the concept of the CM protocol but expressed concerns about some processes (see Friedmann et al., 2008; Rudes et al., 2011). Notably, these concerns centered on informing the individual (offender) of the likely outcomes that might occur with achieving the desired target behaviors. Officers expressed hesitancy because of their concerns about offenders manipulating the system. The officers also raised concerns about sharing information about offenders' risk and need factors because they felt that people should already be aware of their problem behavior. Unexpected barriers in implementing CM centered more around the process of working with the offenders—open communication, clear expectations, and disclosure of criminal justice information—than about the use of tokens or incentives.

Shared Decision Making

In the medical field, most patient education models center around individuals understanding the nature of their disorder, and through this understanding, they are more likely to comply to achieve a better health status. The power between the provider and the individual lies in the balance between the two in making choices concerning the nature of the intervention. These choices are bounded by cost, safety, impact on others, and alignment with values. This type of balance also applies to justice settings. As defined by Légaré et al. (2008: 3):

> The health decision-making process is complex, as it brings together a health professional, considered a scientific content expert, and an individual, considered an expert in his own personal values. It is in this context that there is considerable interest today in the process of shared decision-making (SDM). SDM is defined as a decision-making process jointly shared by patients and their health care provider, and is said to be the crux of patient-centered care. It relies on the best evidence about risks and benefits associated with all available options (including doing nothing) and on the values and preferences of patients, without excluding those of health professionals. Therefore, it includes the following components: establishing a context in which patients' views about treatment options are valued and deemed necessary; reviewing the patient's preferences for role in decision-making; transferring technical information; making sure patients understand this information; helping patients base their preference on the best evidence; eliciting patients' preferences; sharing treatment recommendations; and making explicit the component of uncertainty in the clinical decision-making process.

The shared decision-making process recognizes the individual as a contributor to the process, where individuals determine their own options, within a range. This model is recognized as a key factor to facilitating the role of the individual to be "in charge" of making decisions designed to maximize valued outcomes. In the context of the justice system and its processes, shared decision making is viewed as an opportunity to help an individual understand conforming behavior and the consequences of nonconforming behavior. The justice system uses early release as a reward but does so in a manner where the criteria for release are unclear and vary considerably, therefore not allowing the individual to participate in the decision. Like the procedural justice process where clear consequences are needed, shared decision making is based on a premise of the system actors giving up "power" by allowing the person to make choices for which ultimately he or she is responsible. The theory is that by allowing the person to respond to his or her own risk and need factors, there will be greater ownership to the behavior and greater commitment to behavior change. That is, the sentencing and corrections process could be more appropriately aligned where options are not merely judged based on what is the "least restrictive sanction" or the most severe punishment but from a perspective of being in the best interest of the person. This is similar to the tenets of therapeutic jurisprudence, but the emphasis is more on involving the person in the decision-making process.

Away from the Cattle Call: The Demand for New Processes

Wright and Rosky's (2011) study did the field a huge favor by reminding us that the best intended policies that have one motivation (reducing costs) may not serve to achieve other objectives (recidivism reduction). And, that the criminal justice system works in such a fashion that the laws of physics apply—each part will maintain equilibrium that ultimately results in few changed outcomes. Wright and Rosky identified that there is a need for greater preparation for release from prison, that parole officers may operate in such a fashion to reflect the punitive culture, and that the collective can serve to explain the poor recidivism outcomes for early release offenders. But, the underlying issue is that there is not just a *need* for reentry services, but also there is a *need* for a different process that will facilitate better individual outcomes. That is, the omnipotent punitive culture is the main message and it has the capacity to override all good intentions. Recent evaluation findings confirm that our existing approaches offered within said punitive environments have null to little impact. For example, the Serious and Violent Offender Reentry Initiative, with its emphasis on type of prerelease and early parole supervision process, yielded null recidivism reduction results (Lattimore, Steffey, and Visher, 2009) The same is true for Project Greenlight (it actually had a negative effect), a comprehensive program that was defined as a "kitchen sink" of various services (see Wilson and Davis, 2006; Marlowe, 2006) and Transitional Case Management, a strengths-based case management model with a standard fare of programming (Prendergast et al., 2011). In essence, each approach is based on similar premises that the state or county agency personnel will assess, determine needs, and assign individuals to appropriate programs (the components of RNR). That is, the social control framework, the punitive environment, maintains "power" and authority with state (or county) actors, with little role for the individual offender (client). Most reentry processes are built on case management models where the preference is for the state to determine the needs of the individual and for the individual offender to be *directed (conditioned)* to partake in various services or programs. This approach fails to include our knowledge about operant conditioning, stages of change, change processes, or human developmental growth.

Given the results of nearly 30 years of mass-incarceration–based policies and programming, which has contributed to more of the same accountability approaches to reentry and offender programming, two needs exist: (1) an adoption of a supportive, more offender-change friendly environment; and (2) different theoretical models to guide the next generation of reentry processes and programming that are based on the person having a role in making choices that support changes. The punishment-oriented, social-control–based, and deterrent-based policies that drive contemporary correctional programming account for much of poor outcomes—it is unlikely that a person will or can change in an environment where staff and the system are focused on "looking for failures." Operant conditioning, stages of change, and human development argue for a model that involves the person in the decisions and choices to be made. Often referred to as "client centered," the approach focuses on involving the person in the process in a manner in which the individual is empowered to make choices. As highlighted previously, the reentry process must incorporate dignity and respect for the individual as part of the process—these mechanisms of action are theoretically led to achieve ownership and commitment to changing human behavior. The guiding principle is that effective change-oriented processes/programs cannot tolerate a "process is the punishment" approach (see Feeley, 1992) because this orientation undermines the change processes.

The role of the individual in the process is one strategy to alter the existing framework to reentry. As noted by Tyler in his work on procedural justice, involving the offender in decisions, even if the choices are bounded and associated with clear consequences, should serve only to boost the legitimacy of the processes and increase compliance. Capitalizing on the theoretical framework of procedural justice, contingency management, and shared decision making, these processes need to be associated with clear messages regarding the importance of the individual in making individual decisions, in participating in activities that are of benefit to the individual, and in clarifying how this process can contribute to recidivism reduction outcomes. The message to the individual offender, as well as to the stakeholder community, needs to be clear as to the rationale for the reentry process or programming. Clear social messages about the rationale are important in the process.

Determining the Win

Recidivism reduction is a long haul. Traditionally, a criminological approach focuses mainly on examining technical violations, rearrest, reconviction, and reincarceration (measures of recidivism). Theoretically sound reentry approaches need to recognize that intermediary steps contribute to long-term goals such as participation in treatment and services, employment, stability in the community, and other measures. We need to use a health services approach that considers outcomes to be a function of processes, such as initiation, engagement, and retention in the core processes/programming. The health services model recognizes that each of these processes contributes to outcomes, and that if one desires to assess early impact, then more attention needs to be given to the individual's various levels of participation of appropriate services. The health services framework emphasizes the individual and how the system can facilitate the individual's healthy involvement in programs and services. Measuring these processes is also important to strengthening our knowledge about effective reentry practices. Criminologists in our studies need to adopt these measures to assess whether the reentry process is affecting individual engagement—without the individual making such choices to participate and to take advantage of the programs and services, it is unlikely that we can make good strides. Reentry success then should be determined on these intermediate, proximal steps that are more associated with making gains in the community.

Addressing the Challenges Ahead

Although the evidence-based practices literature has aggressively pushed forward the need to expand the array of clinical services such as cognitive-behavioral therapy and therapeutic community, the challenges of implementation in justice settings cannot be understated. Within the context of implementation, there is a need to understand better the transportability of medical or social service evidence-based practices in the justice system (see Taxman and Belenko, 2011). Procedural justice, contingency management, and shared decision- making emanate from other settings. Although the core components are similar (i.e., clarify expectations and messages, identify target behaviors, and outline consequences), more attention needs to be given to the issues related to operating within justice settings, particularly regarding the compatibility with core legal principles. That is, involving an offender in key decisions in the justice system must be handled in a manner that does not jeopardize civil liberties such as the presumption of innocence and right against self-incrimination. In the past, these legal principles have been used as barriers to including the offender in the process. Although involving individuals in decisions has clear benefits, it is recognized that there is a need to establish an environment where embarrassing, stigmatizing, and even incriminating experiences are minimized. A review of guidelines for confidentiality in clinical practice can be informative such as the National Association of Social Workers' Code of Ethics (18 provisions related to the confidentiality of the therapist/client relationship) or the American Psychological Association's Code of Ethics (10 provisions). That is, although coerced treatment assumes that individuals will participate because they are required to, the reality is that there is a need to create an environment in which the individual feels empowered to collaborate in the change process, and such involvement facilitates behavioral change but does not jeopardize civil rights.

Toward a New Century of Reentry Programming

Reentry processes fail for several reasons—inadequate programming, inadequate resources, punitive approaches, and mixed messages to all sorts of stakeholders. The confusing and conflicting messages of current programming merely serve to delegitimize the reentry processes (and correctional programming), and to contribute to the cynicism that society does not desire for offenders to succeed. Doing more of the same will produce more of the same. In this essay, I examined the social messages to understand that there is likely to be a lot of misconceptions regarding reentry strategies and hot these misconceptions feed unmet expectations. I have also offered a few new frameworks to consider in the reentry processes and programming to actualize the individual's commitment to desired societal goals of reduced offending. As scientists, we need to demand more of our profession to develop and test new processes that can affect the correctional culture, develop programs that are likely to alter offending behavior, conduct studies of organizational change to understand implementation issues better, and highlight the need to alter environments for programs and people to be more successful. In all, to achieve different outcomes, it is apparent that reentry processes should focus on a different position for the individual offender in reentry. A repeat of second-class citizenship, limited options in terms of choices, and programming that fails to address criminogenic needs (such as early release efforts) will not alter the prospects for the future. The next generation of programming should pay greater attention to humanistic approaches that override the current "catch 'em" reentry efforts, even from well-intended policies such as early discharge. And, this includes attention to the organizational culture of reentry programs, probation and parole agencies, social service agencies, and the justice system overall—otherwise, the criminal justice thermodynamics will persist.

References

Blader, Steven L. and Tom R. Tyler. 2003. A four-component model of procedural justice: Defining the meaning of a "fair" process. *Personality and Social Psychology Bulletin*, 29: 747–758.

Feeley, Malcolm M. 1992. *The Process Is the Punishment: Handling Cases in a Lower Criminal Court*. New York: Russell Sage Foundation.

Friedmann, Peter D., Elizabeth C. Katz, Anne G. Rhodes, Faye S. Taxman, Daniel J. O'Connell, Linda K. Frisman, *et al*. 2008. Collaborative behavioral management for drug-involved parolees: Rationale and design of the Step'n Out Study. *Journal of Offender Rehabilitation*, 47: 290–318.

Griffith, James D., Grace A. Rowan-Szal, Ryan R. Roark, and D. Dwayne Simpson. 2000. Contingency management in outpatient methadone treatment: a meta-analysis. *Drug and Alcohol Dependence*, 58: 55–66.

Lattimore, Pamela K., Danielle M. Steffey, and Christy A. Visher. 2009. *Prisoner Reentry Experiences of Adult Males: Characteristics, Service Receipt, and Outcomes of Participants in the SVORI Multisite Evaluation*. Research Triangle Park, NC: RTI International. Retrieved September 7, 2011 from ncjrs.gov/pdffiles1/nij/grants/230419.pdf.

Légaré, France, Glyn Elwyn, Martin Fishbein, Pierre Frémont, Dominick Frosch, Marie-Pierre Gagnon,et al. 2008. Translating shared decision-making into health care clinical practices: Proof of concepts. *Implementation Science*, 3: 2.

Lussier, Jennifer Plebani, Sarah H. Heil, Joan A. Mongeon, Gary J. Badger, and Stephen T. Higgins. 2006. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction*, 101: 192–203.

Marlowe, Douglas B. 2006. When "what works" never did: Dodging the "scarlet M" in correctional rehabilitation. *Criminology & Public Policy*, 5: 339–346.

National Institutes on Drug Abuse. 2000. *Principles of Drug Addiction Treatment: A Research-Based Guide*. NIH Publication No. 00–4180. Rockville, MD: Author.

Paternoster, Raymond, Robert Brame, Ronet Bachman, and Lawrence W. Sherman. 1997. Do fair procedures matter? The effect of procedural justice on spouse assault. *Law & Society Review*, 31: 163–204.

Petry, Nancy M., Sheila M. Alessi, David M. Ledgerwood, and Sean Sierra. 2010. Psychometric properties of the contingency management competence scale. *Drug and Alcohol Dependence*, 109: 167–174.

Prendergast, Michael, Linda Frisman, Joann Y. Sachs, Michele Staton-Tindall, Lisa Greenwell, Hsiu-Ju Lin, et al. 2011. A multi-site, randomized study of strengths-based case management with substance-abusing parolees. *Journal of Experimental Criminology*, 7: 225–253.

Prendergast, Michael, Deborah Podus, John Finney, L. Greenwell, and John Roll. 2006. Contingency management for treatment of substance use disorders: A meta-analysis. *Addiction*, 101: 1546–1560.

Rudes, Danielle, Faye S. Taxman, Shannon Portillo, Amy Murphy, Anne Rhodes, Maxine Stitzer, Peter Luongo, Peter D. Friedmann, M.D. (in press, 2011), Adding positive reinforcements in justice settings: Acceptability & feasibility. *Journal of Substance Abuse Treatment.*

Skogan, Wesley G. and Kathleen Frydl. 2004. *Fairness and Effectiveness in Policing: The Evidence*. Washington, DC: National Academies Press.

Stitzer, Maxine L., Nancy M. Petry, and Jessica Peirce. 2010. Motivational incentives research in the National Drug Abuse Treatment Clinical Trials Network. *Journal of Substance Abuse Treatment*, 38: S61–S69.

Taxman, Faye S. and Steven Belenko. 2011. *Implementing Evidence-Based Practices in Community Corrections and Addiction Treatment*. New York: Springer-Verlag.

Tyler, Tom R. 2000. Multiculturalism and the willingness of citizens to defer to law and to legal authorities. *Law & Social Inquiry*, 25: 983–1019.

Tyler, Tom R. 2003. Procedural justice, legitimacy, and the effective rule of law. *Crime and Justice: A Review of Research*, 30: 283.

Tyler, Tom R. 2010. Legitimacy in corrections. *Criminology & Public Policy*, 9: 127–134.

Tyler, Tom R. and E. Allan Lind. 2002. Procedural justice. In (Joseph Sanders and V. Lee Hamilton, Eds.), *Handbook of Justice Research in Law*. Boston, MA: Kluwer Academic.

Wilson, James A. and Robert C. Davis. 2006. Good intentions meet hard realities: An evaluation of the Project Greenlight reentry program. *Criminology & Public Policy*, 5: 303.

Wright, Kevin A. and Jeffrey W. Rosky. 2011. Too early is too soon: Lessons from the Montana Department of Corrections Early Release Program. *Criminology & Public Policy*. This issue.

Faye S. Taxman is a University Professor in the Criminology, Law and Society Department and Director of the Advancing Correctional Excellence Center at George Mason University. Dr. Taxman is health services criminologist with an expertise in implementation and translational science. She has published over 120 articles and book chapters, and is authored of a translational work on using evidence-based practices in supervision, of *Tools of the Trade: A Guide to Incorporating Science into Practice* (http://nicic.gov/Library/020095) as well as a recent book on *Implementing Evidence Based Practices in Community Corrections and Addiction Treatment* (with Steve Belenko).

Transitional jobs program
Putting employment-based reentry programs into context

Robert Apel
Rutgers University

Employment is a major point of intervention in an offender's criminal career, and employment-based reentry programs have obvious appeal as a policy lever intended to slow the "revolving door" of prison. Indeed, both President George W. Bush and President Barack Obama advocated for federal funding of prisoner reentry initiatives that include employment training provisions—a clear illustration that resolving the employment challenges faced by ex-prisoners is a decidedly bipartisan issue. The astounding scale of contemporary prisoner reentry—several hundred thousand individuals leave prisons annually, not to mention more than one million additional individuals who leave jails—means that a very large number of individuals will invariably return to the community and experience difficulty finding and maintaining stable employment. This issue has important implications from the standpoint of public safety because recidivism studies routinely find that ex-prisoners who maintain stable employment are significantly less likely to be rearrested.

Employment is therefore strongly linked with criminal desistance—both theoretically and empirically. Disappointingly, evaluations of employment-based reentry programs suggest that they tend to yield minimal impacts on the employment and recidivism prospects of targeted individuals (see Bushway and Reuter, 2004). Yet the authors of one of the most recent meta-analyses of employment programs for ex-prisoners lamented the absence of programs inspired by contemporary thinking regarding the best practices in correctional intervention (Visher, Winterfield, and Coggeshall, 2005). The unambiguous conclusion

Direct correspondence to Robert Apel, School of Criminal Justice, Rutgers University, 123 Washington Street, Newark, NJ 07102 (e-mail: robert.apel@rutgers.edu).

DOI:10.1111/j.1745-9133.2011.00781.x

is that a great need exists for the development and evaluation of modern, innovative employment programs.

A comprehensive employment-based reentry program must have a dual focus on the *employment* and *employability* of ex-prisoners. Although employment per se—the "give 'em a job" approach to reentry that characterizes subsidized work programs—is an essential part of any viable reentry strategy, it must be supplemented by efforts to strengthen the work orientation and work readiness of ex-prisoners. Programs that focus solely on job provision and search assistance, in the absence of a skills component, have a poor track record of success in improving the employment and recidivism prospects of ex-prisoners (see Bushway and Reuter, 2004). The emphasis on employability is especially important in light of findings from the National Supported Work Demonstration that, among participants who were recently incarcerated, 33 percent were fired from their program job, and another 20 percent were terminated for other negative reasons such as reinstitutionalization (MDRC, 1980).

Transitional work programs are a very promising avenue for employment-based reentry programming. These programs combine job provision and search assistance with skills training and a variety of other support services (for a broad overview of transitional work programs, see Bloom, 2010). Transitional work programs thus have the necessary dual focus on employment and employability, providing for temporary, subsidized work augmented by mentoring in the "soft skills" that can help ex-prisoners anticipate and meet the demands of the workplace.

The subject of the article by Janine Zweig, Jennifer Yahner, and Cindy Redcross (2011, this issue)—the New York City-based Center for Employment Opportunities (CEO)—is an example of an innovative approach to the design of a transitional work program (see also Redcross, Bloom, Azurdia, Zweig, and Pindus, 2009). The CEO program is comprehensive indeed. After referral by a parole officer, the program begins with a 4-day job readiness class. This is followed by assignment to a subsidized, minimum wage job in the public sector (e.g., building maintenance and cleaning at city and state agencies) with a supervised work crew. During the subsidized work period, 4 days each week are spent at the worksite, and on the fifth day of each work week, participants meet with office-based CEO staff for job coaching and other support services, including counseling that focuses on child support and family relationships. When he or she is deemed "job ready," the participant is assigned to a job developer who matches the ex-prisoner with a permanent position. Finally, for the first year of unsubsidized employment, the parolee is eligible to receive ongoing support and employment incentives.

What is perplexing about the CEO evaluation is that, although the program significantly reduced the probability of recidivism during the first 2 years (for program participants compared with controls who were given only search assistance, but they also were eligible to seek non-CEO employment assistance), these reductions apparently did not develop because participants were more successful at acquiring an unsubsidized job. No significant program effects on employment outcomes were found after the subsidized work period, for which we have several plausible explanations. One distinct possibility, considered by Zweig et al. (2011) in the current study, is that the program participants were decidedly heterogeneous and therefore not equally likely to benefit from what the program had to offer. Specifically, their analysis reveals that the highest risk ex-prisoners—generally, those who were younger and had more extensive arrest histories—were the sole beneficiaries of the CEO program effects with respect to recidivism reduction.

To situate the findings within the broader context of correctional interventions, Latessa (2011, this issue) and Gaes and Bales (2011, this issue) in their policy essays build on insights rooted in the "risks, needs, responsivity" (RNR) model, focusing especially on the risk principle (Andrews and Bonta, 2010; Andrews, Bonta, and Hoge, 1990). As implied by the terminology, the risk principle concerns the identification of the characteristics that put certain individuals at higher risk of criminal recidivism relative to others, and the targeting of such individuals for the most intensive supervision and intervention. As observed by these essayists, the finding that the CEO program effects were limited to the highest risk offenders is actually anticipated by the RNR model. Moreover, both essayists caution that, although the differences were not statistically significant, the CEO program potentially had a perverse impact on recidivism among the lowest risk ex-prisoners, a result also anticipated by the RNR model.

The fact that hundreds of thousands of people leave the nation's prisons each year lends urgency to a renewal of innovative programming designed to ease their transition back into the community. Employment-based reentry programs such as CEO are central to such efforts, and steps taken to understand for whom—and why—these programs are effective will yield dividends for future reentry policy. I, for one, remain optimistic that such efforts will bear fruit by improving the lives of ex-prisoners, reducing their burden on the criminal justice system, and protecting society at large.

References

Andrews, Don A. and James A. Bonta. 2010. *The Psychology of Criminal Conduct*, 5th edition. Cincinnati, OH: Anderson.

Andrews, Don A., James A. Bonta, and Robert D. Hoge. 1990. Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17: 19–52.

Bloom, Dan. 2010. Transitional Jobs: Background, Program Models, and Evaluation Evidence. Unpublished manuscript. New York: MDRC. Retrieved September 21, 2011 from mdrc.org/publications/553/full.pdf.

Bushway, Shawn and Peter Reuter. 2004. Labor markets and crime. In (James Q. Wilson and Joan Petersilia, eds.), *Crime: Public Policies for Crime Control*. Oakland, CA: Institute for Contemporary Studies.

Gaes, Gerry G. and William D. Bales. 2011. Deconstructing the risk principle: Addressing some remaining questions. *Criminology & Public Policy*. This issue.

Latessa, Edward. 2011. Why the risk and needs principles are relevant to correctional programs (even to employment programs). *Criminology & Public Policy*. This issue.

Manpower Demonstration Research Corporation [MDRC]. 1980. *Summary and Findings of the National Supported Work Demonstration*. Cambridge, MA: Ballinger.

Redcross, Cindy, Dan Bloom, Gilda Azurdia, Janine Zweig, and Nancy Pindus. 2009. *Transitional Jobs for Ex-Prisoners: Implementation, Two-Year Impacts, and Costs of the Center for Employment Opportunities (CEO) Prisoner Reentry Program*. New York: MDRC. Retrieved September 21, 2011 from mdrc.org/publications/529/full.pdf.

Visher, Christy A., Laura Winterfield, and Mark B. Coggeshall. 2005. Ex-offender employment programs and recidivism: A meta-analysis. *Journal of Experimental Criminology*, 1: 295–315.

Zweig, Janine, Jennifer Yahner, and Cindy Redcross. 2011. For whom does a transitional jobs program work? Examining the recidivism effects of the Center for Employment Opportunities program on former prisoners at high, medium, and low risk of reoffending. *Criminology & Public Policy*. This issue.

Robert Apel received his Ph.D. in criminology and criminal justice from the University of Maryland in 2004, and is currently Associate Professor in the School of Criminal Justice at Rutgers University. His research specialties include employment and criminal behavior, incarceration and employment, and violent victimization and injury.

TRANSITIONAL JOBS PROGRAM

Overview of: "For whom does a transitional jobs program work?

Examining the recidivism effects of the Center for Employment Opportunities program on former prisoners at high, medium, and low risk of reoffending"

Janine Zweig
Jennifer Yahner
Urban Institute

Cindy Redcross
MDRC

Research Summary

This study documents that a transitional jobs program for former prisoners had its strongest reductions in recidivism among those in the program with the highest risk of reoffending. The New York City-based Center for Employment Opportunities (CEO) is a transitional jobs program designed to help former prisoners increase longer term employment and, consequently, reduce recidivism. Interim results from MDRC's rigorous impact evaluation of CEO showed reduced recidivism in both the first and the second year of follow-up. The current study

- *expanded on the interim results by using regression-based analysis to identify whether CEO had its greatest impact among low-, medium-, or high-risk offenders—with risk levels defined by participants' characteristics before random assignment that are associated with recidivism after random assignment.*
- *found that CEO had its strongest reductions in recidivism for former prisoners who were at the highest risk of recidivism. For high-risk former prisoners, participation in CEO reduced significantly the probability of rearrest, the number of rearrests, and the probability of reconviction 2 years after random assignment to the program.*

DOI:10.1111/j.1745-9133.2011.00766.x

Policy Implications

The findings suggest important implications for policy, practice, and future evaluation research.

1. *The limited resources available to transitional jobs programs for former prisoners should be targeted toward people at the highest risk of recidivating because they are helped most by this intervention.*
2. *Age and criminal history are critical determinants of recidivism risk, and programs for former prisoners should consider assessing the likelihood of reoffending using tools that measure both characteristics. Although specific to our sample, the average-aged offender (33 years old) was considered at a high risk of recidivism if he had nine or more prior arrests; similarly, those with seven prior arrests (the sample average), were at high risk of recidivism if 28 years old or younger.*
3. *Because CEO interim evaluation results did not show an effect on increases in unsubsidized employment, it is not clear what is causing the recidivism effects. Thus, future evaluation research should examine the mechanisms by which this transitional jobs program reduces recidivism among its clients, and particularly, the highest risk clients.*

Keywords

transitional; *employment*; *prisoners*; *recidivism*; *risk*

For whom does a transitional jobs program work?

Examining the recidivism effects of the Center for Employment Opportunities program on former prisoners at high, medium, and low risk of reoffending

Janine Zweig
Jennifer Yahner
Urban Institute

Cindy Redcross
MDRC

Each year, more than 700,000 individuals are released from prisons nationwide (Sabol, West, and Cooper, 2009). Many former prisoners have lengthy criminal backgrounds and struggle to avoid recidivating while reintegrating into the communities to which they return. A person's criminal history, age, and gender all contribute to the likelihood of future crime (Gendreau, Little, and Goggin 2006; Levinson, 2002). Among released prisoners, younger males with extensive criminal histories are often at greatest risk of future recidivism (Langan and Levin, 2002).[1]

This research paper was prepared by the Urban Institute and MDRC as part of the Enhanced Services for the Hard-to-Employ project funded by the Department of Health and Human Services and the Department of Labor through Contract Number HHS-233–01–0012. The authors thank David Butler, Dan Bloom, and Charles Michalopoulos of MDRC; reviewers from the Office of Planning, Research and Evaluation in the Administration for Children and Families; and reviewers from the Office of the Assistant Secretary for Planning and Evaluation. The opinions expressed in this document are those of the authors and do not necessarily represent the official position or policies of the Department of Health and Human Services, or of the Urban Institute, its trustees, or its funders. Portions of this article were reported to the Department of Health and Human Services in the form of a research brief as per contractual obligations. Direct correspondence to Janine Zweig, Urban Institute, 2100 M Street, NW, Washington, DC 20037 (e-mail: jzweig@urban.org).

1. As was the case in this study, released prisoners across the United States tend to be a relatively older (age 30 years and older) group with a higher number of prior arrests (more than seven) (Langan and Levin, 2002). Thus, when one talks about "younger" individuals with "extensive" criminal histories being

DOI:10.1111/j.1745-9133.2011.00767.x

According to social control theories, employment helps prevent criminal activity by providing individuals with legitimate ties to conventional society (Piehl, 2003; Sampson and Laub, 1993). Relatedly, social capital theorists argue that the interpersonal relationships individuals form through employment—the positive social networks—can aid desistance from criminal behavior (Baron, Field, and Schuller, 2001; Boeck, Fleming, and Kemshall, 2008; Evans, 2002; Farrall, 2004). Indeed, research has shown that stable employment is an important predictor of postprison reentry success (Visher and Travis, 2003; Visher, Winterfield, and Coggeshall, 2005). Using a multistate longitudinal design, Visher and colleagues found that former prisoners who worked more weeks and had higher earnings the first few months after release were less likely to be reincarcerated 1 and 3 years after release (Visher and Courtney, 2007; Visher, Debus, and Yahner, 2008; Yahner and Visher, 2008).

It is also widely believed that program intervention soon after prison release can be critical to long-term reentry success (see, e.g., Baer et al., 2006; Johnston-Listwan, Cullen, and Latessa, 2006; Solomon et al., 2008). Accordingly, programs across the country have focused on finding jobs for former prisoners after their release. However, the results from previous evaluation efforts have shown that such employment programs have limited ability to reduce recidivism. Meta-analyses of employment programs for former prisoners have found little if any effect on postprison criminal activity (Aos, Miller, and Drake, 2006; Visher et al., 2005). Yet, many studies included in these meta-analyses had limited methodological designs and those with the most rigorous designs were conducted decades ago. For example, Uggen's (1999) oft-cited reanalysis of the National Supported Work Demonstration relied on data collected in the 1970s.[2] Consequently, Visher et al. (2005) called for stronger evaluations of more current employment programs, specifically pointing to programs that provide transitional jobs as those in most need of rigorous evaluation.

Although some past research has shown that subsidized work programs for youth do not promote noncriminal behavior (e.g., Bushway and Reuter, 2002; Piliavin and Masters, 1981), transitional jobs models have emerged as a promising approach to intervention

at greatest risk of recidivism, it is important to understand that the meaning of these descriptors varies from their use among a general population. What is equally important to understand, however, is that prior research on individuals of many different ages and criminal backgrounds has found that the relationship among these three factors—age, prior arrests, and recidivism—seems constant (younger individuals with a higher number of prior criminal events are more likely to recidivate than the converse) (Levinson, 2002).

2. Notably, Uggen (1999) found that randomly assigned employment reduced self-reported recidivism among older (age 26 years or older) but not younger former prisoners. Although the average age of the National Supported Work sample was 25, which is approximately 8 years younger than former prisoners in the current sample, these findings are consistent with those reported later in this article to the extent that Uggen's measure of age was a proxy for prior arrests (age and prior arrests are typically highly correlated, with older offenders having had greater opportunity to accumulate more prior arrests).

for reducing recidivism among the growing population of adult former prisoners. Since 2008, President Obama has specifically cited transitional jobs as a priority in efforts to reduce poverty during his administration (Obama and Biden, 2008). Earlier that same year, President Bush signed the Second Chance Act, which will lead to substantial funding for programs related to prisoner reentry in the coming years—including funding for transitional jobs programs—and will provide opportunities for policy makers and others to identify best practices for reducing recidivism and helping former prisoners transition after release.

Given this national attention yet dearth of rigorous research on transitional jobs programs, the evaluation of the New York City-based Center for Employment Opportunities (CEO) is particularly relevant and important. CEO is a transitional jobs program designed to help former prisoners obtain earnings and work experience soon after release and to obtain permanent unsubsidized employment, in order to improve longer term recidivism outcomes. CEO has been part of a long-term, random assignment study funded by the Administration for Children and Families in the U.S. Department of Health and Human Services and the Office of the Assistant Secretary for Planning and Evaluation, with additional funding from the U.S. Department of Labor. The evaluation, as led by MDRC and its partner, the Urban Institute, assesses the impact of CEO on employment and recidivism for program participants compared with control group participants. Two-year interim results have been published and are described in the preliminary findings section of this article; 3-year results are to be released in 2011.

The CEO Program Evaluation

Program Description

The goal of the CEO program model is to provide former prisoners with (a) immediate work and pay through a day-labor approach, (b) necessary work experience for finding more permanent jobs, and (c) a way to build work-related soft skills (Redcross, Bloom, Azurdia, Zweig, and Pindus, 2009). The focus is not on training clients in a particular field, but instead it provides participants the chance to create a recent work history and to develop work behaviors that can help them find and maintain unsubsidized positions.

The CEO program model includes a 4-day, preemployment life skills class to prepare participants for the transitional job, job searches, and employment after the transitional job. Participants begin their transitional jobs after they finish the class and are assigned to daily work crews for 4 days a week, each with its own CEO supervisor. In general, the work includes maintenance of buildings and groundskeeping for city and state agencies at several dozen sites around New York City, including sweeping, mopping, dusting, cleaning bathrooms, breaking down boxes for recycling, and outdoor maintenance (Nightingale and Trutko, 2008).

Along with daily work, the CEO program helps build soft skills to facilitate long-term employability in two ways. First, it provides on-the-job coaching at the worksite by on-site supervisors. Second, it provides support and guidance through office-based job coaches with whom participants formally meet once a week on the day that they are not assigned to a work crew. Once a participant is considered job ready, he or she begins meeting with a job developer once a week, while maintaining the position in the transitional job work crew. In an effort to find clients permanent employment, job developers link employers who have open positions with clients who have the skill sets that fit the positions. After being placed into a permanent unsubsidized position, a participant becomes eligible for a CEO's job retention incentive program, which offers noncash incentives for continued employment—such incentives as gift certificates from various stores and Metrocards for use in the city's mass transit system. CEO tracks participants' continued employment via pay stubs, which they can bring to the CEO offices for a reward every 30 days through 1 year, at which point they are eligible for a $250 gift certificate for retaining their position.

Another major goal of the CEO program was its provision of support to learn soft skills and of social support to participants. Job coaches checked on individuals' progress weekly and helped refer those in need to a variety of supportive services, from transportation assistance to substance abuse treatment. Partway through the evaluation, CEO implemented a *Passport to Success* program to enhance the supervision and support provided by staff. The passport system required participants to carry their own evaluation booklets daily, which were completed on transitional workdays by the worksite supervisors, directly in front of participants. Therefore, on a daily basis, supervisors were connecting with workers about how they were doing on site. Job coaches reviewed the completed booklets with clients during their weekly meetings to address any issues or concerns and to give participants positive feedback about their performance as appropriate. In addition to support from job coaches and worksite supervisors, daily contact with other transitional jobs participants provided peer-level social support. According to the chief executive officer of CEO, regular contact among CEO participants might function as a positive peer support group over time (Mindy Tarlowe, personal communication, November 10, 2010).

Evaluation Design

In 2004 and 2005, 977 former prisoners on parole who reported to CEO each week were assigned, at random, to a program group ($n = 568$) or control group ($n = 409$).[3] Intake sessions occurred every Friday during the study period, and the groups were sent immediately to separate floors of the building once assigned to the program or control groups.

3. This research sample was a portion of the approximately 2,000 parolees whom CEO serves each year and representative of the regular parolee population.

Random assignment was only conducted during those weeks where there was a surplus of clients—meaning that more parolees came to CEO than slots available on transitional job work crews. At this time, participants also completed initial baseline information forms that included a limited set of questions about demographic information, as well as employment and educational histories. Life skills classes for members of both research groups started immediately on their being assigned to either the program or the control condition.

The control group received limited services compared with those of the CEO program model, including a 1.5-day preemployment life skills class that focused on securing identification documents needed for employment, job search strategies, and interview concepts. Approximately 37% of control group members completed the class. Participants were then given access to a resource room monitored by a staff person from whom they could seek help and where they could use such equipment as computers, fax machines, and phones with voicemail accounts to engage in job search activities. Few control group members visited the resource room more than two or three times.

The program group received CEO's full-service package, as described previously. Almost 80% of the program group completed their 4-day preemployment life skills class and almost everyone who completed the class—72% of the full program group—worked at least 1 day in a work crew in the transitional jobs program. Of the participants who worked in a work crew, 29% worked 1 to 4 weeks; 40% worked 5 to 12 weeks; and 24% worked 13 weeks or more. The average length of time in a work crew was 8 weeks. Sixty percent of the total program group met with a job coach at least once, and 22% of the group met with a job coach more than four times. Similarly, 57% met with a job developer at least once and approximately 20% met with a job developer more than four times.

However, if clients never work in a transitional job, then we would not expect them to meet with a job coach or a job developer. In fact, whether clients worked a transitional job and how long they worked that job mattered in terms of whether they met with job coaches and job developers and how many times they met with them. For example, among those who worked a transitional job for 13 weeks or more, 94% met with a job coach at least once and 72% met with a job coach more than four times; 99% met with a job developer at least once and 71% met with a job developer more than four times.

Only a portion of those in the program group found permanent positions. Independent of the number of times participants met with job developers, 30% of the program group were placed in permanent jobs or found permanent jobs through their own efforts. However, when one considers the amount of time participants worked in a transitional job, the placement rates were higher. For example, 56% of those who worked in a transitional job 13 weeks or more found permanent positions.

Sample Characteristics

The parolee participants in this study were referred by a parole officer, but not mandated by parole orders, to participate in the CEO program.[4] Referred parolees who reported to CEO were then randomly assigned to either the program or control groups, as described previously. Table 1 shows the characteristics of these sample members as reported by participants on baseline information forms, whereas Table 2 describes the criminal histories of the sample members at the time of random assignment using administrative records from the New York State Division of Criminal Justice Services and the New York City Department of Correction. The sample is mostly male (93%), 64% are Black, and 31% are Hispanic. The average age of the group at baseline was 33 years. The sample had an average of 7.5 arrests (including 4.5 felony arrests), 6.7 convictions (including 2.6 felony convictions), and 60 months spent in state prisons. Few statistically significant differences were found between the program and control groups at baseline, none of which seem to have affected the findings.

Preliminary Findings

Interim results from MDRC's impact evaluation of CEO showed that the program had a short-term effect on employment, mostly driven by the subsidized transitional jobs that participants received (Bloom, Redcross, Zweig, and Azurdia, 2007; Redcross et al., 2009). However, the program did not affect whether clients obtained unsubsidized employment. Figure 1 demonstrates these findings. It presents quarterly employment rates over the follow-up period for jobs covered by unemployment insurance. In quarter 1, for example, 66% of the program group worked for at least 1 day compared with only 26% of the control group—for an impact of 40 percentage points, mostly represented by the subsidized jobs that the CEO provided. Differences between the groups diminished over time as a result of a decline in employment among the program group. These findings are in line with research showing that subsidized work programs tend to have large increases in employment for their clients during their time in transitional jobs but that these programs tend not to have long-term effects on employment (Bushway and Reuter, 2002).

Despite the limited employment effects, CEO had long-lasting effects on the rate of recidivism for program participants. As shown in Table 3, the CEO group had lower rates of recidivism in both year 1 and year 2 of follow-up, compared with the control group. More specifically, the program group members were less likely to be convicted

4. Some criminal justice populations that CEO serves are mandated to participate in the program (for example, the New York State Shock Incarceration participants). Because they were mandated to receive CEO's full service package, they were not eligible to be included in this study sample.

TABLE 1

Selected Characteristics of Sample Members at the Time of Random Assignment, by Research Group

Characteristic	Program Group	Control Group	Full Sample	*p* Value
Age (%)				0.842
18 to 24 years	19.0	20.3	19.6	
25 to 30 years	23.8	23.7	23.8	
31 to 40 years	31.4	30.3	30.9	
41 years or older	25.7	25.7	25.7	
Average age (years)	33.7	33.7	33.7	0.854
U.S. citizen (%)	74.6	73.6	74.2	0.113
Race/ethnicity (%)				0.936
White, non-Hispanic	1.4	2.2	1.8	
Black, non-Hispanic	64.3	64.5	64.4	
Hispanic	31.2	29.8	3.6	
Other	3.0	3.4	3.2	
Male (%)	91.4	95.3	93.0	0.020*
Has any children less than age 18 (%)	48.1	47.9	48.0	0.441
Lives with any children less than age 18 (%)	16.3	15.2	15.8	0.872
Ordered to provide child support to a child less than age 18 (%)	18.9	19.9	19.3	0.917
Education (%)				0.798
High-school diploma	9.5	11.4	1.3	
General Educational Development (GED) certificate	42.6	43.9	43.1	
Beyond high school	4.8	3.5	4.3	
None of the above	43.1	41.2	42.3	
Housing status (%)				0.889
Rents or owns home	16.6	19.9	18.0	
Lives with friends or relatives	59.1	55.1	57.4	
Lives in transitional housing	12.4	11.2	11.9	
Lives in emergency housing or is homeless	3.7	5.4	4.4	
Other	8.1	8.4	8.3	
Marital status (%)				0.892
Married, living with spouse	8.1	9.3	8.6	
Married, living away from spouse	7.4	7.7	7.5	
Unmarried, living with partner	21.8	20.1	21.1	
Single	62.6	63.0	62.8	
Ever employed (%)	81.1	81.2	81.2	0.784
Employed 6 consecutive months for one employer (%)	59.9	62.7	61.1	0.200
UI-covered employment in the quarter prior to random assignment[a] (%)	14.9	11.7	13.6	0.370
UI-covered employment in the year prior to random assignment[a] (%)	24.1	24.0	24.0	0.873
Sample size	568	409	977	

Source. MDRC calculations using data from the baseline information form and unemployment insurance wage records from New York State.
Notes. The data in this table are unweighted, but the results for the statistical significance test are weighted by the week of random assignment. To assess differences in characteristics across research groups, chi-square tests were used for categorical variables, and *t* tests were used for continuous variables. The significance level indicates the probability that one would incorrectly conclude that a difference exists between research groups for the corresponding variable.
[a] This measure was created using data from unemployment insurance wage records from New York State.
$*p < 0.05$.

TABLE 2

Criminal History at the Time of Random Assignment, by Research Group

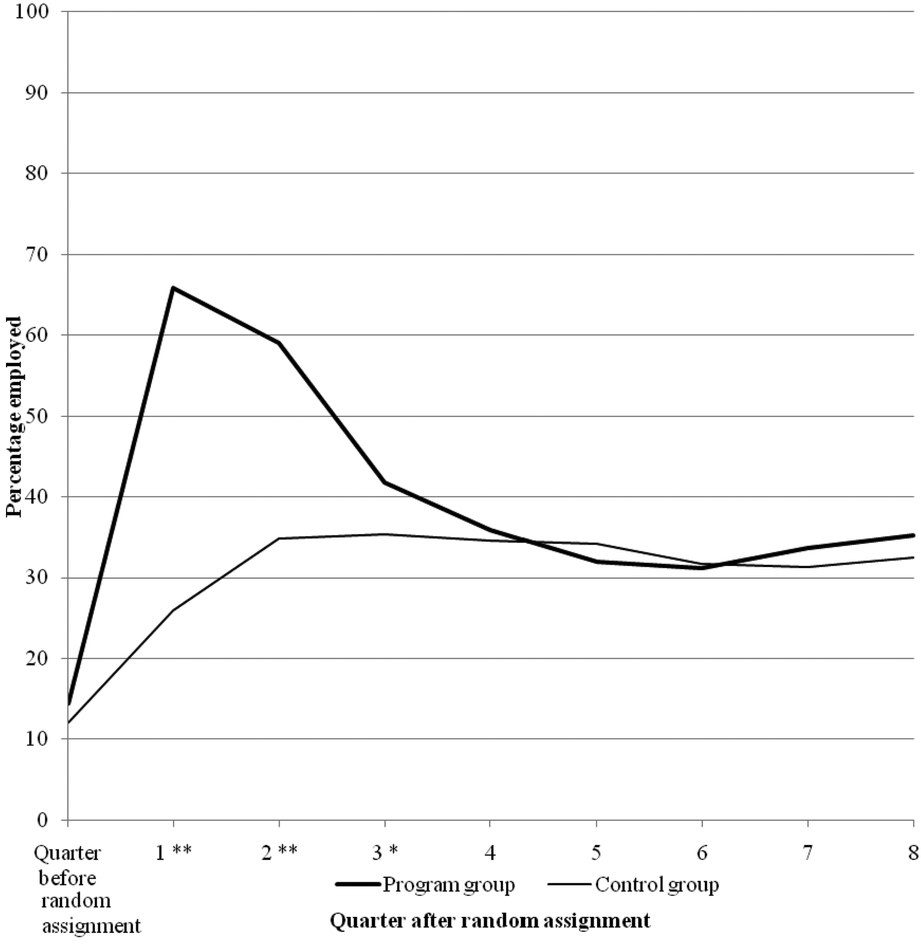
Characteristic	Program Group	Control Group	Full Sample	p Value
Arrest history				
Any prior arrests (%)	100	100	100	N/A
Average number of arrests[a]	7.4	7.7	7.5	0.611
Number of prior felony arrests	4.5	4.6	4.5	0.546
Number of prior misdemeanor arrests	2.8	3.1	2.9	0.842
Ever arrested for a violent crime[b] (%)	67.5	67.5	67.5	0.882
Number of prior arrests for a violent crime	1.4	1.4	1.4	0.770
Conviction history				
Any prior conviction[c] (%)	100	100	100	N/A
Average number of prior convictions[d]	6.6	6.9	6.7	0.683
Number of prior felony convictions	2.7	2.5	2.6	0.002**
Number of prior misdemeanor convictions	3.6	4.1	3.8	0.848
Convicted of a violent crime (%)	51.7	50.9	51.4	0.854
Convicted of a drug-related crime (%)	73.1	73.9	73.4	0.806
State prison history				
Lifetime number of months in state prison[e]	60.6	59.1	60.0	0.760
Months between latest state prison release and random assignment[f] (%)				
1–3 months	41.4	39.4	40.6	0.259
4–6 months	14.7	13.5	14.2	0.318
7–9 months	10.8	11.7	11.2	0.382
More than 9 months	33.0	35.4	34.0	0.166
Parole				
Months remaining on parole	34.2	32.9	33.6	0.012*
Sample size	561	409	970	

Source. MDRC calculations using data from the New York State Division of Criminal Justice Services (DCJS).
Notes. t tests were used to assess differences in characteristics across research groups. The significance level indicates the probability that one would conclude incorrectly that a difference exists between research groups for the corresponding variable. Because of missing data, seven sample members are missing prior criminal histories. The sample sizes vary from 924 to 970 because of missing data. Criminal history includes the arrest, conviction, and incarceration related to the offense for which one was on parole for at the time of random assignment (the current offense).
[a]Each arrest date is counted only as a single event. If there are multiple crimes or charges on the same date, then only the most serious charge is recorded in the analysis.
[b]The violent crime indicator is defined by Penal Law 70.02 and includes underlying offenses.
[c]This outcome excludes convictions where a final disposition was not found.
[d]Each conviction date is counted only as a single event. If there are multiple convictions on the same date, then only the most serious charge is recorded in the analysis.
[e] "Lifetime" includes historical data from as early as 1970.
[f]A total of 48 sample members are missing the latest prison release date and are excluded from this measure.
*$p < 0.05$; **$p < 0.01$.

of a felony and were less likely to be incarcerated in prison for a new crime during the first year after random assignment, and they were less likely to be arrested and less likely to be convicted of a misdemeanor during the second year. Although the mechanism through which CEO reduced recidivism is unclear, the fact that CEO had significant effects on

FIGURE 1

Quarterly Impacts on Employment



Sources. MDRC calculations from the National Directory of New Hires (NDNH) database and UI wage records from New York State.

Notes. The results in this table are weighted by week of random assignment and adjusted for prerandom assignment characteristics. The significance level indicates the probability that one would incorrectly conclude that a difference exists between research groups for the corresponding variable. The sample size is 973. Four sample members are missing Social Security numbers and therefore could not be matched to UI data.

$^{*}p < 0.05$; $^{**}p < 0.01$.

recidivism is unusual, given that research on employment programs has shown a limited effect on crime (Bushway and Reuter, 2002; Visher et al., 2005). The subgroup analysis described subsequently was conducted as part of an effort to improve the understanding of the nature of CEO's impacts on recidivism.

TABLE 3

CEO's Impact on Recidivism in Year 1 and Year 2

Outcome	Program Group	Control Group	Difference (Impact)	*p* Value
Arrested[a] (%)				
Year 1	21.7	22.9	−1.3	0.638
Year 2	22.8	27.5	−4.6	0.098†
Convicted of a felony (%)				
Year 1	1.4	3.1	−1.7	0 .071†
Year 2	5.2	4.5	0.7	0.630
Convicted of a misdemeanor (%)				
Year 1	11.8	12.1	−0.3	0.897
Year 2	14.5	21.5	−7.1	0.004**
Convicted of a violent crime (%)				
Year 1	1.9	1.4	0.5	0.542
Year 2	2.5	4.5	−1.9	0.106
Admitted to state prison[b] (%)				
Year 1	11.0	14.0	−3.0	0.166
Year 2	16.7	17.5	−0.8	0.733
Incarcerated in prison for a new crime (%)				
Year 1	0.8	3.0	−2.2	0.012*
Year 2	3.4	3.8	−0.4	0.742
Incarcerated in prison for a technical parole violation (%)				
Year 1	8.0	9.5	−1.5	0.412
Year 2	11.2	9.5	1.7	0.395
Total days incarcerated in prison				
Year 1	12.0	13.0	−1.0	0.583
Year 2	34.0	40.0	−7.0	0.228
Sample size	568	409		

Source. MDRC calculations using data from the New York State Division of Criminal Justice Services.
Notes. The results in this table are weighted by week of random assignment and adjusted for prerandom assignment characteristics. The significance level indicates the probability that one would incorrectly conclude that a difference exists between research groups for the corresponding variable.
[a] Each arrest date is counted only as a single event. If there are multiple crimes or charges on the same date, then only the most serious charge is recorded in the analysis.
[b] This includes all reasons for incarceration, such as sentences for new crimes, technical violations of parole, detainee (jail), and other reasons. Therefore, incarcerations for new crimes and parole violations do not sum to the percentage incarcerated.
†$p < 0.10$; *$p < 0.05$; **$p < 0.01$.

Current Research Questions

The CEO impact evaluation was an important first step in helping to identify what works for former prisoners as a whole. Although the evaluation included some subgroup analyses, the current paper sought to expand on those efforts by using a regression-based approach to identify subgroups of former prisoners for whom CEO had its greatest impact. The objective was to determine whether limited transitional jobs program resources are best targeted to low-, medium-, or high-risk offenders—with risk levels defined by offenders' prerandom assignment characteristics associated with postrandom assignment recidivism. Specifically, the primary research questions were as follows:

- Which types of former prisoners were most likely to recidivate—as measured by rearrest, reconviction, and reincarceration—after participation in the CEO evaluation, based on their characteristics before random assignment?
- Did participation in CEO reduce recidivism more among former prisoners who were at low, medium, or high risk of reoffending?[5]

Methods

Our methodological approach for answering the primary research questions mirrored and built on that described in Kemple and Snipes (2001). We focused on former prisoners' probability of rearrest, reconviction, and reincarceration in the 2 years following random assignment.[6] Our goal was to differentiate former prisoners into low-, medium-, and high-risk subgroups depending on their risk of recidivism as predicted prior to study participation, and then to examine, within each subgroup, where CEO had its greatest impact on recidivism.

Description of the Data and Measures Used

Data for the current analysis were derived from (a) a baseline questionnaire collected from all study participants at the time of random assignment, (b) criminal history and recidivism data collected from New York State Division of Criminal Justice Services and New York City Department of Correction 1 and 2 years after random assignment, and (c) data from the New York State Department of Labor and the National Directory of New Hires—both of which track quarterly earnings in jobs covered by unemployment insurance. In this section, we describe briefly the domains of measures used in this analysis to develop the most parsimonious model predicting recidivism.

5. This research paper focused on CEO's recidivism effects because of the program's limited impact on unsubsidized employment. However, an exploratory test investigated whether CEO had significant employment impacts for any of the subgroups defined by risk of recidivism, and no significant effects were found in follow-up years 1 or 2 or in the 2 years combined. These results confirmed CEO's overall lack of long-term employment impacts as described in the 2-year evaluation report (Redcross et al., 2009).

6. Recidivism outcomes were defined using data from the New York State Division of Criminal Justice Services and the New York City Department of Correction. Rearrest includes any "unsealed" arrest after random assignment. (An unsealed arrest either was not adjudicated or was disposed before trial, or the arrest resulted in conviction.) Reconviction refers to any conviction with a disposition date after random assignment. Reincarceration includes any admission after random assignment to a New York State prison facility or detention at Rikers Island, New York City's large jail facility, regardless of the length of stay.

Demographics. Demographic information was collected from participants at the time of study entry and included their age, gender, and race/ethnicity. The regression models in this analysis control for participants' age, gender, and race because of their associations with recidivism found in prior criminological studies (see, e.g., Levinson, 2002, for a review).

Time since release. Because former prisoners came to participate in the CEO study for a variety of reasons, their random assignment occurred at varying lengths of time since their prison release. Notably, in the interim CEO evaluation report, MDRC researchers found that participants who came to the study early after release—within 3 months—showed better employment and recidivism outcomes (Bloom et al., 2007; Redcross et al., 2009). Thus, we controlled for time since release in regression models by incorporating a dummy variable measuring those who entered the study within three months of prison release (43% of the sample) and those who did not (57% of the sample).

Education and employment. Information on study participants' prior education was self-reported at baseline; slightly more than half of the participants reported having a GED or high-school diploma, and roughly the same proportion reported having participated in GED courses while in prison (see Table 1). Participants' past employment experiences were both self-reported and assessed using data from the New York State Department of Labor and the National Directory of New Hires—both of which track quarterly earnings in jobs covered by unemployment insurance (UI). Although nearly two thirds of the sample reported working at least 6 consecutive months at some time prior to RA, only one fifth were working in a UI-covered job during the year prior to random assignment.

Partner relationships and children. Marital status and cohabitation information were collected from study participants in the baseline questionnaire. As shown in Table 1, nearly two thirds were single and not living with a partner, few were married, and one fifth was living with a partner but unmarried. The number of minor children (less than age 18) whom participants' had, lived with, and financially supported was also collected. Nearly half of the participants had children, but only one sixth were living with them prior to random assignment.

Housing. Based on study participants' self-reports, few lived in their own home or apartment and well more than half were living in a relative's or friend's house at the time of random assignment (see Table 1).

Criminal history. Many variables measuring study participants' criminal histories prior to random assignment were collected from the New York State Division of Criminal Justice Services and the New York City Department of Correction. Because of the high degree of collinearity among these variables, special efforts were taken during model specification to ensure that only those with the strongest (most statistically significant) relationships to recidivism outcomes were retained. Ultimately, the number of prior arrests emerged as the best predictor of rearrest, reconviction, or reincarceration in the 2 years after random assignment. Study participants had approximately seven prior arrests, on average, at the time of study entry (see Table 2).

Parole status. Former prisoners participating in the study had nearly 3 years remaining on parole at the time of their random assignment, according to information collected from the New York State Division of Criminal Justice Services (see Table 2).

Recidivism. Recidivism data for the first and second year after random assignment was collected for all 977 study participants from the New York State Division of Criminal Justice Services and the New York City Department of Correction. The data covered all rearrests and reconvictions in New York State, plus all reincarcerations in both New York State prisons and New York City jails. Overall, nearly 60% of the sample was rearrested, reconvicted, or reincarcerated within 2 years of random assignment (see Table 3 for specific recidivism rates in each year).

Analytic Strategy

Given the random assignment research design of the CEO evaluation, the baseline characteristics of former prisoners assigned to the control group reflect, on average, those of former prisoners assigned to the program group as well, and this should be true of both observed and unobserved traits. We capitalized on this opportunity presented by experimental data to estimate the risk of recidivism for former prisoners in the program group, using characteristics measured prior to program participation, based on our observations of such risk in the control group. We then classified all participants into low-, medium-, and high-risk subgroups based on these risk scores and evaluated the impact of CEO on recidivism within each subgroup.[7]

Toward this end, our analytic strategy was fourfold. First, we selected a recidivism outcome to use in classifying respondents into low-, medium-, and high-risk subgroups. Given the multiple measures of recidivism (e.g., rearrest, reconviction, and reincarceration) available in the CEO evaluation data, we could have used any one or several outcomes to classify participants. However, for practical purposes, it was more desirable to present and interpret results using only one measure of recidivism. To ensure one measure would capture risk appropriately, we estimated several models predicting recidivism risk using different measures and examined the degree to which the participants fell into the same risk subgroups regardless of the recidivism measure we were using. Ultimately, the global recidivism measure of any rearrest, reconviction, or reincarceration in prison or jail within 2 years of random assignment emerged as that with the highest, average degree of the same participants falling into subgroups created by any individual outcomes. This global measure showed an average of 88% overlap with high-risk participants produced by any individual measure, 70% overlap with medium-risk participants, and 80% overlap with

7. In all analyses, we used weights to adjust for the differential proportion of people randomly assigned to the program/control groups from week to week (Bloom et al., 2007; Redcross et al., 2009).

low-risk participants.[8] For that reason, we felt confident that the global recidivism risk scores herein adequately represented the recidivism risks faced by former prisoners in this study for the purposes of the current analysis. Furthermore, the characteristics we found to predict such recidivism were directly in line with those found by other researchers in several recidivism studies (see, e.g., Levinson, 2002, for a review).

Second, we examined the predictive associations between all baseline characteristics and the global recidivism outcome, measured only within a random half of the control group. Had the analysis followed Kemple and Snipes (2001) exactly, we would have relied on the entire control group. A key limitation of their approach, however, was the potential that this strategy created for overestimating positive program effects among the high-risk group or for overestimating negative program effects among the low-risk group—by misclassifying some program participants into a higher or lower risk group than was true for them. To address and overcome this limitation, we instead randomly selected one half of the individuals in the control group and estimated the model predicting recidivism among those individuals only. These associations represented, theoretically, (a) those that would have been found in any similar sample of former prisoners, especially those that would have been observed in the program group had they not participated in CEO and, equally as likely, (b) those that also would have been observed in the other half of the control group. This step culminated in the derivation of the best-fitting, most parsimonious logistic regression model predicting the probability of recidivism—absent programmatic influences—that we could specify given the data available.[9] We then applied the coefficients from that model to the other half of the control group, as well as to the entire program group, to generate risk subgroups for all study participants (minus the randomly selected subset).

Thus, our third step was to estimate the probability (risk) of recidivism for both program and control participants using regression coefficients from the model run on a randomly selected half of the control participants. For each study participant, we generated a risk of recidivism score and used it to create subgroups of low-, medium-, and high-risk offenders.[10] Our cutoffs for low and high risk were the 25th and 75th percentile risk scores among the ranked distribution for the control group (following Kemple and Snipes, 2001). Thus, low-risk participants had risk scores below the 25th percentile, high-risk participants

8. We assessed the degree of overlap by running a series of crosstabs between subgroups defined using the global recidivism measure and those defined using individual measures. For example, a crosstab between the high-risk subgroup produced using the global recidivism measure and that produced using the rearrest measure showed that 88% of the 977 study participants were identically classified as either high risk or not high risk, whereas 12% were classified in some different combination (e.g., high risk according to global recidivism measure but medium risk according to rearrest measure).

9. We used logistic regression, which measures the probability of an event occurring, because our recidivism outcome was binary (yes/no).

10. So that results can be more easily interpreted and presented for use by practitioners, we take a subgroup-based approach rather than use the continuous risk score index.

had scores above the 75th percentile, and medium-risk participants had scores in between the 25th and 75th percentiles.

Fourth, we analyzed the impact of CEO within each subgroup by estimating a series of regression models—both logistic and ordinary least squares, depending on the nature of the outcome analyzed. We used the same predictors in each outcome model as those in the model estimating risk scores but included an additional variable measuring CEO program group status. From each model's output, we generated adjusted outcomes for the program and control participants to show the size of CEO's impact, while determining the significance of the impact by the p value associated with the program variable's coefficient in each outcome model.

Findings

What Predicts Risk of Recidivism?

Having defined our outcome of interest as the global recidivism measure (any rearrest, reconviction, or reincarceration within 2 years of study entry), we then moved to the second step of our analytic strategy and estimated the risk of such recidivism for all participants. Toward this end, we identified significant bivariate correlates of recidivism from the available baseline characteristics listed in Tables 1 and 2. We then narrowed this list of significant bivariate correlates to those that remained significant ($p < 0.05$) predictors of recidivism after statistical controls measuring participants' demographics (e.g., age, gender, and race/ethnicity) and time since release (less than 3 months or more) were added to the model. From the list of multivariate correlates of recidivism that remained, we worked to specify the best-fitting, most parsimonious model showing only our controls and any other predictors that added significantly to the model. The significance of each predictor's contribution to the model was evaluated by examining the Wald statistic associated with that predictor's coefficient as well as a comparison of the full and reduced model likelihoods, or the likelihood ratio test (see, e.g., Casella and Berger, 2002; Darlington, 1996). Notably, because several baseline predictors were highly correlated with each other and were even subsets of one another (e.g., prior arrests, prior misdemeanor arrests, and prior drug arrests), we included only the more statistically significant of these sets of predictors and omitted the least—especially if the latter did not contribute significantly to the model.[11]

Our analysis culminated in identifying age, gender, and prior arrests as important predictors of recidivism, with statistical controls for race and time since release.[12] Although some prior research also points to the importance of crime type as relevant to recidivism risk

11. We aimed for parsimony in deriving the final model, rather than inclusivity of all possible baseline predictors to reduce the risk of parceling the data into cell counts that were too low for the subgroup sample sizes to support.
12. Although race and time since release are not significant in the model, they are included as controls to determine contributions to recidivism less the effects of these two constructs.

TABLE 4

Logistic Regression Predicting Recidivism 2 Years after Random Assignment[a]

Predictor	Estimate	*p* Value	Odds Ratio	Relative Risk[b]
Age of the sample member	−0.100***	0.0001	0.904	0.951
Male sex	1.756*	0.0315	5.787	2.406
Race is Black/non-Hispanic	0.657	0.3536	1.929	1.389
Race is Hispanic	0.366	0.6183	1.442	1.201
Three months between release and RA	−0.382	0.2697	0.682	0.826
Number of arrests prior to RA	0.238****	<0.0001	1.269	1.126
Intercept	0.356	0.7633		
Pseudo R^{2c}	0.1806			

Source. Calculations using data from the MDRC baseline information form and client survey, and data from the New York State Division of Criminal Justice Services and New York City Department of Correction.
Notes. Model was estimated on $n = 204$ control group participants (randomly selected half). The results are weighted by the week of random assignment and are adjusted for prerandom assignment characteristics.
[a]Recidivism is measured as any rearrest, reconviction, or reincarceration in the 2 years post-RA.
[b]Represents the average relative risk, which equates to the square root of the odds ratio (see Liberman, 2005).
[c]Because R^2 is not available for logistic regression models, we requested the likelihood-based pseudo R^2 provided in SAS output.
*$p < 0.05$; ***$p < 0.001$; ****$p < 0.0001$.

(e.g., Langan and Levin [2002] found that property offenders have higher recidivism rates), we find that the general measure of prior arrests emerged as the strongest predictor of our general measure of recidivism (i.e., rearrest, reconviction, or reincarceration for any crime) in the current study. Notably, although we tested several measures of employment before random assignment, we found that none remained significant predictors of recidivism in our final models, perhaps in part because the employment measures available were indicators of employment status rather than quality, which many criminologists argue is associated more strongly with criminal behavior (e.g., Sampson and Laub, 1993; Uggen, 1999).[13] None of the other available baseline predictors (e.g., housing status and children) was significant in the final model predicting recidivism.

Table 4 shows results from the final model predicting recidivism, as estimated on a randomly selected half of the control group, to capture participants' risk absent any programmatic influences. In this final model, we used mean imputation for the few missing observations on baseline predictors. As shown in Table 4, older former prisoners were less likely to recidivate than younger former prisoners, all else equal. Also, former prisoners who were male and those with more prior arrests were more likely to recidivate than females and those with fewer prior arrests.

13. The employment measures tested included "ever employed," "employed 6 consecutive months," "number of quarters employed," "employed in quarter prior to random assignment," and "employed in year prior to random assignment."

Using the parameter estimates shown in Table 4, we computed the estimated risks of recidivism for both the program and control group participants. We examined the distribution of risk scores for the control group to identify the 25th (lowest) and 75th (highest) percentile scores. We used these scores as cutoffs to divide all study participants into low-, medium-, and high-risk subgroups. Those whose scores were below the 25th percentile (0.495) were said to be at lowest risk of recidivism (n = 194), and those whose scores were above the 75th percentile (0.756) were said to be at highest risk of recidivism (n = 210). All scores in between were classified as medium risk of recidivism (n = 369). The risk scores for both program and control group respondents ranged from a minimum of 0.011 to a maximum of 0.983. The average probability of recidivism among participants in the low-risk subgroup was 0.321, whereas the average probability in the high-risk subgroup was 0.864.

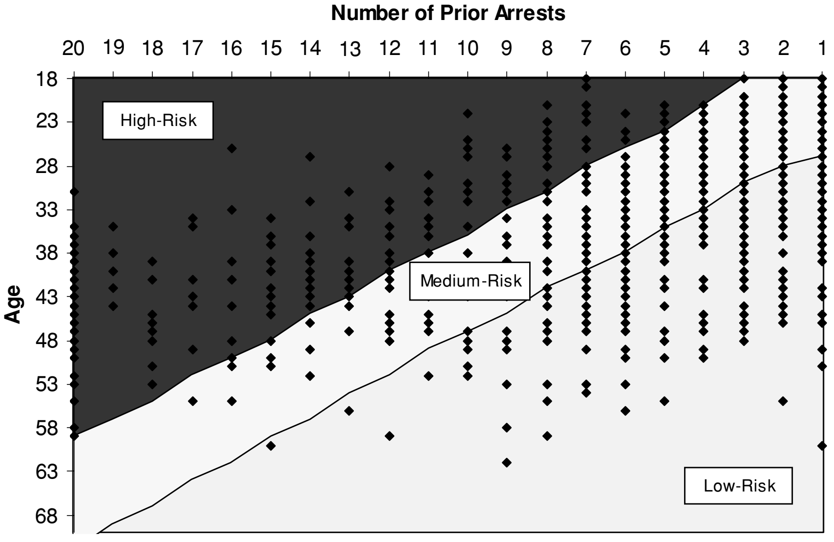
In Figure 2, we illustrate how study participants fall into each of the risk subgroups based on their age and arrest history, while holding constant at the sample means gender, race, and time since release. As shown in the figure, for the average-aged participant (who was 33 years old), those with nine or more prior arrests were placed in the high-risk subgroup; those with five to eight prior arrests were categorized as medium risk; and those with one to four prior arrests were categorized as low risk. Similarly, for participants who have the sample average of seven prior arrests, those who were age 28 years or younger were categorized as high risk; those who were age 29 to 40 years were categorized as medium risk; and those who were age 41 years or older were categorized as low risk. Another way of viewing Figure 2 is to focus on how movement along either axis affects one's probability of recidivism. Movement along the age axis shows that younger offenders, all else equal, had higher probabilities of recidivism (were more likely to be categorized as high risk) than did older offenders. The only factor that kept younger offenders out of the high-risk category was a low number of prior arrests. Conversely, movement along the prior arrests axis shows that those with a higher number of prior arrests, all else equal, were more likely to be categorized as having a high risk of recidivism. Only the oldest of those with many prior arrests (e.g., older than age 50 years) were categorized as being at medium or low risk of recidivism.

Does CEO's Impact on Recidivism Vary by Low, Medium, or High Risk of Reoffending?

Next, we examined whether CEO had a differential impact on study participants' probability of rearrest, reconviction, and reincarceration in the 2 years after random assignment based on their initial risk of such recidivism. Using the same methods employed in the larger CEO evaluation, we estimated the rates of recidivism for each of the predefined risk subgroups. The predictors included in each regression model were those identified previously as controls or as predictors of recidivism risk subgroups (as shown in Table 4), along with a variable measuring CEO program group status. Table 5 shows the adjusted outcomes for the average

FIGURE 2

Definition of Risk Subgroups, by Age and Number of Prior Arrests



Notes. This figure shows former prisoners in the sample and their categorization as high, medium, or low risk of recidivism as determined by regression-based calculations of recidivism risk for different combinations of age and prior arrests, while holding constant at the sample means gender, race/ethnicity, and time since release. As shown in the figure, for the average-aged participant (who was 33 years old), those with nine or more prior arrests are placed in the high-risk subgroup; those with five to eight prior arrests are categorized as medium risk; and those with one to four prior arrests are categorized as low risk. Similarly, for participants who had the sample average of seven prior arrests, those who were 28 years or younger are categorized as high risk; those who were 29 to 40 years are categorized as medium risk; and those who were 41 years or older are categorized as low risk.

program and control participants within each risk subgroup.[14] The table also shows the difference between program and control group outcomes (the impact of CEO) and its statistical significance (*p* value), as derived from the "program group status" variable included in each regression model. The rightmost column of the table gives the *Z* statistic, which measures the significance of the difference in program impacts between the low-risk and the high-risk subgroups (Clogg, Petkova, and Haritou, 1995; Paternoster, Brame, Mazerolle, and Piquero, 1998).

While viewing Table 5, several key findings emerge. First, for former prisoners in the high-risk subgroup, CEO reduced significantly the probability of rearrest, the probability

14. Parameters and odds ratios from all regression models estimated are available from the authors on request.

of reconviction, and the number of rearrests in year 2 after random assignment. Also, for the high-risk subgroup, the impact of CEO in reducing the number of rearrests in years 1 and 2 combined was marginally significant. Furthermore, as shown by the Z statistics, CEO's impact on reducing several recidivism outcomes for those in the high-risk subgroup was significantly greater (better) than its impacts on those same outcomes for the low-risk subgroup. Specifically, for former prisoners in the high-risk subgroup, compared with those in the low-risk subgroup, CEO significantly reduced the probability of rearrest in year 2 and in years 1 and 2 combined; the probability of reconviction in year 2; the probability of reincarceration in year 2; the probability of rearrest, reconviction, or reincarceration in year 2; the number of rearrests in year 2 and years 1 and 2 combined; and the number of misdemeanor reconvictions in years 1 and 2 combined.[15, 16]

Second, there were no significant program impacts on recidivism in year 1 for those in the high-risk subgroup. The impact of CEO on high-risk former prisoners did not emerge until the second year following random assignment to the program.

Last, there were few program impacts on recidivism—of any type—for former prisoners in the low-risk and medium-risk subgroups (see Table 5). For the medium-risk subgroup in the program, there was one statistically significant finding: a decrease in the probability of

15. Former prisoners who entered the CEO study did so at varying points of time after their release from prison. Given that past studies of prisoner reentry have found the highest rates of recidivism during the first year after release (see, e.g., Langan and Levin, 2002), we felt it important to rerun all outcome analyses—within each risk subgroup—focusing on former prisoners who entered the CEO study within 3 months of their prison release. Notably, the interim CEO evaluation report showed that the program had greater impacts on this subsample of study participants (Redcross et al., 2009). For space reasons, we do not present results from this supplemental analysis but do discuss our general findings. CEO's impacts on recently released former prisoners generally mimicked the pattern of findings from the full sample, albeit with some differences. For example, among former prisoners in the recently released high-risk subgroup, CEO significantly reduced the probability of reincarceration in years 1, 2, and 1–2, whereas in full sample results, CEO only showed a lower reduction in year 2 reincarcerations (when the high-risk and low-risk subgroups were compared). In addition, CEO showed a significant reduction in the number of felony reconvictions but not in the overall probability of any reconviction for the recently released high-risk subgroup.

16. As found in the interim CEO evaluation (see Redcross et al., 2009), CEO led to a significant reduction in reincarcerations the first year after random assignment. Empirically, this means that potentially fewer control participants were available (on the street) for possible rearrest, reconviction, or reincarceration during the second year of the follow-up period. With fewer control participants at risk of recidivism, there is the potential for program effects in the second year of study entry to be minimized. We explore whether an analysis limited to study participants who were on the street (i.e., not incarcerated) for at least 9 of the 12 months in year 2 altered our findings at all. We note that this subsample analysis may compromise the random research design of the evaluation, to the extent that unincarcerated study participants are different from those who spent most of year two in prison or jail. For that reason and to preserve space, we briefly discuss our general findings rather than present the full table of subsample results. We found that CEO again had its greatest impacts among the high-risk subgroup of study participants. Among former prisoners who were on the street (not incarcerated) for at least three fourths of year 2, CEO significantly reduced the probability of rearrest, reconviction, and/or reincarceration in year 2, as well as the number of rearrests.

TABLE 5

CEO's Impact on Recidivism Outcomes, by Risk of Recidivism[a]

Outcome	Low Risk: Program Group ($n = 145$)	Low Risk: Control Group ($n = 49$)	Low Risk: Difference/Impact	Low Risk: p Value	Medium Risk: Program Group ($n = 275$)	Medium Risk: Control Group ($n = 94$)	Medium Risk: Difference/Impact	Medium Risk: p Value	High Risk: Program Group ($n = 148$)	High Risk: Control Group ($n = 62$)	High Risk: Difference/Impact	High Risk: p Value	Z Statistic[b]
Rearrested													
Year 1	0.121	0.130	−0.009	0.6534	0.227	0.238	−0.011	0.8250	0.312	0.316	−0.004	0.9400	−0.422
Year 2	0.200	0.079	0.121†	0.0850	0.233	0.245	−0.012	0.9166	0.260	0.420	−0.160*	0.0341	2.512*
Years 1–2	0.271	0.154	0.117	0.1567	0.393	0.406	−0.012	0.8147	0.477	0.562	−0.085	0.3642	1.682†
Reconvicted													
Year 1	0.094	0.147	−0.053	0.2081	0.146	0.143	0.003	0.9433	0.256	0.251	0.005	0.7721	−1.209
Year 2	0.168	0.065	0.103	0.1226	0.218	0.280	−0.062	0.3018	0.231	0.379	−0.147*	0.0426	2.296*
Years 1–2	0.230	0.172	0.058	0.5070	0.315	0.399	−0.084	0.1384	0.403	0.513	−0.110	0.2085	1.256
Reincarcerated													
Year 1	0.216	0.183	0.033	0.7600	0.356	0.464	−0.108	0.1059	0.485	0.504	−0.019	0.9703	0.230
Year 2	0.313	0.173	0.140	0.2102	0.420	0.463	−0.043	0.6606	0.534	0.668	−0.134	0.1106	1.942†
Years 1–2	0.368	0.261	0.107	0.4405	0.513	0.597	−0.084	0.2015	0.609	0.723	−0.114	0.1971	1.412
Rearrested, reconvicted, or reincarcerated													
Year 1	0.234	0.266	−0.032	0.5241	0.408	0.530	−0.122*	0.0453	0.540	0.528	0.013	0.6452	−0.786
Year 2	0.358	0.210	0.148	0.1715	0.491	0.531	−0.040	0.6475	0.575	0.700	−0.126	0.1589	1.943†
Years 1–2	0.402	0.326	0.076	0.7108	0.593	0.679	−0.086	0.1372	0.658	0.755	−0.097	0.3151	0.943
Number of rearrests													
Year 1	0.147	0.143	0.004	0.8124	0.273	0.266	0.007	0.9092	0.512	0.428	0.084	0.6731	−0.484
Year 2	0.313	0.110	0.203	0.1858	0.324	0.319	0.005	0.9394	0.469	0.925	−0.456**	0.0069	2.994**
Years 1–2	0.460	0.253	0.207	0.3745	0.597	0.585	0.012	0.8988	0.981	1.353	−0.372†	0.0927	1.899†
Number of felony reconvictions, years 1–2	0.103	0.034	0.069	0.1765	0.159	0.150	0.009	0.6312	0.188	0.216	−0.028	0.6744	1.113
Number of misdemeanor reconvictions, years 1–2	0.602	0.310	0.292	0.3254	0.632	0.750	−0.118	0.5776	0.998	1.601	−0.602	0.1436	1.768†
Number of days reincarcerated													
Year 1	21.29	16.88	4.41	0.7662	38.59	37.96	0.63	0.6145	54.87	48.64	6.22	0.5924	−0.262
Year 2	42.14	30.36	11.79	0.5094	67.34	65.92	1.42	0.5890	87.09	98.64	−11.56	0.4087	1.059
Years 1–2	63.43	47.24	16.20	.5570	105.92	103.88	2.04	.5484	141.95	147.29	−5.34	.7528	0.614

Source: Calculations using data from the MDRC baseline information form and client survey, and data from the New York State Division of Criminal Justice Services and New York City Department of Correction.

Notes. The results in this table are weighted by the week of random assignment and adjusted for prerandom assignment characteristics.

[a]Risk of recidivism is defined as the likelihood of a new arrest, conviction, or incarceration in the 2 years after the date of random assignment.

[b]The Z statistic, as used herein, measures the significance of the difference in program impacts between the low- and high-risk subgroups (see Clogg, Petkova, and Haritou, 1995; Paternoster, Brame, Mazerolle, and Piquero, 1998).

†$p < 0.10$; *$p < 0.05$; **$p < 0.01$.

rearrest, reconviction, or reincarceration in year 1.[17] There was also a marginally significant finding for the low-risk subgroup: an increase in the probability of rearrest in year 2.

Study Limitations

There are two possible limitations to the primary analysis of interest in this article, which we identify here and describe how we attempted to address. First, despite the study's random design, the relationships of predictor variables to recidivism in the control group are not exactly identical to those we would have observed in the program group (had we been able to do so) because of random sampling error. Therefore, had we used coefficients from the control group at face value to classify all participants into different risk subgroups, it would have been possible to misclassify some program participants into a higher or lower risk group than was true for them. This misclassification could have led to an overestimation of positive program effects in the high-risk group or an overestimation of negative program effects in the low-risk group.

To address and overcome this limitation, we instead randomly selected one half of the control group participants, estimated the logistic regression model predicting recidivism among them only, and applied coefficients from that model to the other half of the control group and the program group to generate risk subgroups. In this way, we treated the randomly selected subsample of control group participants as a separate independent sample, one that was similar in all ways to the other half of the control group and to the program group as well, prior to program participation. By applying coefficient estimates from this random subsample to the unselected control group participants and program participants alike, we avoided the potential for misclassification bias—because the likelihood of such misclassification was distributed equally across the control and program group participants who remained in the final analysis.

Second, given the study's random design, we began with general equality in baseline characteristics between the program and control groups. However, once study participants were stratified into low-, medium-, and high-risk subgroups, it is possible that this effect of random research design was altered (in other words, that program and control participants

17. As described previously and shown in Table 3, the interim results from MDRC's impact evaluation of CEO showed two significant year 1 recidivism effects for program group members, as a whole: reduction in the rate of felony convictions ($p < 0.10$) and reduced likelihood of incarceration for a new crime ($p < 0.05$) (Redcross et al., 2009). Although we do not report risk subgroup results for these two specific outcomes, it was the case—as is implied by the outcomes we do report (namely, number of felony convictions and incarceration for any reason)—that there were no significant year 1 effects for these outcomes within any risk subgroup. A subsequent examination of this apparent discrepancy pointed to only one potential reason: the reduced sample sizes of the data analyzed in this article (reduced because we examined subgroups and reduced because we excluded a random half of the control group). Across both outcomes, the year 1 effects of CEO program participation on these two outcomes in the largest risk subgroup, those at medium risk, was in the appropriate direction but failed to achieve statistical significance ($p > 0.10$).

within one or more risk subgroups were no longer similar to each other, on average, at baseline).

To assess the extent to which this was true, we estimated two logistic regressions within each of the three subgroups. Both regressions predicted the probability of program group status, and the first included virtually all of the baseline characteristics identified in Tables 1 and 2 as predictors, whereas the second included only the subset of baseline characteristics shown in Table 4 to predict recidivism. The results indicated that there was a mixed bag of differences in baseline characteristics for the high-risk group, no overall differences for the medium-risk subgroup (i.e., baseline characteristics for program and control participants were similar), and one difference for the low-risk group. Although participants in the high-risk subgroup showed no differences by program group status for the subset of variables included in Table 4, they did show some differences in the model including all baseline characteristics. Program participants in the high-risk subgroup were somewhat less likely to be married, had more felony convictions, and had fewer misdemeanor and parole violation convictions than control participants in the high-risk subgroup. However, this imbalance was countered by the fact that program and control participants in the high-risk subgroup showed no significant differences in the number of prior felony or misdemeanor arrests, prior convictions for any reason, likelihood of having served a prior prison term, and number of times acquitted of prior charges. Thus, it is unclear whether these differences had any meaningful effects on results. For the low-risk subgroup, the results indicated that although most baseline characteristics were similar between the two types of participants, the program group had a somewhat higher number of prior arrests. This difference may explain, in part, the lack of significant CEO impacts on program participants in the low-risk group (as shown in Table 5).

Discussion and Policy Implications

Visher et al. (2005) called for a rigorous evaluation of current employment programs, particularly transitional jobs programs. The CEO evaluation is responsive to this call, and its findings indicate that CEO has long-term impacts on participants' likelihood of recidivating. Furthermore, the current analysis indicates that CEO's greatest impacts are for those at highest risk of recidivating.

These findings have important implications for policy and practice related to transitional jobs programs for former prisoners. First, in line with previous arguments by parole and reentry experts (for example, Petersilia, 2004; Solomon et al., 2008), and as advanced by proponents of the "risk principle" (Andrews and Bonta, 2003; Andrews and Dowden, 2006), we found that high-risk offenders benefited most from the CEO transitional jobs program.[18] If these results are confirmed by other studies of transitional jobs programs, then

18. The *risk principle* asserts that criminal behavior can be predicted and that treatment should focus on those at the highest risk of reoffending (Bonta and Andrews, 2007).

we can conclude that limited program resources should be targeted toward those at highest risk for recidivating because they are the people helped most by this intervention.[19] The high-risk offenders who participated in the CEO program were less likely to be rearrested, had fewer rearrests, and were less likely to be reconvicted of crimes than high-risk offenders who did not have a chance to participate in the program. Furthermore, those in the low-risk category who participated in CEO had outcomes that were similar to those of the control group. Some researchers have argued that offenders with a low likelihood of recidivating may not require intervention or that they may adopt crime-supportive attitudes and behaviors if they become involved in programs with other offenders (Cullen and Gendreau, 2000, as seen in Petersilia, 2004; Lowenkamp and Latessa, 2004). Clearly, the current analysis offers support for both types of arguments: that offenders at highest risk of recidivism should be targeted when implementing transitional jobs programs and that those at lowest risk are not likely to benefit from such programs. Importantly, however, the current study cannot determine whether program impacts would definitively differ if the program were limited to those at highest risk of recidivating.

Second, in line with the first policy implication, the current analysis contributes information toward future assessments of former prisoners' risk of reoffending. In this sample, a person's age and number of prior arrests were most predictive of recidivism. In particular, for the average-aged person in the sample (33 years old at random assignment), the subgroup at a high risk of recidivism consisted of those with nine or more prior arrests. Similarly, for those in the sample who had seven prior arrests (the average), the high-risk subgroup included those who were 28 years old or younger. These categories offer insight into the types of former prisoners who are best suited for transitional jobs programs similar to CEO's. Furthermore, these findings add weight to past research showing that age and criminal history, in particular, should be included as part of any assessment of recidivism risk.[20] In line with Solomon et al.'s (2008) recommendations on the best strategies for producing successful supervision and reentry, we recommend that supervision agencies and program providers use assessment tools that analyze risk, with two specific elements of risk being individual age and criminal history. Using assessment tools strengthens the ability of staff to make decisions about individual risk better than basing them on personal judgments alone (Cullen and Gendreau, 2000, as seen in Petersilia, 2004; see also, Solomon et al., 2008). Program providers and supervision agencies that are focused on facilitating reentry success often heed this advice already and use assessment tools to measure risk and to

19. This is not to say that high-risk offenders should be "rewarded" with more services than those at low risk, but such intervention may not help or may actually hinder the reintegration process of low-risk offenders.

20. In fact, some (e.g., Coid et al., 2009) have argued that criminal history alone can predict the likelihood of recidivism as well if not better than clinical risk assessments—although this determination cannot be made by nor is it the focus of the current study, given the lack of data on participants' clinical psychological profiles.

target resources. Practitioners in the criminal justice arena are often readily familiar with the assessment tools designed to indicate the probability that an offender will recidivate (Gendreau et al., 2006).

The findings from the current study also have implications for future evaluation efforts. Foremost, it is unclear how the CEO program actually works to reduce recidivism. Interim evaluation results of CEO showed that program participation did not increase one's likelihood of obtaining unsubsidized employment but that it *did* reduce recidivism outcomes, and from this analysis, we saw that this effect was concentrated among high-risk former prisoners (Bloom et al., 2007; Redcross et al., 2009). From a social control/social capital theoretical perspective, it is possible that the participants' daily interactions with transitional jobs program staff—including daily support and guidance from worksite supervisors and job coaches—as well as peer support provided by other transitional jobs participants, affected recidivism outcomes. This type of mentorship, which CEO participants were more likely than control group members to report having received, may have led to changes in the behavior and outlook of participants, even without a lasting impact on unsubsidized employment (Redcross et al., 2009). Notably, a cross-tabulation of CEO service receipt and offender risk level among CEO program participants revealed no significant ($p < 0.10$) differences between the high-risk and low-risk subgroups in the number of weeks worked in a transitional job or in the percentage who met with a job developer or job coach. This leads one to conclude that it was not that high-risk CEO participants received more services than those at low risk, but rather that their *reaction* to the services and social support received differed from that of low-risk CEO participants.

Other research has shown that having social support and developing close relationships with others, in general, can help prevent recidivism (e.g., Visher, Knight, Chalfin, and Roman [2009] showed that close social ties—that is, being married or living with someone as if married—reduced the likelihood of recidivism for a sample of former prisoners). Given that CEO reduced significantly the rates of rearrest, reconviction, and reincarceration, understanding exactly how it accomplished this is an important direction for future research. Similarly, if future interventions are successful at improving employment outcomes, then it will be interesting to observe whether recidivism effects improve substantially in conjunction with employment impacts.

Last, employment programs are just one subset of the programs and resources available to assist former prisoners in a successful reentry process. Our analysis indicates that offender risk of recidivism may be an important element of any reentry program's evaluation. Crucial to this possibility, however, is the rigor of the evaluation. The current analysis' approach was only possible because of the experimental design of the CEO evaluation and its creation of groups of individuals who were similar on background characteristics prior to random assignment. Rigorous evaluation designs strengthen not only the internal validity of findings but also permit more in-depth explorations of for whom such programs work best. When resources permit, future evaluation efforts should be similarly rigorous.

References

Andrews, Don A. and James Bonta. 2003. *The Psychology of Criminal Conduct*, 3rd Edition. Cincinnati, OH: Anderson.

Andrews, Don A. and Craig Dowden. 2006. Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, 50: 88–100.

Aos, Steve, Marna Miller, and Elizabeth Drake. 2006. *Evidence-Based Corrections Programs: What Works and What Does Not*. Olympia: Washington State Institute for Public Policy.

Baer, Demelza, Avinash Bhati, Lisa Brooks, Jennifer Castro, Nancy La Vigne, Kamala Mallik-Kane, *et al*. 2006. *Understanding the Challenges of Prisoner Reentry: Research Findings from the Urban Institute's Prisoner Reentry Portfolio*. Washington, DC: Urban Institute.

Baron, Stephen, John Field, and Tom Schuller (eds). 2001. *Social Capital: Critical Perspectives*. Oxford, UK: Oxford University Press.

Bloom, Dan, Cindy Redcross, Janine Zweig (Urban Institute), and Gilda Azurdia. 2007. *Transitional Jobs for Ex-Prisoners: Early Impacts from a Random Assignment Evaluation of the Center for Employment Opportunities (CEO) Prisoner Reentry Program*. New York: MDRC. Retrieved January 7, 2011 from mdrc.org/publications/468/full.pdf.

Boeck, Thilo, Jennie Fleming, and Hazel Kemshall. 2008. Social capital, resilience, and desistance: The ability to be a risk navigator. *British Journal of Community Justice*, 6: 5–20.

Bonta, James, and Don A. Andrews. 2007. *Risk-need-responsivity model for offender assessment and rehabilitation*. Prepared for Public Safety Canada. Retrieved January 7, 2011 from publicsafety.gc.ca/res/cor/rep/_fl/Risk_Need_2007--06_e.pdf.

Bushway Shawn and Peter Reuter. 2002. Labor markets and crime. In (Lawrence Sherman, David Farrington, Brandon Welsh, and Doris MacKenzie, eds.), *Evidence-Based Crime Prevention*. New York: Rutledge Press.

Casella, George and Roger L. Berger. 2002. *Statistical Inference*, 2nd Edition. Pacific Grove, CA: Duxbury Press.

Caspi, Avshalom, Bradley R. E. Wright, Terrie E. Moffitt, and Phil A. Silva. 1998. Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood. *American Sociological Review*, 63: 424–451.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100: 1261–1293.

Coid, Jeremy, Min Yang, Simone Ullrich, Tianqiang Zhang, Steve Sizmur, Colin Roberts, *et al*. 2009. Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, 77: 337–348.

Cullen, Francis, and Paul Gendreau. 2000. Assessing correctional rehabilitation: Policy, practice, and prospects. In (J. Horney, ed.), *NIJ Criminal Justice 2000: Changes in Decision Making and Discretion in the Criminal Justice System* (pp. 109–175). Washington, DC: U.S. Department of Jsutice, National Institute of Justice.

Darlington, Richard B. 1996. *How Many Covariates to Use in Randomized Experiments?* Retrieved January 7, 2011 from psych.cornell.edu/darlington/covarnum.htm.

Evans, Karen. 2002. Taking control of their lives? Agency in young adult transitions in England and the New Germany. *Journal of Youth Studies*, 5: 245–269.

Farrall, Stephen. 2004. Social capital and offender reintegration: Making probation desistance focused. In (Shadd Maruna and Russ Immarigeon, eds.), *After Crime and Punishment: Pathways to Offender Reintegration*. Devon, UK: Willan.

Gendreau, Paul, Tracy Little, and Claire Goggin. 2006. A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34: 575–608.

Johnston-Listwan, Shelley, Francis T. Cullen, and Edward J. Latessa. 2006. How to prevent prisoner re-entry programs from failing: Insights from evidence-based corrections. *Federal Probation*, 70: 19–25.

Kemple, James J. and Jason C. Snipes, with Howard Bloom. 2001. *A Regression-Based Strategy for Defining Subgroups in a Social Experiment*. New York: MDRC. Retrieved January 7, 2011 from mdrc.org/publications/110/full.pdf.

Langan, Patrick A. and David J. Levin. 2002. *Recidivism of Prisoners Released in 1994*. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics. Retrieved January 7, 2011 from bjs.ojp.usdoj.gov/content/pub/pdf/rpr94.pdf.

Levinson, David, ed. 2002. *Encyclopedia of Crime and Punishment*, Volume 4. Thousand Oaks, CA: Sage.

Liberman, Akiva M. 2005. How much more likely? The implications of odds ratios for probabilities. *American Journal of Evaluation*, 26: 253–266.

Lowenkamp, Christopher T. and Edward J. Latessa. 2004. Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. *Topics in Community Corrections*, pp. 3–8. Retrieved January 7, 2011 from nicic.gov/pubs/2004/period266.pdf.

Nightingale, Demetra and John Trutko. 2008. *Two Models of Transitional Jobs: CEO and TWC*. Unpublished document, prepared for MDRC: Enhanced Services for Hard-to-Employ Demonstration.

Obama, Barack and Joseph Biden. 2008. *Blueprint for Change*. Retrieved January 7, 2011 from barackobama.com/pdf/ObamaBlueprintForChange.pdf.

Paternoster, Raymond, Robert Brame, Paul Mazerolle, and Alex Piquero. 1998. Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36: 859–866.

Petersilia, Joan. 2004. What works for prisoner reentry? Reviewing and questioning the evidence. *Federal Probation*, 68: 4–8.

Piehl, Anne. 2003. *Crime, Work, and Reentry*. Paper presented at the Employment Dimensions of Reentry: Understanding the Nexus between Prisoner Reentry and Work, Urban Institute Reentry Roundtable, May 19–20, New York, NY. Retrieved January 7, 2011 from urban.org/uploadedpdf/410856_piehl.pdf.

Piliavin, Irving and Stanley Masters. 1981. *The Impact of Employment Programs on Offenders, Addicts, and Problem Youth: Implications from Supported Work*. Madison: University of

Wisconsin, Institute for Research and Poverty Discussion. Retrieved January 7, 2011 from eric.ed.gov/PDFS/ED213804.pdf.

Redcross, Cindy, Dan Bloom, Gilda Azurdia, Janine Zweig, and Nancy Pindus. 2009. *Transitional Jobs for Ex-Prisoners: Implementation, Two-Year Impacts, and Costs of the Center for Employment Opportunities (CEO) Prisoner Reentry Program*. New York: MDRC. Retrieved January 7, 2011 from mdrc.org/publications/529/full.pdf.

Sabol, William J., Heather C. West, and Matthew Cooper. 2009. *Bureau of Justice Statistics Bulletin: Prisoners in 2008*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.

Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points through Life*. Cambridge, MA: Harvard University Press.

Solomon, Amy, Jesse Jannetta, Brian Elderbloom, Laura Winterfield, Jenny Osborne, Peggy Burke, *et al*. 2008. *Putting Public Safety First: 13 Strategies for Successful Supervision and Reentry*. Washington, DC: Urban Institute. Retrieved January 7, 2011 from urban.org/UploadedPDF/411800_public_safety_first.pdf.

Uggen, Christopher. 1999. Ex-offenders and the conformist alternative: A job quality model of work and crime. *Social Problems*, 46: 127–151.

Visher, Christy A. and Shannon M. E. Courtney. 2007. *One Year Out: Experiences of Prisoners Returning to Cleveland*. Washington, DC: Urban Institute. Retrieved January 7, 2011 from urban.org/UploadedPDF/311445_One_Year.pdf.

Visher, Christy and Jeremy Travis. 2003. Transitions from prison to community: Understanding individual pathways. *Annual Review of Sociology*, 29: 89–113.

Visher, Christy A., Sara Debus, and Jennifer Yahner. 2008. *Employment After Prison: A Longitudinal Study of Releasees in Three States*. Washington, DC: Urban Institute. Retrieved January 7, 2011 from urban.org/UploadedPDF/411778_employment_after_prison.pdf.

Visher, Christy A., Carly R. Knight, Aaron Chalfin, and John Roman. 2009. *The Impact of Marital and Relationship Status on Social Outcomes for Former Prisoners*. Washington, DC: Urban Institute. Retrieved January 7, 2011 from urban.org/UploadedPDF/411871_returning_prisoners.pdf.

Visher, Christy A., Laura Winterfield, and Mark Coggeshall. 2005. Ex-offender employment programs and recidivism: A meta-analysis. *Journal of Experimental Criminology*, 1: 295–315.

Yahner, Jennifer and Christy A. Visher. 2008. *Illinois' Prisoners Reentry Success Three Years After Release*. Washington, DC: Urban Institute. Retrieved January 7, 2011 from urban.org/UploadedPDF/411748_reentry_success.pdf.

Janine Zweig is a senior research associate in the Justice Policy Center at the Urban Institute. Her research addresses issues relating to vulnerable populations, including former prisoners and victims of violence. She is part of a team of researchers conducting the multisite Transitional Jobs Reentry Demonstration and the Enhanced Services for Hard-to-Employ Evaluation.

Jennifer Yahner is a research associate in the Justice Policy Center at the Urban Institute. Her research interests include studies of prisoner reentry and quantitative methodologies. She is a research analyst on the multisite Transitional Jobs Reentry Demonstration evaluation and the Enhanced Services for Hard-to-Employ Evaluation.

Cindy Redcross is a senior associate in MDRC's Health and Barriers to Employment policy area. Her research focuses on programs serving populations with employment barriers, in particular, former prisoners and long-term welfare recipients. Her current work includes studies of transitional jobs programs, including the multisite Transitional Jobs Reentry Demonstration and the Enhanced Services for Hard-to-Employ Evaluation.

POLICY ESSAY

TRANSITIONAL JOBS PROGRAM

Why the risk and needs principles are relevant to correctional programs (even to employment programs)

Edward Latessa
University of Cincinnati

It is rare today to read an article or study on correctional intervention programs that that does not refer to the work of Andrews, Bonta, and Gendreau and their risk, need, responsivity (RNR) principles (Andrews and Bonta, 1994; Gendreau, 1996). Through the lens of RNR, scholars and practitioners alike have a framework by which they can better study and understand criminal conduct and the effectiveness (or lack thereof) of correctional programs. Indeed, understanding RNR principles provides insight into the findings from Zweig, Yahner, and Redcross (2011, this issue).

Risk Principle

We will start with the risk principle, or the "who" to target—those offenders who pose a higher risk of continued criminal conduct. This principle states that our most intensive correctional treatment and intervention programs should be reserved for higher risk offenders: those with a higher probability of recidivating. Low-risk offenders have a low probability of recidivism and, as such, generally have few risk factors. The question is "why waste our scarce correctional programs on offenders who do not need them?" Furthermore, placing low-risk offenders in intensive programs also can be counterproductive because research has clearly demonstrated that when we place lower risk offenders in our more intensive programs (usually with higher risk offenders), we often increase their failure rates (and, thus, reduce the overall effectiveness of the program (see, e.g, Bonta, Wallace-Capretta, and Rooney, 2000; Lowenkamp, Latessa, and Holsinger, 2006).

Direct correspondence to Edward Latessa, University of Cincinnati, School of Criminal Justice, 508 Dyer Hall, P.O. Box 210389, Cincinnati, OH 45221–0389 (e-mail: edward.latessa@uc.edu).

DOI:10.1111/j.1745-9133.2011.00759.x

Need Principle

The second principle is referred to as the need principle, or the "what" to target—criminogenic factors that have been found to be significantly correlated with criminal conduct. The need principle states that programs should target and focus most efforts on crime-producing needs, such as antisocial attitudes, values, and beliefs; antisocial peer associations; lack of work and financial achievement; substance abuse; lack of problem-solving and self-control skills; as well as other factors that have been found to be correlated with criminal conduct (Andrews and Bonta, 1994; Gendreau, Little and Goggin, 1996).

As noted, one of the domains identified by Andrews and Bonta (1994) is employment and financial achievement. Although Andrews and Bonta were clear that this domain is more than just having employment, the operationalization of this area in practice often is viewed as simply "just get them a job." Unfortunately, as noted in the Zweig et al. (2011) article, some research shows that employment programs alone will not have the desired effects (Bushway and Reuter, 2002; Visher, Winterfield, and Coggeshall, 2005). What we are increasingly learning is that work is much more than just getting someone a job—it involves how work is viewed, the satisfaction one derives from work, how one gets along with coworkers and supervisors, and other work-related aspects linked to attitudes and skills. Indeed, a recent study of parolees in Pennsylvania by Bucklen and Zajac (2009) illustrated these points. Contrary to conventional wisdom, they found little evidence that job acquisition alone was a significant predictor of success or failure on parole. Rather, in the area of employment and finances, they found that those parolees that failed were

- Less likely to have job stability
- Less likely to be satisfied with employment
- Less likely to take low-end jobs and work up
- More likely to have negative attitudes toward employment and unrealistic job expectations
- Less likely to have a bank account
- More likely to report that they were "barely making it" (yet the success group reported more than double median debt).

Furthermore, failures were only slightly more likely to report having difficulty getting a job, and most eventually did. The major contributors to failure included antisocial attitudes, continued association with those with a criminal record, unrealistic expectation about life outside of prison, and poor coping and problem-solving skills (Bucklen and Zajac, 2009).

So although having a job is an important element of reentry and prosocial behavior, we have to look deeper to understand the relationship between work and criminal conduct. Interestingly, in the Bucklin and Zajac (2009) study, successes and failures did not differ in difficulty in finding a place to live after release, and they were equally likely to report eventually obtaining a job. The most important factors centered on attitudes, whether they were about work, behavior, social support systems, or peers.

It also is important to remember that higher risk offenders have multiple risk factors, which is why programs that tend to be one dimensional tend to be less effective than programs that target multiple risk factors (Gendreau, French, and Taylor, 2002). In short, offenders are usually not higher risk because they have one risk factor but because they have multiple risk factors. This issue can clearly be seen with employment. For many offenders, being unemployed is a risk factor, but is it a risk factor for most of us? If we lost our jobs, would we start selling drugs? Steal cars? Mug the elderly? I suspect not. What we would most likely do is find another job. However, if we thought that we could make more money in a day than what others can make in a month, were not into the "9 to 5" grind, not willing to take a lower paying job and work our way up, hung around with friends who think work is for others, and so forth, then not having a job would be a significant risk factor mainly because of these antisocial attributes combined with having a lot of time on our hands to get into trouble. For this reason, correctional programs need to focus on the "big four" predictors of criminal conduct—attitudes, peers, history, and personality (Andrews, Bonta, and Wormith, 2006) before simply sending an ex-offender out to work. Again, research has demonstrated that programs that target several criminogenic factors are much more effective than those that are one-dimensional (Gendreau et al., 2002).

So, For Whom Does a Transitional Jobs Program Work?

The data for the Zweig et al. (2011) article was drawn from a Center for Employment Opportunities (CEO) study, conducted by the MDRC (Redcross, Bloom, Azurdia, Zweig, and Pindus, 2009). This randomized experiment was well designed, and in the end, mixed results were found with regard to employment and recidivism. The CEO did not have a long-term effect on employment; however, there were some significant effects on selected measures of recidivism. The current study goes one step further by separating CEO participants by risk. The findings confirm the risk principle; reductions in recidivism were found for higher risk offenders, with very few differences for the low- and moderate-risk offenders. Furthermore, lower risk offenders actually reported an increase in the probability of rearrest in the second year. Given what we have learned from the risk and need principles, none of these findings were surprising.

Some Lessons Learned

Several lessons can be learned from the Zweig et al. (2011) study. The first lesson is the importance of examining outcomes by risk level. It is likely that over the years some treatment effects have been masked by failing to examine differences by offender risk. The mistake that often is made (by researchers and practitioners) is that we have been looking for treatment effects by comparing low-risk with high-risk offenders rather than by comparing low-risk with low-risk offenders and high-risk with high-risk offenders. Second, by mixing low- and high-risk offenders, programs often can produce iatrogenic effects. This result can occur for several reasons (see Lowenkamp and Latessa, 2004), but having low- and high-risk offenders interact and mingle together is probably not a good idea. Third, relying on one-dimensional programs will unlikely produce the reductions in recidivism that might be obtained by designing multimodality programs. Finally, as noted by Zweig et al., the use of assessment tools to determine risk is much more effective than personal judgments. Taking it one step further, an abundance of evidence shows that using a validated risk/need assessment tool is a key ingredient to meeting the risk and need principles, which in turn helps correctional programs become more effective in reducing recidivism.

References

Andrews, Don A. and James Bonta. 1994. *The Psychology of Criminal Conduct*. Cincinnati, OH: Anderson.

Andrews, Don A., James Bonta, and J. Stephen Wormith. 2006. The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52: 7–27.

Bonta, James, Suzanne Wallace-Capretta, and Jennifer Rooney. 2000. A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice and Behavior*, 27: 312–329.

Bucklen, Kristofer B. and Gary Zajac. 2009. But some of them don't come back (to prison!): Resource deprivation and thinking errors as determinants of parole success and failure. *The Prison Journal*, 89: 239–264.

Bushway Shawn and Peter Reuter. 2002. Labor markets and crime. In (Lawrence Sherman, David Farrington, Brandon Welsh, and Doris MacKenzie, eds.), *Evidence-Based Crime Prevention.* New York: Rutledge Press.

Gendreau, Paul. 1996. The principles of effective intervention with offenders. In (Alan T. Harland, ed.), *Choosing Correctional Options That Work: Defining the Demand and Evaluating the Supply*. Thousand Oaks, CA: Sage.

Gendreau, Paul, Shelia. A. French, and Angela. Taylor. (2002). *What Works (What Doesn't Work) Revised 2002*. Invited submission to the International Community Corrections Association Monograph Series Project.

Gendreau, Paul, Tracy Little, and Claire Goggin. 1996. A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34: 575–608.

Lowenkamp, Christopher T. and Edward J. Latessa. 2004. Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. *Topics in Community Corrections*, pp. 3–8. Retrieved January 7, 2011 from nicic.gov/pubs/2004/period266.pdf.

Lowenkamp, Christopher T., Edward J. Latessa, and Alexander M. Holsinger. 2006. The risk principle in action: What we have learned from 13,676 offenders and 97 correctional programs. *Crime & Delinquency*, 52: 77–93.

Redcross, Cindy, Dan Bloom, Gilda Azurdia, Janine Zweig, and Nancy Pindus. 2009. *Transitional Jobs for Ex-Prisoners: Implementation, Two-Year Impacts, and Costs of the*

Center for Employment Opportunities (CEO) Prisoner Reentry Program. New York: MDRC. Retrieved January 7, 2011 from mdrc.org/publications/529/full.pdf.

Visher, Christy A., Laura Winterfield, and Mark Coggeshall. 2005. Ex-offender employment programs and recidivism: A meta-analysis. *Journal of Experimental Criminology*, 1: 295–315.

Zweig, Janine, Jennifer Yahner, and Cindy Redcross. 2011. For whom does a transitional jobs program work? Examining the recidivism effects of the Center for Employment Opportunities program on former prisoners at high, medium, and low risk of reoffending. *Criminology & Public Policy*. This issue.

Edward Latessa received his Ph.D. from the Ohio State University in 1979 and is a professor and director of the School of Criminal Justice at the University of Cincinnati. Prof. Latessa has published more than 110 works in the area of criminal justice, corrections, and juvenile justice. He is coauthor of seven books, including *Corrections in the Community* and *Corrections in America*. Prof. Latessa has directed more than 100 funded research projects, including studies of day reporting centers, juvenile justice programs, drug courts, intensive supervision programs, halfway houses, and drug programs. Prof. Latessa served as president of the Academy of Criminal Justice Sciences (1989–1990).

Deconstructing the risk principle

Addressing some remaining questions

Gerald G. Gaes
William D. Bales
Florida State University

The article by Zweig, Yahner, and Redcross (2011, this issue) is one of the strongest tests to date of the risk principle. Offenders were randomly assigned to a transitional jobs program. The risk of reoffending levels—high, medium, and low—were based on exogenous factors that were in place prior to the intervention. The results showed that the transitional job program reduced recidivism relative to control group participants for the high-risk offenders but not for the low- or medium-risk offenders. One of the few ways that Zweig et al. could have improved on their design would have been to use preexisting risk levels as a blocking variable and to assign offenders randomly to the intervention and control groups within each of these risk levels.

In this policy essay, we accept the premise of the risk principle, but we pose certain questions that should be addressed by criminologists to further our understanding of the mechanisms at work, and to enhance its utility as a public policy tool. We start by deconstructing elements of the risk principle, acknowledging the original statement by Andrews, Bonta, and Hoge (1990). We also give credit to expositions by Lowenkamp and Latessa (2004) and Lowenkamp, Latessa, and Holsinger (2006) in expressing the relationship among risk, supervision, and program intensity.

Risk Principle Defined

Offenders are distributed along a dimension of risk to commit crimes based on some scalar assessment of that dimension. Based on the risk scale, the risk principle suggests the following actions, especially with regard to community supervision:

1. The higher the risk level, the higher the level of community supervision.
2. The higher the risk level, the greater the required level of rehabilitative intervention.

Direct correspondence to William D. Bales, College of Criminology & Criminal Justice, Florida State University, 634 W. Call Street, Tallahassee, FL 32306–1127 (e-mail: wbales@fsu.edu).

DOI:10.1111/j.1745-9133.2011.00777.x

3. Low-risk offenders require little or no services, and it may be that there is a perverse effect of assigning intense rehabilitation services to low-risk offenders that could create criminogenic effects.

The Higher the Risk Level, the Higher the Level of Community Supervision

Many jurisdictions have adopted risk assessment tools, and they use them to divide their supervised population into different risk levels, assigning the highest levels of supervision to the highest risk groups (e.g., Latessa and Lovins, 2010; Shaffer, Kelly, and Lieberman, 2001). This is an efficient way to assign limited supervision resources. Indeed, in some jurisdictions, such as the states of Maryland and Washington, and New York City, the lowest risk offenders report to kiosks to measure supervision compliance (Baker, n.d.). In earlier work on inmate risk assessment and classification, it was argued that the tools that measured risk had to be normed on the jurisdiction population. We presume the same guidance should hold for different community supervision jurisdictions. Nonetheless, other questions develop that may be jurisdiction specific.

First, what thresholds should a jurisdiction adopt to decompose its supervised population into the various levels of risk? If a jurisdiction has the resources to conduct a normative study within its authority, then the thresholds can be based on the levels of recidivism the jurisdiction establishes. Second, although we do not want to complicate this proposed solution, analysts and practitioners should be aware of the analytical problem of conducting a normative study when offenders are assigned to levels of supervision that are intended to suppress the behaviors predicted by the risk assessment device. This dilemma has been posed by Bushway and Smith (2007), and a solution has been suggested by Rhodes (2010).

As Rhodes (2010: 57) pointed out, "Researchers do not always convey their purpose when developing prediction instruments, but it seems that they intend to estimate the inherent power of observed risk/prediction variables (the X_O) to predict the outcome (Y) in the absence of control and correctional responses." Rhodes (p. 63) also noted that "[t]he more effective are control/correctional responses, the worse are the predictions. Regrettably, the more effective are control/correctional responses, the more important are good predictions. This is perverse: The more important the policy relevance of good predictions, the greater the harm done by following best estimation practices." In other words, the initial assessment of the risk instrument will only be correct to the extent that jurisdictions make no effort to control, monitor, or suppress criminal behavior while they are first developing their instrument. This may work in a jurisdiction that has no supervision policy based on risk assessment, but it will fail in jurisdictions that do have one. Rhodes offered a technical solution to the problem, showing how to model both the individual level factors that determine criminal recidivism and the supervision factors that suppress criminal recidivism, and readers are encouraged to read his paper if they are developing a risk assessment, renorming it, or modifying it. Second, the more effective the jurisdiction's control/supervision strategies, the worse the prediction will be when one renorms or revalidates the instrument without using Rhodes' procedure. Thus, what is assumed to be current best practice may produce invalid risk estimation and bad public policy choices.

As there has never been an empirical test along the lines Rhodes (2010) suggested, it leaves open the possibility that the individual level factors that predict future criminality may be so compelling—have a much higher weight than community supervision factors—that failure to account for these community control processes will not have a major impact on the validity of the prediction instrument. Or, that in fact community supervision is not very effective at suppressing criminal misconduct.

Putting aside this more abstract methodological concern, administrators of community supervision agencies can develop thresholds of high risk and monitor the risk levels of their supervised population over time to see where to allocate the high supervision resources. These latter resources refer to technology such as GPS devices (Bales, Mann, Blomberg, McManus, and Dhungana, 2010) or human resources (Jalbert, Rhodes, and Flygare, 2010) where smaller caseloads and more follow-up can be devoted to the higher risk cases.

The Higher the Risk Level, the Greater the Required Level of Rehabilitative Intervention

Lowenkamp and Latessa (2004) have provided insight into the nature of this risk principle corollary. The reason higher risk individuals are in fact more risky is because they have deficits on multiple domains that affect criminality. These domains include drug dependence, the weight of prior criminality, criminal peers, criminal thinking, criminal attitudes, and other dimensions. To address this multifaceted cocktail of needs, rehabilitation programs must be designed to address such needs. The interaction between proper assessment and proper treatment is the key according to advocates of this corollary.

The Zweig et al. (2011) study does suggest that the services provided under the rubric of transitional jobs programs included more than just employment services, including job coaches who conducted weekly checks of the participants' progress and provided referrals for various support services such as transportation and substance abuse treatment. The authors themselves suggest that learning was occurring along several domains of criminal needs as a result of the work of the job coaches and worksite supervisors, who served as role models. Offender participants also received peer support from other CEO participants.

Although there is theoretical justification for providing more services to high-risk offenders, we are not aware of any empirical investigations that show that by stacking programs on top of one another for high-risk offenders, greater results are achieved to produce marginal reductions in offending. An even more compelling result would be that such stacking produces an interaction effect where the stacking results in a decrease in recidivism beyond the effect of the individual interventions themselves, a kind of multiplier effect.

One of the major studies that we are aware of which addressed the treatment of high-risk offenders that also provided evidence on the types and number of programs offered to offenders and the relationship of program participation to recidivism was the multisite evaluation of the Serious and Violent Offender Reentry Initiative (SVORI; Lattimore and Visher, 2009). According to the final report, "The study participants were high-risk offenders who had extensive criminal and substance use histories, low levels of education and employment skills, and families and peers who were substance and criminal justice system involved" (p. ES-8). The final report shows that although SVORI participants received more services than their non-SVORI counterparts, the level of services were far short of their needs and declined substantially after release. Furthermore, "participation in SVORI programs was associated with moderately better outcomes with respect to housing, employment, substance use, and self-reported criminal behavior, although these improvements were not associated with reductions in official measures of reincarceration" (p. ES-9). In contrast to the current study, which did not demonstrate a positive effect of the transitional job programs on job acquisition, but could demonstrate an effect on recidivism, the SVORI study found an effect of enhanced prison and community services on intermediate outcomes, such as housing, employment, substances abuse, and self-reported criminal behavior, but not on officially recorded criminality. Furthermore, there has been no evidence from SVORI that the stacking of programs produced better results for these high-risk programs.

The bottom line is that we still do not know a great deal about how to design a comprehensive plan to address high-risk offender's needs, and perhaps more importantly, there may not be sufficient resources in most communities to provide for those services in the community. However, results such as the current evaluation should provide us with encouragement that it is possible to have an effect.

Low-Risk Offenders Require Little or No Services, and It May Be That There Is a Perverse Effect of Assigning Intense Rehabilitation Services to Low-Risk Offenders That Could Create Criminogenic Effects

This last corollary of the risk principle suggests that we will do no harm if services are not provided to low-risk offenders. Rather, services may increase the risk of recidivism for low-risk offenders. Although the Zweig et al. (2011) evaluation did not produce statistically significant results supporting this conclusion, close inspection of their Table 5 shows that if there had been a larger sample size, this could have occurred in the low-risk group in this study.

Lowenkamp and Latessa (2004) and Lowenkamp et al. (2006) have explained this phenomenon based on three potential mediating influences. First, the treatment of low-risk offenders in the community probably exposes them to high-risk offenders who are also receiving treatment, and this creates the opportunity for them to learn antisocial behavior from the latter group and form peer associations with individuals who reinforce offending. There is some evidence for this phenomenon in the juvenile literature (Osgood and Briddell, 2006), but it is weak. Second, participation in the program interferes with the normative activities of low-risk offenders disrupting their normal reintegration back into society. Third, enhanced supervision and surveillance with more demanding conditions may increase the probability of violations.

Whether or not participating in community supervision programs actually interferes with the successful integration of low-risk offenders into the community really depends on the way the program is designed. If it is the kind of program that requires the offender to give up his/her job and spend time in intensive supervision under the constant watch of the program provider as well as the community supervision agent, perhaps this is a causal mechanism for failure. However, the transitional job program described by Zweig et al. (2011) did not seem to be designed and implemented in that manner, yet had a criminogenic effect on low-risk offenders. To tease out these relationships, studies should be designed either to evaluate or manipulate the various components of this corollary. Here is the optimal experimental 2 × 2 × 2 design. Take something like the transitional jobs program or some other program that we know works and assign low-risk community supervision cases to participate in the program and assign them to participate with either other low- or high-risk participants. Vary the levels of supervision. Also, vary the setting in which the program occurs, allowing half of the offenders to carry on the program activities after hours promoting minimal interference in reintegration, and the other half to participate in a custodial setting. We do not really anticipate that such a study will ever be conducted either as an experiment or as a quasi-experimental design. Our interest here is pedagogical. However, this is what it would take to sort out these competing influences.

Policy Implications

One policy implication of this policy essay is that jurisdictions that have incorrectly developed or normed their risk assessment instrument, when there is also a correctional response such as increased supervision or rehabilitative services for higher risk offenders, will produce a fundamentally flawed tool for setting risk levels and supervision thresholds. Furthermore, an instrument normed in one jurisdiction may be totally inappropriate for another jurisdiction, not necessarily because the supervision populations are different, as is typically assumed, but because the two different jurisdictions use diverse supervision and rehabilitation strategies. Jurisdictions that either develop their own risk classification devices or use "off-the-shelf" proprietary tools should take care, and perhaps seek, analytical services before applying these tools.

There may be jurisdictions that are unwilling to, or as a result of legal impediments cannot, prevent low-risk offenders from receiving services. At this point, the research community has not been able to tease out the best way of delivering those services, and there is the potential to do more harm than good. The best policy advice at this point is to provide those services with the least amount of interference in the daily lives of low-risk participants, and to try to ensure that low-risk offenders are not mingled with high-risk offenders in a treatment setting.

Finally, although there is no systematic evidence that stacking multiple services for an offender provides a benefit beyond the specific individual services, jurisdictions will have to use common sense in providing those services. If an individual has many needs, proper assessment may point out the most important deficits that require special services or interventions. The science of assessment is still too immature to suggest which needs should be addressed first, which needs are primary, or which combination or sequence of services will produce optimal results.

Summary

The risk principle is close to becoming a law in the study of criminology and criminal justice. It does not have the status of the age/crime curve, but clearly it has permeated the consciousness of many criminal justice practitioners and administrators. Although the Zweig et al. (2011) article is a meaningful contribution to this important area of inquiry, and has clear policy implications for programs and interventions in the community corrections and prisoner reentry arena, we have laid out a set of issues that should be addressed to improve our understanding of the limits and boundaries of the risk principle.

References

Andrews, Don A., James A. Bonta, and Robert D. Hoge. 1990. Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17: 19–52.

Baker, Rosalyn. n.d. Automated kisok reporting for offenders. Retrieved from fdle.state.fl.us/Content/getdoc/01da8494-55f5-4045-b20d-ac42503724af/Baker-rosalyn-final-paper.aspx.

Bales, William D., Karen Mann, Thomas Blomberg, Brian McManus and Karla Dhungana. 2010. Electronic monitoring in Florida. *The Journal of Offender Monitoring*, 22: 5–12.

Bushway, Shawn and Jeffrey Smith. 2007. Sentencing using statistical streatment rules: What we don't know can hurt us. *Journal of Quantitative Criminology*, 23: 377–387.

Jalbert, Sarah K., William Rhodes, and Christopher Flygare. 2010. Testing probation outcomes in an evidence-based practice setting: Reduced caseload size and intensive supervision effectiveness. *Journal of Offender Rehabilitation*, 49: 233–253.

Latessa, Edward J. and Brian Lovins. 2010. The role of offender risk assessment: A policy guide. *Victims & Offenders*, 5: 203–219.

Lattimore, Pamela K. and Christine A. Visher. 2009. *The Multi-site Evaluation of SVORI, Summary and Synthesis*. Grant No. 2004-RE-CX-002, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Retrieved from svori-evaluation.org/documents/reports/SVORI_Summary_Synthesis_FINAL.pdf.

Lowenkamp, Christopher T. and Edward J. Latessa. 2004. Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. *Topics in Community Corrections*, pp. 3–8. Retrieved from nicic.gov/pubs/2004/period266.pdf.

Lowenkamp, Christopher T., Edward J. Latessa, and Alexander M. Holsinger. 2006. The risk principle in action: What we have learned from 13,676 offenders and 97 correctional programs. *Crime & Delinquency*, 52: 77–93.

Osgood, D. Wayne and L. O'Neill Briddell. 2006. Peer effects in juvenile justice. In (Kenneth A. Dodge, Thomas J. Dishion, and Jennifer E. Lansford, eds.), *Deviant Peer Influences in Programs for Youth*. New York: Guilford Press.

Rhodes, William. 2010. Predicting criminal recidivism: A research note. *Journal of Experimental Criminology*, 7: 57–71.

Shaffer, Deborah K., Bridget Kelly, and Joel Lieberman. 2011. An exemplar-based approach to risk assessment: Validating the risk management systems instrument. *Criminal Justice Policy Review*, 22: 167–186.

Zweig, Janine, Jennifer Yahner, and Cindy Redcross. 2011. For whom does a transitional jobs program work? Examining the recidivism effects of the Center for Employment Opportunities program on former prisoners at high, medium, and low risk of reoffending. *Criminology & Public Policy*. This issue.

Gerald G. Gaes, Ph.D, is a Visiting Researcher at Florida State University and a criminal justice consultant. He is former Director of Research for the Federal Bureau of Prisons and has published in *Criminology and Public Policy*, *Justice Quarterly, and the Journal of Experimental Criminology among other crime and policy journals*.

William D. Bales, Ph.D., is an Associate Professor at Florida State University's College of Criminology and Criminal Justice. He has published in *Criminology*, *Criminology and Public Policy*, *Journal of Research in Crime and Delinquency*, and *Justice Quarterly*, among other crime and policy journals.

COMMUNITY-DRIVEN VIOLENCE REDUCTION PROGRAMS

Community-based partnerships and crime prevention

Wesley G. Skogan
Northwestern University

In their article, Jeremy M. Wilson and Steven Chermak (2011, this issue) provide an evaluation of a community-based youth violence prevention program, Pittsburgh's One Vision One Life (OVOL). Facing a record-setting rise in homicide, a local coalition of organizations launched a street-work program that intervened to defuse impending disputes and identified high-risk youth who could be connected with services. Three clusters of neighborhoods were targeted by the program, whereas others served as comparison groups in the evaluation.

Why should this journal, and the research community, devote attention to OVOL and programs like it? First, and foremost, in my view, it promises a *nonenforcement alternative* to violence prevention. If we are going to maintain democracy in a world that is increasingly governed through fear and punishment, it behooves us to err on the side of paying close attention to approaches to peace and stability that rely on civil society rather than on criminal justice institutions. Research on hot spot interventions and problem-oriented criminal justice programs, to pick examples of the latter, is more than "promising." We know a great deal about why, and how, they work. However, the contrast between the vigor and the rigor of research in those fields and this one is striking. So too are the theoretical perspectives from which they operate, with those on the civil society side worrying about individual and collective norm change and harm reduction, not about fear and deterrence.

Second, we should pay attention because OVOL is representative of *how services are delivered* in the United States. Community-based partnerships—coalitions among grassroots organizations, nonprofit service providers, local government agencies, and funders—are to a significant degree replacing the older model of creating public agencies and staffing them

Direct correspondence to Wesley G. Skogan, Institute for Policy Research, Northwestern University, 2040 Sheridan Road, Evanston, IL 60208–4100 (e-mail: skogan@northwestern.edu).

DOI:10.1111/j.1745-9133.2011.00782.x

with public employees who provided services to community members. We are abandoning permanent, professional, and often unionized civil servants (read them as "expensive") for more temporary, well-intentioned, task-oriented contractors who may be dumped when tax revenues falter. They hope to protect themselves from this fate by finding political patrons, hiring local influentials to staff the program, and cultivating friends in the media. With luck they will find themselves—as Andrew V. Papachristos in his policy essay (2011, this issue) dubs it, "too big to fail." This approach is now how we deliver child care, family support services, preventive physical and mental health care, affordable housing and housing rehabilitation, recreation, adult education, employment counseling and job training, and a myriad of services and activities for senior citizens. OVOL is just a case study of programming in the criminological domain. Governance has become contract-letting, whereas in the public sector, even police numbers are at risk.

As Wilson and Chermak (2011) make clear, OVOL is not alone in the world of largely nonenforcement, community-based, crime-prevention partnerships. While explaining the program, and later decoding the findings of the evaluation, Wilson and Chermak detail the relationship between Pittsburgh's program and related projects in other cities, notably, Baltimore, Chicago, and Boston. The apparent commonalities and contrasts among them will provide the reader with a useful picture of this world of prevention services.

Of course, readers also will want to know whether all this can actually work. Mixed evidence is available regarding the effectiveness of this and similar programs, as well as a mixed evaluation record. They are hard to implement, it is challenging to say exactly what they do, and it is difficult to link whatever they did to measured outcomes.

As a reader of this article by Wilson and Chermak (2011) and of evaluations of similar programs will quickly note, they are hard to implement and sustain. Our colleagues who evaluate policing strategies lament the difficulties of funding, organizing, and keeping top managers focused on crackdowns by their troops in a few targeted geographical areas. They should take a peek into the world of community-based partnerships. There, programs depend on the coordinated activities of multiple organizations with distinct agendas, different budget cycles, highly varying degrees of professionalism, sometimes volunteer staffing, and frequent financial crises. Of course, they also often have differing views on how—and whether—they should be evaluated. They are beholden to different and often competing politicians. The "grassroots" components of the partnerships often are as critical and oppositional as they are interested in cooperating with criminal justice agencies, which are not popular among their constituents. David M. Kennedy in his policy essay (2011, this issue) is unwilling to give up the enforcement tools in his criminological tool bag, although he is sensitive to the implications of pulling them out. However, many groups that actually represent their community's views may not be willing to take up the hammer.

It is hard to keep track of what program staff actually does, and the shifting population they work with. OVOL and related programs are not situated in schools or other controlled settings. Strikingly, reviews of successful programs (such as those summarized at the Center for the Study and Prevention of Violence's "Blueprints for Violence Prevention" website) most often identify school-based or pre/postnatal initiatives. However, as one staff member I interviewed as part of my CeaseFire-Chicago evaluation bluntly put it, "Gangsters aren't in school." Instead, they find fellowship as far from the constraints of adult supervision as they can place themselves. OVOL and the others deal with potential offenders "in the wild." They run down clients, who may be of no fixed address, on the street. They work in the night and in places where no institutional-review-board–governed principal investigator would be allowed to send his or her staff.

Also, it is difficult to link these activities plausibly to measured outcomes. Our policy essay commentators on the article by Wilson and Chermak (2011) all lament the shortage of randomized trials in this field, which is a familiar chant. Some, like CeaseFire-Chicago, are at least multisite, which yields some analytic leverage. All of the programs discussed here ended up employing retrospective, data-analysis–intensive designs for detecting program effects, and Megan Ferrier and Jens Ludwig's (2011, this issue) policy essay is a reminder of the fallibility of this approach. But they all came to their evaluations only after the programs had apparently proved their worth. Like program funders, research sponsors also want to minimize the risk of betting scarce dollars on a program that will not pay off, so both groups are adverse to getting in on the ground floor.

Strikingly, one persistent finding of the OVOL evaluation (it occurred in all target neighborhoods areas and in some nearby spillover areas) is that it may have caused violent crime to go *up*, not down. Wilson and Chermak (2011, this issue) and the policy essay commentators mull over this point, as doubtless will the reader. Malcolm W. Klein (2011, this issue) points to the pernicious role that the program's rough-and-tumble staff may have played in glorifying gangs. Other policy essayists point to Klein's own research to infer that OVOL may have *fostered* gang identification and cohesion, in a city where gangs were not particularly well organized before. Kennedy (2011, this issue) observes that the hostility of the street workers to the police may have further undermined the confidence of their clients in the police, and it may have encouraged a "stop snitchin'" attitude. In short, this dialog is a reminder that "first, do no harm" is a relevant injunction in the social service world as well as in medicine.

Where should we go from here? First, the evidence on how, and even whether, community-based client service and street intervention programs work is sufficiently mixed that we should continue to pursue what Ferrier and Ludwig (2011, this issue) describe as "efficacy trials." These evaluate small but carefully developed field tests of closely monitored programs, looking directly inside the "black box" to see what activities actually take place, what is doable and what is not, and what seems to work. It is advisable to conduct several of these, testing various program permutations, with big-bang "effectiveness trials" only coming later. One message of research by Klein (2011, this issue) and our other policy commentators is that what the interventions look like should vary in line with the nature of the gang problem in the community. The model for Pittsburgh might be different from that for Los Angeles or Chicago, with their large, cohesive, and professionally led gangs. Second, clearly, the role of randomized experiments is important at the efficacy-testing stage. These programs usually involve a client-service component, providing a natural home for finding what works in that domain. Third, community-based programs should seriously entertain the possibility that they need a strong law enforcement arm to get the attention of the crime groups and most chronic offenders at work in their target communities. Kennedy (2011, this issue) describes a "mixed-mode" program strategy in Boston that followed these lines, facilitating coordination-at-a-distance between street workers and the police.

References

Ferrier, Megan and Jens Ludwig. 2011. Crime policy and informal social control. *Criminology & Public Policy*. This issue.

Kennedy, David M. 2011. Whither street work? The place of outreach workers in community violence prevention. *Criminology & Public Policy*. This issue.

Klein, Malcolm W. 2011. Comprehensive gang and violence reduction programs: Reinventing the square wheel? *Criminology & Public Policy*. This issue.

Papachristos, Andrew V. 2011. Too big to fail: The science and politics of violence prevention. *Criminology & Public Policy*. This issue.

Wilson, Jeremy M. and Steven Chermak. 2011. Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life. *Criminology & Public Policy*. This issue.

Wesley G. Skogan is a professor of Political Science and Faculty Fellow in the Institute for Policy Research at Northwestern University. Skogan has directed many major crime studies on fear of crime, the impact of crime on communities, public participation in community crime prevention, victimization, and victim responses to crime. Since 1993 he has directed an evaluation of Chicago's experimental citywide community policing initiative. His newest projects include an evaluation of the utilization and impact of information technology in law enforcement, and an evaluation of CeaseFire, a Chicago crime prevention program.

Overview of: "Community-driven violence reduction programs

Examining Pittsburgh's One Vision One Life"

Jeremy M. Wilson
Steven Chermak
Michigan State University

Research Summary

We assessed the effect of the One Vision One Life program on violence by comparing target areas with comparison areas constructed by propensity scores and by program staff recommendations, and by examining areas adjacent to the target areas. We found the program was not associated with changes in homicide but was associated with increases in aggravated and gun assaults. Whereas aggravated assaults increased in one spillover area and decreased in another, gun assaults increased in one spillover area and did not statistically change in the other.

Policy Implications

The findings raise several critical issues for similar and future initiatives. Among others, these include the transferability of success in programs elsewhere and elements missing in the Pittsburgh implementation. Successful results from similar programs suggest the promise of these programs, whereas the Pittsburgh results suggest the need for continued rigorous evaluation.

Keywords

violence, community, gangs, crime prevention, problem solving, quasi-experiment

DOI:10.1111/j.1745-9133.2011.00762.x

Community-driven violence reduction programs

Examining Pittsburgh's One Vision One Life

Jeremy M. Wilson
Steven Chermak
Michigan State University

Despite some evidence of reductions (FBI, 2009a), violent crime remains among the most important social problems affecting the quality of life in communities throughout the United States. Aggregate reductions also mask the variability in violence among and within communities. The total number of persons annually victimized by violence remains high. In 2008, more than 9,000 persons were killed with guns (FBI, 2009b). In 2006, 71,000 persons suffered nonfatal gunshot wounds, and 2.1 million persons sustained an injury requiring emergency-room treatment as a result of a violent incident (CDC, 2010). Overall, more than six million individuals were victimized by crimes of violence in 2006 (BJS, 2007). One comprehensive review of gun research indicated that firearms play a significant role in violence and that young persons are particularly vulnerable to violence and death from firearms (Wellford, Pepper, and Petrie, 2005).

The impact of violent crime on individuals, families, and communities is substantial. Some estimates indicate that the annual costs of gun violence are approximately $100 billion (Cook and Ludwig, 2000). The annual costs of all personal victimization by violence, including intangible losses such as pain, suffering, and reduced quality of life, are more than $450 billion (NIJ, 1996). This figure is dated and likely to be significantly higher today.

This research was supported by Grant 2006-IJ-CX-0030 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice, and by the Richard King Mellon Foundation, and it was conducted under the auspices of the Safety and Justice Program within RAND Infrastructure, Safety, and Environment (ISE). The points of view or opinions expressed in this article are those of the authors and do not necessarily represent the official position of the National Institute of Justice, U.S. Department of Justice, or the Richard King Mellon Foundation. Direct correspondence to Jeremy M. Wilson, School of Criminal Justice, Michigan State University, 560 Baker Hall, East Lansing, MI 48824 (e-mail: jwilson@msu.edu).

DOI:10.1111/j.1745-9133.2011.00763.x

Indeed, Cook and Ludwig (2000: 138) suggested that "the costs of violence are so great that effective interventions essentially pay for themselves."

The extent of violence and its impact highlight a critical need to develop and implement effective programs to reduce it. Many communities have initiated a wide range of responses to violent crime, firearm-related violence, and drug crimes. These interventions cover a wide range of approaches, including public health, media publicity, technology, community-driven, and criminal justice initiatives. Scholars have produced an overwhelming number of studies on these initiatives using data and methods of evaluation that range greatly in quality. Although previous evaluations indicate that there are certain types of strategies and specific programs that are promising, there is still a great need for additional critical evaluations. As the National Institute of Justice (NIJ) (2002: 19) noted, after compiling and analyzing a representative selection of NIJ research on gangs, there remains "a need to know 'what works'. . . too little is known about the relative merits of comprehensive, broad-based interventions." More recently, Weisburd and Neyroud (2011: 11) reiterated that, "what is most striking about policing is that we know little about what works, in what contexts, and at what costs." Moreover, most evaluations of gang interventions examine enforcement strategies that are primarily implemented by law enforcement organizations. In short, a critical need remains for researchers to evaluate promising strategies rigorously, to broaden understanding of promising strategies by replicating them and their evaluations at other sites, to identify why and what about such programs work, and to assess the impact of nonenforcement-related strategies.

In this article, we assess a Pittsburgh, Pennsylvania–based violence-prevention strategy known as One Vision One Life (or One Vision). In 2003, Pittsburgh had a record-setting 70 homicides, a 49% increase over 2002, with the homicide rate that year increasing from 14 per 100,000 to 22. The homicide rate in Pittsburgh in recent years has been higher than that elsewhere in the nation and, since 2001, than in other cities with 250,000 to 500,000 residents. This increase in violence rallied a coalition of community leaders who formed the Allegheny County Violence Prevention Initiative, which became One Vision One Life. Real increases in certain types of crime, as observed in Pittsburgh, as well as perceptions that a type of crime is "getting out of control," can often lead communities and their leaders to adopt well-meaning but not always well-considered responses. One Vision staff, however, planned their response carefully by examining systematically the nature of violence, considering best practices from other communities across the nation, coordinating with key community partners, communicating with law enforcement, and adopting a strategy they felt was appropriate for responding to the problem and consistent with the goals of the initiative.

Borrowing aspects from several promising evidence-based models, One Vision seeks to prevent violence using a problem-solving, data-driven model to inform how community organizations and outreach teams respond to homicide incidents. It also uses street-level intelligence to intervene in escalating disputes and seeks to place youth in appropriate social programs. One Vision shares information with law-enforcement officials, but it is truly a grassroots effort. Its evaluation has practical and theoretical value.

This assessment of One Vision builds on prior research, and policy makers and scholars should be interested in the findings for several reasons. First, although there is a rich literature evaluating various types of violence-reduction strategies, there have been few quality studies of community-initiated actions that could be thought of as an alternative to strictly an enforcement strategy. Most evaluations have focused on interventions led by the criminal justice community, but the initiative discussed in this article was designed to be representative of evidence-based practices that have been shown to work from public health, social services, and criminal justice disciplines. Second, a critical element of this strategy is to involve non–criminal-justice personnel, usually former gang members, in mediating potential violent conflicts. Although the involvement of "street workers" has been part of other well-known violence reduction strategies like Boston's Lever Pulling initiative, few studies are available and some raise concerns about their effectiveness (see Klein, 1971). We discuss these studies in the context of our results. Third, the intervention is modeled after (but does not mirror) a similar strategy that has been implemented in Chicago, Baltimore, and several other cities. In fact, personnel involved in the Pittsburgh program visited Chicago in late 2004 and early 2005 and attempted to model the intervention and their data collection after CeaseFire in Chicago. Fourth, this type of intervention has been evaluated carefully in Chicago and Baltimore (see Skogan, Hartnett, Bump, and Dubois, 2008; Webster, Vernick, and Mendel, 2009), but an additional evaluation of this type of intervention can yield new lessons about the promise and possible pitfalls of such a strategy. Exploring the program's effectiveness relative to variation in implementation, local dynamics, and community characteristics is helpful for assessing the likelihood that this program could succeed elsewhere. Such lessons would be a useful resource for policy makers, practitioners, communities, and researchers. Finally, the results in this article are not only different that what was observed in the other studies, but it seems that this program led to an increase in violence in the target neighborhoods. We discuss the potential reasons for these increases and the implications for these types of strategies.

Literature Review

In this literature review, we review first the literature relevant to understanding the potential impacts of the model. Specifically, we examine research on problem-solving, street workers, and community outreach initiatives. Second, we review the small number of studies that examined the impacts of programs designed similarly.

Problem Solving, Homicide Incident Reviews, and Collaborative Partnerships

One of the most significant developments for initiating change within criminal justice organizations is the application and adoption of problem-solving approaches. The theory behind the approach has been adopted widely and used successfully in multiagency collaborative partnerships (Dalton, 2003). There are many examples of criminal justice officials systematically collecting data to examine a crime problem more completely, to develop and implement innovative responses, and to assess the impact of these responses. New York City's CompStat program is probably the best-known example of formulating this process into everyday organizational decision making (Silverman, 1999), and the Boston Gun Project is often used as a program that demonstrates the potential of systematic data analysis (Kennedy, 1997, 1998; Kennedy, Piehl, and Braga, 1996; NIJ, 2001). Analyses of the Boston Gun Project found several benefits (see Wellford, Pepper, and Petrie, 2005). Violent gang offending slowed dramatically, and youth homicide in Boston fell by two thirds after the strategy was put into place (Kennedy, 1998: 3). The intervention also led to a 63% decrease in the monthly number of youth homicides, a 25% decrease in assaults with firearms, and a 32% decrease in shots fired. Boston experienced a greater (statistically significant) decrease in youth homicide than did 39 other comparison cities (Braga, Kennedy, Waring, and Piehl, 2001; see also Braga and Pierce, 2005). Minneapolis also experienced sharp reductions in homicide after implementing a similar strategy (Kennedy, 1998; Kennedy and Braga, 1998).

This success led NIJ to support efforts to replicate similar Strategic Approaches to Community Safety Initiatives (SACSIs) in ten other cities, ultimately leading to national deployment of the Project Safe Neighborhoods (PSN) initiative by the Department of Justice (Coleman, Holton, Olson, Robinson, and Stewart, 1999; PSN, n.d.). Recently, the principles of problem-oriented policing generally and PSN have been extended to a drug market initiative (see Corsaro, Brunson, and McGarrell, 2009). Although the deployment of this model elsewhere has not been examined as closely as it was in Boston, there is some evidence of similar promise. For example, the Indianapolis Violence Reduction Partnership helped reduce homicides from 155 in 1997 to 101 in 2000, making Indianapolis the only city among six comparison cities to experience a statistically significant change in homicide frequency (Chermak and McGarrell, 2004; Corsaro and McGarrell, 2009; McGarrell, Chermak, Wilson, and Corsaro, 2006). A national evaluation of ten SACSI sites concluded that, when the SACSI approach is implemented effectively, it "is associated with reduction in targeted violent crime in a community, sometimes as much as 50%" (Roehl et al., 2006: 2). Similar positive results are emerging from select PSN sites that have implemented the problem-solving model (McGarrell, Hipple, and Corsaro, 2007; McDevitt, Braga, and Cronin, 2007; Papachristos, Meares, and Fagan, 2007) and from a national assessment of the PSN initiative (McGarrell, Corsaro, Hipple, and Bynum, 2010).

One intriguing element of the Pittsburgh One Vision approach to violent crime is that, although it is only loosely linked to law enforcement, it embraced the problem-solving model. Concerned officials and community leaders completed a systematic review to improve their understanding of the nature of the problem before acting. They discovered an important and familiar pattern: A small group of chronic offenders in just a few neighborhoods accounted for a large share of all homicides. They also found that young

Black males living in several high-crime neighborhoods were significantly more likely to be homicide victims and that more than 60% of the homicides in Pittsburgh occurred in just four neighborhoods. The homicide rate for Black males living in just a few neighborhoods was 423 per 100,000—more than 50 times the U.S. rate (One Vision One Life, 2005). These neighborhoods became some of the target neighborhoods chosen for a strategic response. Violence data continue to guide the program's intervention strategies, as they did when One Vision expanded its Pittsburgh Southside target area when it became clear that incidents in its original target neighborhood were spilling into adjacent neighborhoods.

Conflict Intervention and Mediation: Street Workers and Street Intelligence

One Vision community coordinators use street-level intelligence to become aware of and then intervene in potentially violent altercations. The coordinators, who are selected because of their familiarity with and connections to the targeted neighborhoods and knowledge about rival groups, are trained in dispute resolution, conflict mediation, and culturally sensitive outreach. They work to prevent violence in three direct ways:

1. They attempt to defuse disputes, such as a petty argument or turf battle, before they escalate.
2. They coordinate public and behind-the-scenes responses to every homicide (and shooting, when awareness of the incident is timely) that occurs in the targeted neighborhoods.
3. They connect individuals and specifically youths to critical services.

Responses to homicides include gathering intelligence about the situation and talking with key actors (e.g., the victim's family, the perpetrator, or others who might be involved in any ongoing dispute) to mediate or minimize the violence and disseminating a general antiviolence message by providing resources, materials, and information to residents.

This is similar to the underexamined role that street workers and community organizations played in contributing to the success of the Boston Gun Project. Boston street workers identified at-risk youth and worked to provide them with critical services, such as job training and substance-abuse counseling. They mediated disputes between rival gangs and worked with law enforcement to prevent violent outbreaks (Braga and Kennedy, 2002). These street workers also worked closely with the Boston TenPoint Coalition—a group of activist Black clergy that also tried to link youths with social services and worked with law enforcement to resolve disputes. Few data exist on the work of street workers and community organizations, which was not measured in any substantive way. This is unfortunate especially given contentions that the TenPoint Coalition was critical to the decreases in youth violence through its creation of an "umbrella of legitimacy," providing balance to the inner-city community and law enforcement that did not exist (Winship and Berrien, 1999). Other cities, such as Indianapolis and Rochester, New York, also have implemented a clergy or street-worker coalition as part of a larger violence-reduction strategy. Yet we have little understanding of whether or how these are effective and how they might be transferred to other cities and programs. Importantly, scholars have identified several potential problems and weaknesses in the delivery of service by street workers. For example, Klein's (1971) important study of programs in Los Angeles highlights potential weaknesses, including lack of supervision, lack of focus, and goal confusion. Moreover, he found that gang workers spent only approximately 20% of their time monitoring gang members, concluding "it may be like squeezing blood out of a turnip to think that an average of five minutes per week per boy could somehow result in a reduction of delinquent behavior" (p. 163). Klein (1971: 151) raised the possibility of street workers contributing to a "paradox of programming," whereby meeting with gang members might actually increase delinquency by increasing the potential cohesiveness of the gang. An evaluation of the Pittsburgh program can expand understanding of the impact of street work in that the program uses primarily former gang and other individuals with criminal justice histories.

Community Mobilization and Outreach

One Vision coordinates broadly and to varying degrees with other community and social service agencies, businesses, and law enforcement. Much of the violence in the areas it targeted stemmed from the illicit drug trade. In its broad approach, it is similar to effective programs that addressed neighborhood drug problems from multiple perspectives with a diverse array of resources and that were connected to broader neighborhood quality-of-life issues (Corsaro et al., 2009; Weingart, Hartmann, and Osborne, 1994). A better understanding is needed of how broader efforts, such as that in Pittsburgh, can harness community capacity to combat both relatively narrow problems, such as the drug trade, and broader problems, such as crime.

Macrolevel variables, such as economic inequality, politics, racism, and demographics, certainly have a greater impact on neighborhood crime, disorder, and quality of life than anything law enforcement or community organizations do (see Duffee, Renauer, Scott, Chermak, and McGarrell, 2006; Skogan, 1990; Spergel, 1976; Wilson, 1987). Yet community organizations or law enforcement can still mediate the impact of these broad social forces on residents (Byrum, 1992; Cortes, 1993; Grogan and Proscio, 2000; Sampson, Raudenbush, and Earls, 1997; Spergel, 1976). As Duffee et al. (2006: 2.7) noted, "[t]here are numerous actions that can be and are taken within neighborhoods and between neighborhoods and outsiders that are an effective component of a larger, more encompassing community improvement strategy." For One Vision, these actions include working in the community to build broad-based sustainable partnerships, significantly increasing the community's commitment to its most troubled neighborhoods, reducing the isolation of the residents living in these neighborhoods, and linking residents to social service organizations as well as organizations to each other.

Research on Similar Initiatives

There have been two other evaluations of programs like One Vision. These evaluations are discussed subsequently. Although the results show generally positive effects for such strategies, it is important to test the effectiveness of the model in other cities with different types of offense and program challenges.

CeaseFire Chicago. As noted previously, the individuals involved in the creation of One Vision were significantly influenced by a program administered by the Chicago Project for Violence Prevention called CeaseFire Chicago (Skogan et al., 2008). CeaseFire Chicago began in 1999 and underwent a rigorous NIJ evaluation, led by Wesley G. Skogan, in 2005. The process evaluation included surveys of staff, interviews with clients and collaborators (e.g., community, clergy, business, police, and school representatives), and observation of meetings. The impact assessment compared changes in violent crime, hot spots, and gang-related changes that occurred in seven CeaseFire sites with those that occurred in other matched areas.

The researchers found that the program contributed to statistically significant decreases in shootings and attempted shootings, the size and intensity of hot spots, gang homicide density, reciprocal killings, and gang homicides in many of the research areas evaluated relative to the comparison sites (Skogan et al., 2008). The researchers examined the impact of the program in seven of 25 program areas, comparing the results with matched areas. Although violence in Chicago was generally down in all areas during the evaluation period, the study indicates that the program pushed key violence indicators down even more. Specifically, shootings and attempted shootings decreased in four of the seven areas between 17% and 24%. An analysis of hot spots in the program areas indicated that six of the seven sites were safer, and "there was evidence that decreases in the size and intensity of shooting hot spots were linked to the introduction of CeaseFire in four of these areas" (Skogan et al., 2008: 8–15). A critical component of the analysis was examining the impact on gang-related activities and homicides. The findings indicate that gang homicide density, reciprocal killings, and gang involvement in homicides decreased in about half of the areas examined.

Baltimore Safe Streets Program. To date, there has only been an interim evaluation of Baltimore Safe Streets (Webster et al., 2009). This program was modeled after Chicago CeaseFire. The analysis focused on differences between attitude changes and program effects on violence in the target areas and a comparison area. The analyses indicated that participants' views on gun violence were much different in one of the target areas. The analysis found, even after controlling for other variables, significantly reduced support for gun violence to settle disputes in McElderry Park but no significant change in Ellwood Park. Controlling for various indicators, the results indicated that being a resident in McElderry reduced support of gun violence to settle disputes.

The reduced support for violence in McElderry Park was coupled with overall positive results for the program there. The area had seen "an average of 0.31 homicides per month

(3.7 per year) during the months prior to the implementation of Safe Streets in August 2007, but no homicides during the 14-month follow-up period," a reduction that was also statistically significant (Webster et al., 2009: 9). There was some diffusion of benefits to surrounding communities, where homicides also decreased. The program also led to a reduction of youth homicides in McElderry Park. The evaluation found no effect of the program in Ellwood Park, but there was an upturn in homicides in Union Square. The evaluation found an association with the program and fewer nonfatal shootings in Ellwood Park but with more such shootings in McElderry Park and Union Square. We discuss the Chicago and Baltimore programs subsequently.

Methods

The main focus of our analysis is to examine what impact, if any, One Vision had on violence in the targeted and surrounding communities. It is important to note, however, that we conducted a comprehensive implementation assessment as well, including field observations, interviews, and police ride alongs. These results are available elsewhere (see Wilson, Chermak, and McGarrell, 2010) but will be referenced in the Conclusions in an attempt to improve our understanding of the nature of the impacts.

Impact Assessment

We examined the impact of One Vision on violence using a quasi-experimental design that compared violence trends in the program's target neighborhoods before and after implementation with (a) trends in Pittsburgh neighborhoods where One Vision was not implemented through a propensity-score analysis and (b) trends in specific nontarget neighborhoods whose violence and neighborhood dynamics One Vision staff contended were most similar to those of target neighborhoods.[1] As part of the outcome analysis, we also explored the extent to which violence or violence-suppression benefits "spill over" into neighborhoods that are adjacent to the target neighborhoods. One Vision's primary goals were to reduce homicide and shootings. Given the data were at the neighborhood level, the outcome models assessed intervention effects by comparing the average outcome for the target neighborhoods with the average outcome for the nontarget neighborhoods. This is a standard way of assessing a difference in difference. Consistent with One Vision's first goal, we drew on existing data to incorporate homicides as an outcome variable. Unfortunately, changes in how Pittsburgh police recorded incidents precluded us from directly measuring

1. The untreated control group design with multiple pretests and posttests (Shadish, Cook, and Campbell, 2002) is a widely used quasi-experimental design that accounts for most threats to internal validity except selection bias or the chance that something "unique" and unobserved about the target or comparison areas influenced levels of violence in them and hence measurements of program effectiveness. Fortunately, propensity-score weighting and our ability to examine the impact of One Vision in multiple target areas with multiple start dates using two sets of comparison neighborhoods help limit selection bias.

progress toward the second goal of reducing shootings. For proxy variables, we gathered data on aggravated assaults and aggravated assaults with a gun. Although these categories of violence include shootings and might indicate program effects, they also include other forms of violent acts and hence are not a precise measure of One Vision's success in reducing shootings.[2]

The Pittsburgh Bureau of Police provided incident-level data for homicides occurring between January 1, 1997 and December 31, 2007, as well as for aggravated assault and gun assaults between January 1, 1996 and December 31, 2007. We aggregated these data into monthly counts for each neighborhood. The Pittsburgh Department of City Planning provided all remaining variables, which were extracted from the 2000 census (Department of City Planning, 2006).

Analyzing the effect of One Vision posed several challenges. Chief among these was that the implementation of the program was not random but based on levels of violence and expert opinion of the areas most suitable for it. This created the possibility that something particular about the neighborhoods chosen, aside from the One Vision program, could account for any change in levels of violence—or, specifically, in homicides, aggravated assaults, and aggravated assaults with a gun—after implementation.

To help control for the possibility of such selection bias, we used the statistical method of propensity scores (Rosenbaum, 2002; Rosenbaum and Rubin, 1983) to find the most appropriate (simulated counterfactual) neighborhoods to compare with the One Vision neighborhoods. For a sensitivity analysis, we used expert opinions in a subsequent analysis to select a second set of counterfactual neighborhoods and compared them with the One Vision neighborhoods. Finally, to assess whether One Vision had an impact beyond the target neighborhoods and into the neighborhoods surrounding them, we conducted a spillover analysis. Next, we summarize our approach to these analyses and present the results of them.

Defining One Vision's Target Neighborhoods

Pittsburgh is made up of 89 officially recognized neighborhoods that vary in size from 39 to 14,507 residents. One Vision was implemented in three target areas, each of which contained multiple neighborhoods. Becoming the target neighborhoods, the Northside included 18 neighborhoods, whereas the Hill District and Southside contained

2. In conducting our impact analyses, we attempted to minimize type I and II errors. To minimize the probability of rejecting a null hypothesis when it is true (type I error), we used a .05 alpha level, a standard benchmark, as the criterion to determine statistical significance. The probability of not rejecting the null hypothesis when it is false (type II error) relates to the ability to detect whether One Vision was associated with some change in the violence measures when it actually was. Such error is a function of sample size. We attempted to minimize it by expanding our sample as much as possible. We compiled longitudinal data on each neighborhood in our analysis. This yielded at least 3,036 observations (and as many as 10,512 observations) for each of our impact models. See the Outcome Models section for an example of how the sample size for each model is calculated.

6 and 8 neighborhoods, respectively. One Vision began operating in its Northside and Hill District target neighborhoods in May 2004; it expanded to eight Southside neighborhoods in May 2005.[3] The differential start dates enabled us to assess the impact of One Vision at two unique intervention points, strengthening the validity of our analysis and reducing the chance that some other unseen variable was the true cause of any program effects.

Designing the Simulated Counterfactual

Comparison Neighborhoods. Assessing the impact of a violence-prevention strategy, or any social program, requires comparing the actual experience of an area where a program was implemented to some benchmark on what likely would have occurred there without it. One of the greatest challenges to gauging a strategy's effectiveness is choosing or designing a comparison or counterfactual that best represents what a target neighborhood would experience without any sort of intervention. Ideally, an intervention would be assigned randomly to a large number of neighborhoods so that the intervention and nonintervention neighborhoods are statistically equivalent, meaning that any preexisting differences would be simply a result of chance. This standard is difficult to attain in field settings. In the case of One Vision, for example, community leaders chose target neighborhoods based on their assessment of which had the greatest propensity for violence and highest likelihood for One Vision to work effectively. So researchers instead select for comparison neighborhoods that are similar or are somehow matched to the target neighborhood on key dimensions related to the outcome variables (in this case, measures of violence). As a quasi-experiment, such a design cannot rule out every threat to validity (i.e., the ability to link outcomes to the intervention). Nevertheless, when conducted properly, quasi-experiments represent the best available option for assessing program effectiveness.

To begin evaluating One Vision's effect on violence, we weighted the 55 nontarget neighborhoods (i.e., all other Pittsburgh neighborhoods not chosen as a target) based on how well they matched the target neighborhoods. These nontarget neighborhoods represented a simulated counterfactual for the target neighborhoods without the intervention. All nontarget neighborhoods were used in the analysis, so we lost no cases in the matching process. Here, we used the method of propensity scores (Rosenbaum, 2002; Rosenbaum and Rubin, 1983) to reduce selection bias. This strategy has been used previously to assess neighborhood effects (Tita and Ridgeway, 2007; Tita et al., 2003). The method of propensity scores can produce causal estimates using observational data by weighting or

3. In May 2004, One Vision also started working in the neighborhoods of Beltzhoover and Saint Clair, which are traditionally considered "Southside" neighborhoods. However, we excluded these from the analysis because, given the different start date from the other Southside neighborhoods, they would need to be modeled independently from the other Southside neighborhoods and with only two neighborhoods the model may have produced unreliable estimates. Given these neighborhoods received One Vision services that could have affected violence, they were also inappropriate to use as counterfactual neighborhoods. We therefore excluded them from our analyses.

TABLE 1

Comparison of Target and Nontarget Neighborhood Characteristics

	Target Neighborhoods		Nontarget Neighborhoods Before Propensity Weighting			Nontarget Neighborhoods After Propensity Weighting		
	Mean	***SD***	**Mean**	***SD***	***p* Value**	**Mean**	***SD***	***p* Value**
Homicide rate in 2003	0.41	0.56	0.61	2.43	.38	0.63	1.59	.90
Aggravated assault rate in 2003	13.95	26.41	9.25	31.18	.01	18.60	46.27	.94
Gun assault rate in 2003	5.18	9.82	2.31	5.37	.01	4.65	7.68	1.00
Population density	8.22	9.26	6.33	4.41	.42	6.35	3.83	.94
% population aged 15–24 years	14.97	9.23	16.60	12.36	.37	16.06	10.22	.57
% no high-school grad	24.90	9.81	20.39	11.77	.04	24.27	9.29	.61
% Black	45.47	35.00	29.13	32.71	.06	47.11	36.39	.92
% professionals	0.25	0.12	0.34	0.16	.00	0.27	0.12	.87
% income < $25,000	53.95	17.76	45.50	14.93	.03	51.49	13.17	.74
% in poverty with child	11.27	13.40	6.34	7.06	.01	9.50	7.75	.77
% public assistance	10.55	9.63	6.76	8.37	.00	9.31	6.68	.60
% vacant housing unit	18.97	14.49	11.97	8.61	.01	17.11	8.89	.76
% moved in 5 years	42.67	13.55	38.19	11.94	.11	39.73	9.88	.74

matching different neighborhoods in a way such that target and nontarget neighborhoods have similar characteristics, thereby reducing selection bias in the process of comparison.[4] The propensity score for a neighborhood is the probability that a neighborhood with a particular set of features is a member of a target neighborhood. We employed a two-step process for estimating the propensity scores. First, we sought to control for as many neighborhood characteristics as possible; yet we were sensitive to our sample size and the available power to detect statistically significant differences. We therefore employed logistic regression using the backward selection method to identify the variables that should be used in the estimation of propensity scores. Initially beginning with 30 socioeconomic-demographic neighborhood characteristics, the selection process identified 13 characteristics useful for calculating propensity scores. These are listed in Table 1. Second, we estimated the propensity score with generalized boosting methods (GBM) using the 13 neighborhood

4. See Apel and Sweeten (2010) for an overview of the propensity-score methodology and its use in criminology.

characteristics potentially correlated with the violence rate in a neighborhood.[5] When fitting this model, the outcome was an indicator of whether a neighborhood was a target neighborhood, and the covariates were the neighborhood characteristics. Table 1 illustrates that after propensity score weighting, no statistically significant differences were found between target and comparison neighborhoods relative to these characteristics. We used the resulting model to predict the probability of intervention assignment for every neighborhood in the sample.[6]

A second way we tested for an impact of One Vision was to compare changes in the outcome variables in the target neighborhoods with a set of neighborhoods One Vision staff advised were most like the target neighborhoods. One Vision staff suggested 17 neighborhoods for this. We used these neighborhoods to create another comparison area, which permitted an additional test of impact that had face validity as determined by local experts.

Spillover Areas. In addition to intervention effects in the target neighborhoods, it is possible that the One Vision program produced displacement effects in nearby neighborhoods. The program might have shifted violence from neighborhoods where outreach and other program activities were focused to surrounding neighborhoods where they were not. Conversely, some researchers (Clarke and Weisburd, 1994; Eck, 1993; Weisburd et al., 2006) contend that interventions might extend crime-suppression benefits. Accounting for such possible "spillover" effects is necessary to gauge the true benefits, or possible drawbacks, of the program.

We analyzed the possible spillover effects for the Hill District and Southside. We did not do so for Northside because it is largely surrounded by the Ohio and Allegheny Rivers, which, local experts contended, largely separate the area from the rest of the city. Our methods for the spillover analysis were similar to those for our counterfactual comparison analyses. We determined the extent of a spillover effect through change in violence in the

5. Following McCaffrey, Ridgeway, and Morral (2004), we used GBM to estimate propensity scores. GBM is a flexible nonparametric approach to modeling $\log(p_i / (1 - p_i))$ that handles a large number of variables in an automated and systematic manner. Ridgeway and McCaffrey (2007) showed that it provides estimated propensity scores that yield better estimates of effects than other approaches do. In particular, GBM automatically selects parameters for inclusion in the model and does not arbitrarily exclude potentially important predictors. It also allows for interaction and nonlinearity in the propensity scores. With p_i estimated for each neighborhood, we used $w_i = 1 / p_i$ as the weight to be used in the Poisson regression model.

6. A common method for selecting comparison neighborhoods among all candidate nontarget neighborhoods involves matching every target neighborhood with the nontarget neighborhoods that have the most similar propensity score. This process eliminates nontarget neighborhoods that are dissimilar to the target neighborhoods. The nontarget neighborhoods matching a target neighborhood are used as simulated counterfactual neighborhoods without the program. In our analysis, we used an improved version of the propensity-score method called doubly robust (Kang and Schafer, 2007; Robins and Rotnitzky, 2001) because it can yield more consistent estimates.

neighborhoods that were each adjacent to the Hill District (6) and Southside (6) relative to all other nontarget neighborhoods at the time One Vision was implemented (43).

Outcome Models

To estimate the outcome models, we employed Poisson regression, which often is used to model information on counts, such as the number of homicides in a neighborhood, where lower bound values are truncated at zero and upper bound values have no limit. Because the neighborhoods differ in size, we modeled for violence rates, or the number of incidents per 100,000 residents. For the outcome Y_{it}, the number of homicides (or aggravated assaults or gun assaults) in a given month or year t in neighborhood i, for example, the probability of observing any specific number of crimes depends on a unique parameter, the mean number λ_{it} of crime, which for this distribution, turns out to be the same as the variance of the distribution. We model the count of incidences using the regression

$$\log\left(\frac{\lambda_{it}}{N_{it}}\right) = \mu_i + \alpha_1 \text{Treat}_{it} + \alpha_2 \text{Post}_{it} + \alpha_3 (\text{Treat}_{it} X \text{Post}_{it}) + \beta\, \text{Month}_{it} + \beta\, \text{Year}_{it} + \beta\, \text{X}_{it,} \mu_i \sim N(\theta, \tau^{2,})$$

where X_{it} represents neighborhood characteristics including the population density per square mile and the proportions of employed residents in a professional occupation, housing units that were vacant, population aged 15 to 24 years old, residents aged 5 years or older who lived elsewhere 5 years previously, households with public assistance income, and households with an annual income less than $25,000. Treat$_i$ represents the treatment of interest, taking a value of 1 for target neighborhoods and 0 for nontarget neighborhoods. With monthly homicide data collected from January 1997 through December 2007 and monthly aggravated assault and gun assault data collected from January 1996 through December 2007 by neighborhood, an indicator (POST$_{it}$) of the crime data before and after implementation also is included as well as an interaction between the treatment and the postimplementation that allows for an estimation of the change in crime between treatment and nontreatment neighborhoods, a difference in difference.

This model controls for a month and year effect to capture trends and serial dependence, as well as a random neighborhood effect μ_i normally distributed with mean θ and standard deviation τ. Because some neighborhoods were more populated than others, we used the population size N_{it} in a neighborhood at time t as an offset. It allowed for the estimation of rate of crime per person. $e^{\alpha 1}$, the exponential of the treatment regression estimate $\alpha 1$, the main effect, is the ratio of the rate of crime between target and nontarget neighborhoods (when the treatment of interest is the One Vision program). Because our interest was in α_3 the interaction effect, which is a straightforward difference of difference in the case of a linear model, and because we used a Poisson regression, we used the method of predictive margins to turn our estimates into expected count of crime per 100,000 persons in a neighborhood.[7] This yields a difference-in-difference equivalence to the interaction effect. From the Poisson regression model, this estimated an average count of crime "hypothetically assuming" that all of the neighborhoods were nontarget neighborhoods and then estimated an average "if hypothetically" One Vision was implemented in all neighborhoods. The difference between those obtained crime counts is equivalent to the main effect, α_1. We did a similar transformation for the pre–post effect as well as the interaction (i.e., the difference in difference).

The number of observations used to estimate each model is a function of the number of neighborhoods in the particular analysis and the number of months for which we have data. For example, the model used to estimate the impact of One Vision by comparing homicides in Northside with those in all other nontarget neighborhoods is 9,636. This is calculated by multiplying the number of neighborhoods in the analysis, 73 (18 target plus 55 nontarget), by the number of months for which we have data, 132 (11 years of 12 months each). We had full data for each neighborhood, so we lost no cases in the analysis. Census-derived socioeconomic-demographic variables were constants and not adjusted or interpolated in any way.

One Vision's Impact

General Violence Trends

Before exploring the empirical impact of One Vision, it may be helpful to review the general trend of violence in the target and comparison neighborhoods. Figures 1–3 show the annual counts of homicides, aggravated assaults, and aggravated assaults with a gun (or gun assault) in the neighborhoods that comprise the three target areas and the nontarget area. The behavior of the trend is much more illustrative than the aggregate level as the number of neighborhoods differs in each area (e.g., the nontarget area has the highest level of violence in each of the figures because it contains many more neighborhoods than the target areas). These illustrate general increases over time. Keeping in mind the frequency is low, homicide levels spiked in 2003 and then temporarily fell in 2004 (Figure 1). At this point, they generally increased in the nontarget and Hill District neighborhoods and fell in the Northside neighborhoods. Figures 2 and 3 highlight that aggravated and gun assaults spiked in the nontarget and Northside neighborhoods in 2002 and in the Hill District and Southside neighborhoods in 2003. In 2006, Hill District aggravated assaults spiked again. By 2007, Northside aggravated assaults spiked, whereas they fell in the Hill District (they remained relatively stable in the nontarget and Southside neighborhoods). From 2004 to 2007, the neighborhoods in each of the areas exhibited different gun assault patterns.

7. Because the Poisson regression coefficients can be interpreted only as the expected increase (or decrease) in the log count of violence per population size as a result of One Vision, we converted the regression estimates into the estimated number of count per 100,000 people using the method of predictive margins (Graubard and Korn, 1999).

FIGURE 1

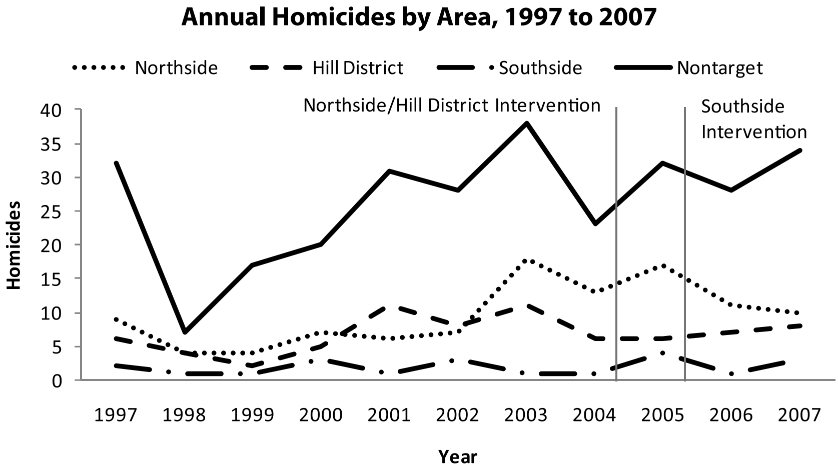
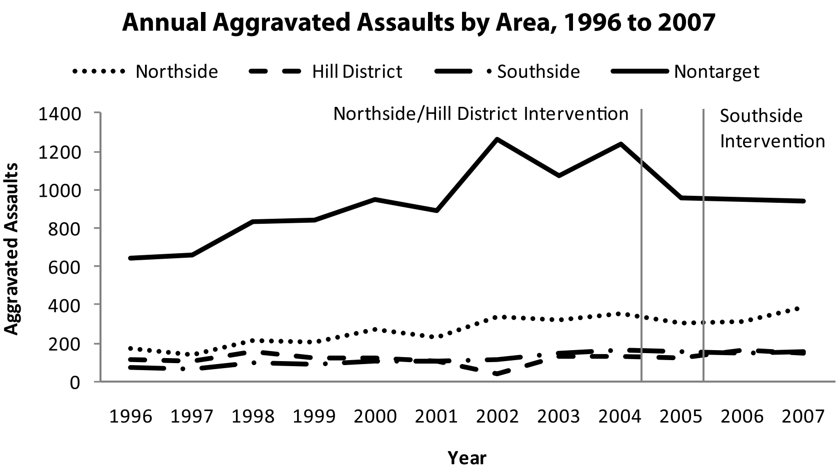


FIGURE 2



The patterns in nontarget neighborhoods substantively declined and leveled off. The Northside patterns spiked in 2005, fell in 2006 and then spiked again to its highest level in 2007. This Hill District patterns substantively increased and the Southside remained relatively stable until both fell in 2007.

Impact Relative to the Propensity-Based Comparison

Although the One Vision initiative was implemented during a time of increasing violence, its effect is best assessed by comparing changes in crime in the target neighborhoods with those in the comparison areas. Such analysis must control statistically for other variables that could

FIGURE 3

Annual Gun Assaults by Area, 1996 to 2007

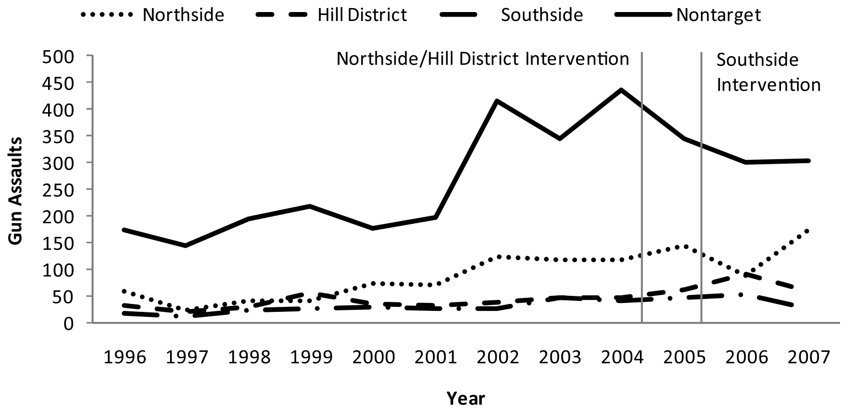


TABLE 2

Test of One Vision Intervention Effects, Propensity Score–Weighted Counterfactual Neighborhoods

Outcome	Predicted Monthly Rate Change (per 100,000 residents)	*p* Value
Northside		
Homicide	0.0219	0.7432
Aggravated assault	25.2095	0.0000
Gun assault	9.2824	0.0000
Hill District		
Homicide	−0.6710	0.3374
Aggravated assault	7.7365	0.0255
Gun assault	5.2893	0.0012
Southside		
Homicide	−0.2540	0.6976
Aggravated assault	25.3953	0.0000
Gun assault	4.9865	0.0015

explain changes in violence, including the time period of observation and neighborhood conditions.

Table 2 summarizes the outcomes of the models used to assess the impact of One Vision on violence in all the target neighborhoods compared with nontarget neighborhoods.[8]

8. As noted in the Outcome Models section, these models control for seven neighborhood characteristics. To preserve space, we do not provide the full results of the outcome models. However, they are all available upon request.

TABLE 3

Test of One Vision Intervention Effects Relative to Comparison Neighborhoods Suggested by One Vision Staff

Outcome	Predicted Monthly Rate Change (per 100,000 residents)	*p* Value
Northside		
Homicide	0.2845	0.7588
Aggravated assault	26.7970	0.0000
Gun assault	14.6100	0.0000
Hill District		
Homicide	–0.9174	0.2681
Aggravated assault	6.4579	0.1922
Gun assault	9.4336	0.0016
Southside		
Homicide	–0.6288	0.7438
Aggravated assault	25.0327	0.0000
Gun assault	4.8154	0.0057

We weighted the nontarget neighborhoods by propensity scores on how closely they matched the target neighborhoods. As noted previously, the sample sizes, calculated by multiplying the number of neighborhoods in the particular analysis by the number of months for which we had data, varied from 8,316 to 10,512. The results show that One Vision was not associated with any change in homicide rates relative to all Pittsburgh neighborhoods not served by One Vision. They show aggravated assault and gun assault rates increased in the target neighborhoods relative to the comparison neighborhoods after program implementation. The table presents effects in predicted change in a monthly rate of occurrence per 100,000 residents. These data suggest that the rates of aggravated assault increased similar amounts in the Northside and Southside (approximately 25 per month) but by a smaller rate in the Hill District (approximately 8 per month). Gun assault monthly rates increased more in the Northside (approximately 9 per month) than in the Southside and the Hill District (approximately 5 per month).

Impact Relative to the One Vision-Suggested Comparison

As a second way to assess the impact of One Vision, we examined changes in violence in the target neighborhoods compared with neighborhoods that One Vision staff suggested were most similar based on their intimate familiarity with the neighborhoods. Table 3 highlights these results, again controlling for the time period and differing neighborhood characteristics. The number of observations for each of these models ranged from 3,036 to 5,040. With one exception, the assessment of the program's impact on violence was essentially the same as shown previously. This analysis showed that One Vision did not

TABLE 4

Test of Spillover Effects, Propensity Score–Weighted Counterfactual Neighborhoods

Outcome	Predicted Monthly Rate Change (per 100,000 residents)	*p* Value
Hill District		
Homicide	−0.5546	0.6483
Aggravated assault	−14.2040	0.0379
Gun assault	6.1647	0.0979
Southside		
Homicide	−0.8695	0.8012
Aggravated assault	28.7132	0.0000
Gun assault	5.5715	0.0072

have an effect on homicide rates. It showed One Vision was associated with increases in the monthly rate of aggravated assaults in the Northside (approximately 27) and Southside (approximately 25) areas but was statistically unrelated to changes in the rate of aggravated assaults in the Hill District. This comparison with areas suggested by One Vision staff also showed increased gun assault rates in the target neighborhoods areas relative to those not targeted.

The Impact of One Vision on Violence in Adjacent Neighborhoods

To account for potential spillover effects of One Vision's implementation, either displaced-violence or violence-suppression benefits, we used impact analyses to assess change in violence in the neighborhoods adjacent to the Hill District and Southside relative to all the remaining nontarget neighborhoods in the city (matched to the spillover neighborhoods by propensity scores). As Table 4 shows, the models detected no spillover effects for homicide as a result of One Vision's implementation (the sample sizes of the models varied from 6,486 to 7,050). By contrast, the table shows One Vision was associated with spillover effects in aggravated and gun assaults. After One Vision was introduced, neighborhoods adjacent to the Hill District saw a reduction in aggravated assaults but no statistically significant change in gun assaults relative to other comparison neighborhoods. The neighborhoods surrounding Southside experienced increases in both aggravated and gun assaults. The suppression benefit to the neighborhoods contiguous to the Hill District was approximately 14 aggravated assaults per 100,000 residents per month. The increased rate of this offense in the neighborhoods next to Southside was nearly 29 per 100,000 residents. The detrimental spillover effect of One Vision on gun assault rates per month was approximately 6 incidents per 100,000 residents in the neighborhoods adjacent to the Southside.

Discussion

The Overall Impact of One Vision

Using two forms of comparison, each of which controlled for neighborhood attributes, seasonal effects, and trends over time, we found no quantitative evidence One Vision helped reduce violence. We found no effect of the program on homicide rates. We did find that the onset of One Vision efforts was associated with increases in aggravated assaults and gun assaults in all three target areas (excepting the comparison of aggravated assaults in the Hill District and the comparison area suggested by One Vision staff).

Our spillover analyses also indicated that the introduction of One Vision was associated with no change in homicide rates. We did find introduction of One Vision associated with an increase in aggravated assaults in the Southside spillover neighborhoods and a decrease in such assaults in the Hill District spillover neighborhoods. We found that the program was associated with increases in gun assaults in the Southside spillover neighborhoods but had no effect in the Hill District.

It is a challenge to explain why a program did not produce any effect, but it is an even greater test to discuss why a program had a negative effect. Before attempting to explain the negative effects, we contrast features of the Pittsburgh program with similar ones in Chicago (Skogan et al., 2008) and Baltimore (Webster et al., 2009). We also contrast the One Vision street-worker program with the original Boston Gun Project that involved street workers as part of a broader violence-reduction strategy. Considering the different findings, we think such comparisons are critical to help policy makers think through the implications of adopting such strategies and identify key implementation strategies. We also think it helps set up the need for additional research.

Comparing One Vision with Other Initiatives

CeaseFire Chicago. As noted previously, the individuals involved in the creation of One Vision were influenced significantly by a program administered by the Chicago Project for Violence Prevention called CeaseFire Chicago (Skogan et al., 2008). The design of CeaseFire Chicago reflected research documenting the success of various public health strategies. The goals of this program include disrupting the cycle of violence and changing attitudes and norms about specific behaviors. The program invested considerable resources in communicating, particularly to high-risk individuals, the costs of being involved in violence; in connecting individuals to services that might provide an alternative to violence; and in directly confronting individuals (usually gang members) who might resort to violence to resolve a conflict. CeaseFire used various community mobilization, education, and mentoring strategies to communicate the dangers of violence. A critical aspect of the program provided "on-the-spot" alternatives to violence and intervention before a conflict escalated in violence. The program also sought to influence perceptions about the risks and costs of involvement in violence.

Several key individuals and groups were critical to implementation of the Chicago program. First, the program employed outreach workers in each targeted community. Each outreach worker had a caseload of approximately 15 clients identified and assessed as being in need. These workers lived in or knew the neighborhoods where they worked and thus had street credibility and a good sense of individuals who were in need. Outreach workers worked the streets by talking with individuals, identifying clients, and then counseling and connecting these clients to needed services. It seems that working with clients was their primary task, but they also were expected to distribute information about the program and its "stop the violence" message to groups and individuals. Outreach workers mediated conflicts as well. Skogan et al. (2008) concluded that the outreach workers succeeded at identifying and working with high-risk clients. In fact, interviews with the clients indicated that, "after their parents, their outreach worker was typically rated the most important adult in their lives" (Skogan et al., 2008: 8–10). Nevertheless, it is difficult to assess the comparable levels of risk clients had in Chicago, Pittsburgh, and other cities with similar programs.

Second, the program employed another group of street-savvy individuals that focused specifically on conflict mediation. These individuals, who are called violence interrupters, were former gang members, who had street credibility because of their past. They were expected to use their understanding of the individuals and groups living in a neighborhood to prevent violence. The violence interrupters identified brewing conflicts or reacted to shootings that occurred and would gather intelligence about these conflicts and then attempt to mediate nonviolent solutions. They talked with gang members, as well as friends and families of gang members and shooting victims, focusing "on affecting risky activities by a small number of carefully selected members of the community, those with a high chance of either 'being shot or being a shooter' in the immediate future" (Skogan et al., 2008: ES-1). A significant amount of their time focused on responding to retaliatory shootings. Skogan et al. (2008: 8–11) estimated that "40 percent of intervener's mediation efforts concerned potential shootings that would have been in retaliation for an earlier imbroglio."

Third, other key contributors to CeaseFire Chicago included community members, social service organizations, and clergy. The program attempted to build and enhance community partnerships. These partnerships were valuable for many reasons, including the access to jobs and services they offered to clients and the legitimacy partners gave the program and its antiviolence message.

Fourth, police and prosecutors were frequent collaborators with CeaseFire Chicago staff. The role of police and criminal justice partners in changing the perceived risk of illegal gun carrying was a formal part of the Chicago CeaseFire logic model. Additionally, police shared information with the program after an incident so that staff could calculate a response. Police also collaborated with them for marches and vigils, walking with program staff and assisting with traffic and crowd control.

Baltimore Safe Streets Program. The Baltimore Safe Streets program was implemented in three high-crime neighborhoods—McElderry Park (East Baltimore), Union Square

(South Baltimore), and Ellwood Park (East Baltimore)—in mid-2007 and early 2008. Safe Streets, like the programs in Chicago and Pittsburgh, attempted to decrease violence by communicating to residents and high-risk individuals the impacts of violence on their communities; reaching out to persons in need, especially high-risk youth; and identifying and then intervening in potentially violent conflicts.

The interim evaluation focuses on the first 14 months of program implementation. It discusses implementation of the program and its effects on attitudes toward gun violence as well as on the number of homicides and shootings. Two different community groups implemented the program model in Baltimore. The implementation in Union Square was abbreviated because problems caused the program to cease after 5 months. Each implementing group was to collect data on the ratio of outreach workers to clients and the number of face-to-face contacts with clients, referrals for services, mediations of disputes, flyers distributed, and violence responses initiated.

The results indicate that the number of clients and face-to-face contacts increased as expected after the implementation of the program. Outreach workers made 450 face-to-face contacts in McElderry Park and just fewer than 100 contacts in Ellwood Park in August 2008. Outreach workers also made a large number of referrals to various services, an average of 26 per month. Most referrals were for employment issues. There was "considerable month-to-month variation" in the number of conflicts mediated (Webster et al., 2009: 6). Between August 2007 and August 2008 in one target area, the number of mediations ranged from six to eight in some months to less than two in others. There were no statistically significant changes in Ellwood Park, but the authors found that there was not a single homicide in the McElderry Park neighborhoods for at least 17 consecutive months (p. 14), but nonfatal shootings decreased less here than in the comparison areas.

However, it also is important to note that although these researchers note that there were problems with implementation in the Union Square neighborhood that resulted in the program being eliminated, the 5 months of activity here was associated with an increase in homicides and shootings in the targeted neighborhoods relative to the comparison areas. The authors discuss the problems with the implementation but unfortunately do not attempt to explain why there might have been increases in shootings or homicides and if the program might have contributed to the increases.

One Vision Versus Chicago and Baltimore. Although the amount of information on the Baltimore program is somewhat limited given that it has only an interim evaluation, several noteworthy differences exist between One Vision on the one hand and the Chicago and Baltimore programs on the other that highlight the difficulties in evaluating programs that are on paper very similar but in practice are quite different. First, although it is difficult to detect dosage of such programs, the organization documents we reviewed for the implementation assessment point to some limitations in the administration of the program model in Pittsburgh.

Specifically, it does not seem that One Vision used the documentation of activities in any systematic way to select actions for the targeted neighborhoods or to monitor the performance of the community coordinators. In contrast, the Chicago program in particular seemed to rely on the information of these documents as an accountability mechanism. The Chicago Project for Violence Prevention (CPVP) essentially supported local organizations to administer the model and then monitored the activities and coordinated with these local programs. Completed forms were a key source of accountability in Chicago. Moreover, Skogan et al. (2008: 2–25) reported:

> During the evaluation period we saw a tightening of policies and procedures on the part of CPVP that reflected the adoption of a more centralized management role. CPVP took a more active role in regulating program activities and reviewing site records. CPVP staff made an increasing number of site visits to ensure better program implementation, and new central office positions were created to handle program implementation and documentation issues. Sites were held more accountable to meeting standards regarding shooting responses, client caseload size, and other program activities. CPVP also became more assertive about the hours that sites were to be open, to parallel the hours when violent crime actually occurs.

Second, both the observations and the organizational documents to some extent reveal that the street workers were involved in a variety of important activities and worked to help people in dire need. Nevertheless, the clients with whom the Pittsburgh community coordinators worked and the types of conflicts mediated seemed to be different from those in Chicago or Baltimore. Specifically, Baltimore and Chicago workers focused almost exclusively on the activities of and conflicts between high-risk violent individuals. Indeed, in Chicago, the clients of CeaseFire workers had extensive criminal histories, which were consistent with those most at risk for being involved in homicides as both victims and offenders (Skogan et al., 2008: ES-10). Similarly, in Baltimore:

> [O]utreach workers logged hundreds of contacts with these high-risk individuals during which they encouraged alternatives to violence, mediated conflicts, provided informal mentoring, and made referrals for services that could decrease risks. The outreach workers interfaced with dangerous gangs with access to guns that operated under circumstances where the odds of lethal altercations are alarmingly high. (Webster et al., 2009: 14)

In McElderry Park, a site that did not have any homicides during the evaluation period, outreach workers intervened in 53 high-stakes disputes and altercations. In Chicago, violence interrupters estimated that 40% of the conflicts mediated could have resulted in retaliation shootings (Skogan et al., 2008). In contrast, few of the conflicts mediated in

Pittsburgh were specifically directed at retaliations. In our field research, we found that One Vision staff, especially the executive staff, attempted to assist shooting victims and discourage retaliations, but that the street workers were not working potential violent conflicts. Data indicated that only 1.8% of the conflict mediations were in response to a potential retaliatory event. We found that street workers did interact with gang members, but they did not necessarily intervene in gang conflicts. Many simply focused on protecting specific gang-affiliated individuals. The implementation data also indicated that the street workers responded to conflicts unsystematically. They typically mediated a conflict when coming into contact with involved individuals in the regular course of their day. They rarely focused on systematically identifying key violence threats and developing responses to them.

Third, Pittsburgh street workers had a variety of responsibilities that made it difficult to manage their workload. They were the heart and soul for program implementation, expected to intervene and mediate conflicts, assist clients, attend violence responses, and participate in community programming. Each of these tasks required different skills and training. As a result, many street workers might have emphasized what they enjoyed doing and those things at which they were most effective and ignored other responsibilities. The Chicago model, in which outreach workers focus primarily on working with clients and mentoring individuals and violence interrupters focus on responding to gang conflicts and responding to shootings, has much more potential for allowing workers to specialize and perhaps become more effective with specific tasks.

Fourth, one difficult challenge of quasi-experiments is the inability to control for other variables that might have contributed to program outcome. Communities with high rates of violent crime might have multiple simultaneous programs and strategies. In the McElderry Park area of Baltimore, there were other law-enforcement initiatives, including "close monitoring of individuals with histories of gun offending, increased police presence in areas with the highest numbers of shootings, and efforts to suppress illegal gun possession and sales" (Webster et al., 2009: 15). Similarly, there were several other initiatives, such as PSN, occurring in Chicago at the same time as CeaseFire.

One Vision Versus Comprehensive "Pulling Levers" Programs

In the original Boston Gun Project, street workers were part of a broader antiviolence strategy that was driven by a multiagency criminal justice task force. The overall mission, reducing homicide and gun violence, was consistent with One Vision, but the tactics included a comprehensive effort to change the perceived risk of groups of chronic offenders from both violent victimization and incarceration. Like those in One Vision, street workers sought to convince at-risk individuals not to carry guns and to avoid conflict and retaliation. Unlike those in One Vision, Boston street workers were backed by direct communication from police, district attorneys, and federal prosecutors on the consequences for illegal gun possession and use. This strategy, which is known as *pulling levers*, had a significant impact on homicide and gun violence in Boston (Braga et al., 2001; NIJ, 2001; but also see Berk, 2005; Rosenfeld, Fornango, and Baumer, 2005a, 2005b; Weisberg, 2005) and in cities that have attempted to replicate the Boston model. These include Indianapolis, Indiana; Lowell, Massachusetts; Stockton, California; and Los Angeles, California (Braga, 2008; McDevitt et al., 2007; McGarrell et al., 2006; Tita et al., 2003). They also include Chicago, where the pulling-levers strategy was a key aspect of a PSN program, which led to a 37% reduction in homicide (Papachristos, Meares, and Fagan, 2007).

In evaluations of complex interventions, it is difficult to identify what elements are critical to success or failure. Like the One Vision strategy, pulling levers as implemented in Boston and other locations consisted of many different elements, making it difficult to identify the elements that produced changes in violent offending behaviors. There were multiple parts of the strategy, including a communication campaign, the work of ministers, home visits, and other police strategies.

Both Chicago and Baltimore had active PSN programs at the time of their street-worker programs. These included efforts to communicate a message aimed at felons against carrying firearms and to increase the federal prosecution of felons possessing or using firearms. Although there was also a PSN program in Pittsburgh, there is no evidence of coordination among the police, the PSN task force, and One Vision. We do not have evidence of such coordination in Baltimore or Chicago. Nevertheless, the fact that Chicago CeaseFire was occurring during a time when Chicago's PSN initiative was holding face-to-face offender notification meetings with high-risk individuals, albeit in targeted neighborhoods, might indicate that the street-worker intervention is more powerful when supported by the credible threat of prosecution for illegal gun carrying and use. It is interesting to note that PSN offender-notification meetings were also occurring in Baltimore during its street-worker program, although we do not have evidence of coordination between PSN and the program.

Assessing the Negative Effects

The problems in implementation in Pittsburgh and some of the differences between One Vision and the strategies in Chicago and Baltimore might help explain the null effects we found in regard to homicide but would not be consistent with the data that showed increases in aggravated and gun assaults. An important question to consider is whether it is plausible that the program contributed to these increases. Importantly, the increases that were uncovered in Pittsburgh are not completely in contradiction to the findings in Baltimore. That is, researchers in Baltimore found significant increases in homicides and shootings in one of the target neighborhoods, but they did not explain why this might have occurred in the Union Square neighborhood. The potential that a program such as what was examined here contributed to increases in violence is a serious concern and worth exploring, and the variation in results that were found when comparing Pittsburgh to Chicago and within Baltimore point to a critical need for more research on this topic. Subsequently, we provide some thoughts about what might be occurring that account for these differences.

Malcolm Klein's research is important here. Klein (1971, 1995) discussed many critical issues that help us better understand street gangs, but his ideas on the "centrality of cohesiveness" were particularly important to thinking through reasons for increased levels of violence after an intervention. In an effort to understand a program that was introduced in Los Angeles in the 1960s called the Group Guidance Project, Klein evaluated closely the activities of gang workers. Similar to what the community coordinators were asked to do in Pittsburgh, these gang workers tried to assist gang members by organizing group activities, assisting them with building skills, and advocating for them as they interacted with criminal justice and social service bureaucracies. What he discovered was that the introduction of these street workers actually increased delinquency and isolated the youth from the community. Importantly, he found when two of the most aggressive, antipolice gang workers took on different responsibilities, the gangs they had been working with essentially disintegrated. Klein wrote (1995: 45), "The original two workers had inadvertently become the focus of the gangs' cohesion. Their active group programming, their antipolice attitudes, their total commitment to the groups had become even stronger glue than the members' original need to come together for identity, status, and belonging." He found also that in the areas where the gang workers were retained, cohesiveness continued to increase.

In Pittsburgh, like most other cities, gangs are generally not very cohesive entities, and thus street workers might increase cohesiveness and, therefore, levels of violence between and among them. It is thus possible that we might be observing a "paradox of programming" (see Klein, 1971)—the presence of outreach workers increased the cohesion of gangs, making some groups more organized, in turn leading to increased violence. Comparisons of programs like those implemented in Pittsburgh, Chicago, and Baltimore might vary the nature and type of gang structures that exist in a particular city or even with neighborhoods within that city. For example, the gang networks in Chicago and Baltimore might be very stable, making it straightforward to identify and mediate conflicts. There are many important studies that explore the evolving gang structure in Chicago, which might best be described as a "chronic gang city" (Tita and Ridgeway, 2007: 233; see also Venkatesh, 1997). The gang structure in Pittsburgh, by contrast, consists of loose conglomerations of groups and would be better described as an emerging gang city (Tita and Ridgeway, 2007). It is possible that the intervention might have brought more people into the gangs by providing opportunities for more individuals to be exposed to gangs, and more clearly defining conflicts as group-based threats.

The process evaluation related to this project revealed several important things about the nature of the "street work" done by the community workers. First, the community coordinators we observed were working to provide activities for the youth in the targeted neighborhoods. They focused on creating activities (e.g., summer basketball leagues, cookouts, etc.) and client-centered outreach. Such activities were deemed to be important to providing alternatives to youths at risk, but they might serve as central meeting places that resulted in the development and enhancement of social networks. Second, although an analysis of the organizational documents revealed that the street workers had contact with gang members, the results showed that their contacts with known gang members was actually lower than expected. Approximately 41% of the conflicts that were targeted by street workers involved gang activities. This means that they were working with many youth who were not known gang members but were bringing them into programming that was designed to assist gang members and provided opportunities to make connections with others. Third, one of the critical ways that Pittsburgh was different from Chicago was in the involvement of the police. In Chicago CeaseFire, the police and related criminal justice partners were an explicit component of the logic model. Specifically, the police and criminal justice system were considered key components in changing the perceived risk and costs for illegal gun possession and use (Skogan et al., 2008). Pittsburgh police were certainly not absent in the targeted neighborhoods during the study period. Nevertheless, there was not the type of coordination between One Vision and the police that Chicago enjoyed. In addition, our ethnographer observed outright hostility between the community coordinators and police in Pittsburgh. Although there seemed to be a good relationship between the executive staff of One Vision and the police, this did not translate to what was occurring on the street. Thus, as community coordinators were in contact with youth and active gang members, they might have pushed them toward being isolated even more from the community as Klein (1971, 1995) observed in Los Angeles.

Our results are tentative as there are other plausible explanations for the increase. For example, preceding conditions in Pittsburgh might have contributed to the effects we saw in some of the violent-crime measures. Other unique social conditions also might have contributed to these results. Violence was increasing prior to the implementation of the program in the targeted neighborhoods. It is possible that One Vision simply did not work as intended (to reduce crime and violence), and the target neighborhoods did in fact realize a marked increase, whereas the comparison neighborhoods did not (or at least less of an increase). The analysis might then suggest what we found, although there were some other neighborhoods that One Vision staff thought would be problematic as well, and they were comparison neighborhoods. But one has to be concerned about the iatrogenic effects we discovered. What is interesting is that in Baltimore there was also variation in effects by neighborhood, and there seemed to be some effects in one of the targeted neighborhoods, no effect in the other, and negative effects in a third area. Chicago, however, produced positive effects in all neighborhoods evaluated: Shootings and gang-related homicides decreased in all of the targeted neighborhoods. Although we find no effects on homicide, we do find that aggravated and gun assaults went up in the targeted neighborhoods and in part also increased in the spillover areas. An important question for future research is why such different results? These contradictory results call out for a need for additional research especially in the area of understanding issues related to cohesiveness and change after the implementation of such programs.

The logic behind using street-savvy individuals to respond to and manage potentially violent conflicts—from identifying to understanding to searching for solutions to them—may be appealing as individuals can use their street credibility to a positive way and help organizations, including the police, monitor brewing conflicts and beefs more effectively. Because the escalation of a street conflict or "beef" to an act of violence may take some time, there is an opportunity for prevention with better intelligence. It is clear that being close to the street and "in the game" is required to obtain good intelligence, but being so forces the street worker to walk a fine line. How much standing should one have? Too little and the worker might be ineffective and in danger. Too much and the worker might become corrupt. This points to the importance of talk and street gossip and its impact on conflicts and ultimately violence—a topic that has received very little attention (but see Lauger, 2010). It is important to think through how a street worker, when even attempting to do right in responding to and addressing beefs, might cause what is a complete but inadvertent effect. The workers used in Pittsburgh, as in other places, were chosen because they had street credibility and in theory could use their status as a starting point in communicating with rival gang members and others in a neighborhood. Their position as a worker might even legitimize their status and their street credibility. They seem to have connections and thus can help individuals in many different ways. If they are confronted with a conflict and chose to intervene, it is plausible that the connections that they have and the communications they make could deescalate the beef. However, it is equally plausible, unless the gang worker is adequately trained and prepared, that their communications might escalate the conflict. The gang worker becomes a critical node of communication across gangs, and what they say, how they say it, and what they ignore all can impact how rival gang members react. The result could certainly increase the hostility and violence between groups.

Policy Implications

Innovative programs are critical to addressing the major issues facing our most disadvantaged cities. Anderson's (1999) important work describing the code of the streets demonstrates how individuals living in these neighborhoods are affected by their environment. The people who live in these neighborhoods adapt to their environment in different ways. The code of the streets becomes a guide to living their lives (see Stewart and Simons, 2009)—adopting a lifestyle that, for many, includes violent criminal activities. There have been many attempts to inject programs into these communities. One Vision represents one of the strategies implemented in Pittsburgh to address concerns about violence. Yet, our evaluation found the onset of the program to be associated with increases in violence in these Pittsburgh neighborhoods. In this section, we discuss several issues related to the limitations of implementing such programs.

First, the implementation of the One Vision program deviated in several ways from ideal implementation. One Vision lacked consistent documentation; the completion of documentation was sporadic and varied by areas. One Vision staff seemed to rarely use the documentation in any systematic way to guide program actions. Street workers focused more on persons in need than on those at risk. This contributed to street workers having a broad variety of tasks and workloads that were difficult to manage. Finally, program actions were neither as frequent nor as focused on gangs and drugs as had been expected. In particular, it does not seem that One Vision routinely focused on the most serious offenders and highest risk individuals.

Second, the program did not intervene with the group or gang structure generating violence. It seems that Chicago CeaseFire, likely reflecting the prevalence of gangs in Chicago, focused on gangs explicitly. The original Boston Gun Project and the successive programs in Indianapolis, Lowell, High Point, and Stockton included a group accountability component. Gangs, cliques, or groups of chronic offenders were told that they would be held accountable for the continued violence of any of their members. As evident in other programs, this form of intervention calls for a greater law-enforcement component.

A related but alternative explanation is that the gang structure in Pittsburgh might require a different approach. Gangs in other cities where similar initiatives apparently have succeeded have more stable and persistent structures than are evident in Pittsburgh. Pittsburgh gangs seem to be far less cohesive, perhaps making it more difficult to identify and mediate conflicts among them. Outside Los Angeles and Chicago, such a fluid gang structure seems to be the norm (NIJ, 2002; Weisel, 2002).

Among the key components of the Chicago and Baltimore programs are the following:

1. Change the norms about the acceptability of violence
2. Increase the perceived costs of involvement in behaviors associated with violence (e.g., illegal gun carrying)
3. Increase the perceived legitimacy and fairness of antiviolence interventions
4. Hold groups of offenders accountable for continued violence
5. Increase linkages to a variety of social supports and legitimate opportunities ("widen decision alternatives"; Skogan et al., 2008: ES-2)

The questions raised about One Vision relate to target populations, dosage, and comprehensiveness. One Vision emphasized the first, third, and fifth components listed previously. It is not clear whether its work with the highest risk groups was intense enough to help reduce overall violence. That is, although One Vision might have had some success in working with individuals in the target areas, these successes might not have been on a scale sufficient to change the levels of violence as measured in this evaluation. One Vision did not partner with local police and prosecutors to communicate a consistent and credible deterrent message that might have changed the perceived risk associated with illegal gun carrying and use, nor did it explicitly focus on influencing social networks of at-risk individuals. The lack of a systematic and integrated law-enforcement component to complement One Vision's activities might, in part, explain its inability to demonstrate a measurable reduction in violence.

One also cannot help but to wonder to what extent community conditions matter in the selection of target areas. The three chosen and studied in this evaluation were thought by community leaders to have significant violence problems and attributes conducive to the activities One Vision sought to implement. However, our examination of the violence data suggested, for example, that the frequency of homicide in the Southside was substantially less than in the Northside and the Hill District. Although it raises the question as to whether One Vision could have had a measurable impact on violence in the Southside because of its amount of observed violence, the answer is obscured given that we did not detect violence reductions in the target neighborhoods with more per capita violence.

One Vision was established to address the serious problem of lethal violence in particular neighborhoods of Pittsburgh. The program leaders looked to CeaseFire Chicago to follow a "promising practice" model for implementation in Pittsburgh. The program staff was trained in the CeaseFire approach. The finding that One Vision did not have an impact on violence in the target neighborhoods raises several critical issues for a field attempting to move toward evidence-based practices. Are the CeaseFire Chicago results stable over time? Are they transferable to other communities that differ from Chicago in gang structure or parallel systems (such as community policing) coordinating with CeaseFire? If the results are stable and not unique to Chicago, then what was missing in Pittsburgh?

We speculated on some of the potential differences in the One Vision program; yet these are post hoc observations. The results from the Baltimore evaluation will be important in addressing these questions. The results from Chicago and some of the results from Baltimore suggest the promise of street-worker programs. The results from Pittsburgh and one of the target neighborhoods in Baltimore suggest the need for continued rigorous evaluation. Taken together, there seems to be enough promise for continued programmatic experimentation but also enough questions that future programs should be coupled with continued evaluation. One critical area to study is the impact of such programs on cohesiveness and thorough analysis of the changing social networks after the implementation of such programs is warranted. This research is needed to assess the efficacy of this type of program in reducing community violence as well as to identify program components associated with violence reduction.

Study Limitations

All studies have limitations that should be considered in interpreting their findings. Evaluations of the sort used in this article face difficulties in identifying best comparison areas, measuring program delivery and performance, and isolating program effects from other effects. True random designs are generally not possible for such social programs. Quasi-experimental designs can approach the rigor of random selection and experimental analysis. Nevertheless, they cannot control for some variables, such as other ongoing initiatives or community changes that might contribute to program outcome. It is possible that the rise in violence we observed was caused by some other change in the target communities that we could not identify and separate from the assessment of program effects.

Similar to design challenges, there are several measurement limitations. First, as noted previously, One Vision's main focus has been on reducing homicide and shootings in its target areas. Although we had data on homicides, changes in the Pittsburgh Bureau of Police's reporting policies precluded us from gathering and assessing longitudinal shooting data. As a consequence, we analyzed the broader categories of aggravated assaults and aggravated assaults with a gun. Although it is possible that these measures could detect changes in shootings, they include other forms of violence whose changing levels might mask program effects on shootings. Second, our data did not permit us to assess gang and group violence and how One Vision's efforts have affected it. Third, homicide is a rare occurrence. Detecting measurable changes in variables with low frequency and variation is generally difficult. Further distinguishing these offenses to examine only those that are gang or group related would make an analysis even more problematic. Finally, our control measures are not as precise as we would like. Necessarily, we drew on U.S. Census Bureau data for socioeconomic and demographic data of the neighborhoods in our analysis. These data illustrate variation among the neighborhoods in 2000 but do not identify changes in them since then.

References

Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York: W. W. Norton.

Apel, Robert J. and Gary Sweeten. 2010. Propensity score matching in criminology and criminal justice. In (Alexis Russell Piquero and David Weisburd, eds.), *Handbook of Quantitative Criminology*. New York: Springer.

Berk, Richard. 2005. Knowing when to fold 'em: An essay on evaluating the impact of *Ceasefire, Compstat,* and *Exile*. *Criminology & Public Policy*, 4: 451–465.

Braga, Anthony A. 2008. Pulling levers focused deterrence strategies and the prevention of gun homicide. *Journal of Criminal Justice*, 36: 332–343.

Braga, Anthony A. and David M. Kennedy. 2002. Reducing gang violence in Boston. In National Institute of Justice, *Responding to Gangs: Evaluation and Research.* Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, NCJ 190351. Retrieved April 22, 2010 from purl.access.gpo.gov/GPO/LPS37771.

Braga, Anthony A., David M. Kennedy, Elin J. Waring, and Anne Morrison Piehl. 2001. Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's operation ceasefire. *Journal of Research in Crime and Delinquency*, 38: 195–225.

Braga, Anthony A. and Glenn L. Pierce. 2005. Disrupting illegal firearms markets in Boston: The effects of Operation Ceasefire on the supply of new handguns to criminals. *Criminology & Public Policy*, 4: 717–748.

Bureau of Justice Statistics. 2007. Table 2: Criminal Victimization, Numbers and Rates, 2006. In (Shannan Catalano and Michael Rand, eds.), *Criminal Victimization, 2006*. Washington, D.C.: Author, NCJ 219413. Retrieved April 22, 2010 from bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=765.

Byrum, Oliver E. 1992. *Old Problems in New Times: Urban Strategies for the 1990s*, 2nd Edition. Chicago, IL: Planners Press.

Centers for Disease Control and Prevention. 2010. Welcome to WISQARS™. In *Injury Prevention and Control: Data and Statistics (WISQARS™)*. Retrieved August 28, 2010 from cdc.gov/injury/wisqars/index.html.

Chermak, Steven and Edmund McGarrell. 2004. Problem-solving approaches to homicide: An evaluation of the Indianapolis violence reduction partnership. *Criminal Justice Policy Review*, 15: 161–192.

Clarke, Ronald V. and David Weisburd. 1994. Diffusion of crime control benefits: Observations on the reverse of displacement. In (Ronald V. Clarke, ed.), *Crime Prevention Studies*, Volume 2. Monsey, NY: Criminal Justice Press.

Coleman, Veronica, Walter C. Holton Jr., Kristine Olson, Stephen C. Robinson, and Judith Stewart. 1999. Using knowledge and teamwork to reduce crime. *National Institute of Justice Journal*, October: 16–23. Retrieved June 10, 2009 from ncjrs.gov/pdffiles1/jr000241d.pdf.

Cook, Philip J. and Jens Ludwig. 2000. *Gun Violence: The Real Costs*. New York: Oxford University Press.

Corsaro, Nicholas, Rod K. Brunson, and Edmund F. McGarrell. 2009, October 14. Problem-oriented policing and open-air drug markets: Examining the Rockford pulling levers deterrence strategy. *Crime & Delinquency* [epub ahead of print] Retrieved August 20, 2011 from http://cad.sagepub.com/content/early/2009/10/14/0011128709345955.abstract.

Corsaro, Nicholas and Edmund F. McGarrell. 2009. Testing a promising crime reduction strategy: Re-assessing the impact of the Indianapolis "pulling levers" intervention. *Journal of Experimental Criminology*, 5: 63–82.

Cortes, Ernesto. 1993. Reweaving the fabric: The iron rule and the IAF strategy for power and politics. In (Henry G. Cisneros, ed.), *Interwoven Destinies: Cities and the Nation*. New York: W. W. Norton.

Dalton, Erin. 2003. *Lessons in Preventing Homicide*. East Lansing: Michigan State University, School of Criminal Justice. Retrieved April 22, 2010 from cj.msu.edu/~outreach/psn/erins_report_jan_2004.pdf.

Department of City Planning. 2006. *Census 2000: Census: Pittsburgh—A Comparative Digest of Census Data for Pittsburgh Neighborhoods*. Pittsburgh, PA: Author. Retrieved April 22, 2010 from city.pittsburgh.pa.us/cp/assets/census/2000_census_pgh_jan06.pdf.

Duffee, David E., Brian C. Renauer, Jason D. Scott, Steve Chermak, and Edmund F. McGarrell. 2006. *Measuring Community Building Involving the Police (Final Report)*.

Washington, D.C.: U.S. Department of Justice. Retrieved April 22, 2010 from ncjrs.gov/pdffiles1/nij/grants/213135.pdf.

Eck, John E. 1993. The threat of crime displacement. *Criminal Justice Abstracts*, 25: 527–546.

Federal Bureau of Investigation. 2009a. Table 12: Crime trends by population group, 2007–2008. In *Crime in the United States 2008*. Washington, D.C.: U.S. Department of Justice. Retrieved October 6, 2009 from fbi.gov/ucr/cius2008/data/table_12.html.

Federal Bureau of Investigation. 2009b. Table 20: Murder by state, types of weapons, 2008. *Crime in the United States 2008*. Washington, D.C.: U.S. Department of Justice. Retrieved October 6, 2009 from fbi.gov/ucr/cius2008/data/table_20.html.

Grogan, Paul S. and Tony Proscio. 2000. *Comeback Cities: A Blueprint for Urban Neighborhood Revival*. Boulder, CO: Westview Press.

Graubard, Barry I. and Edward L. Korn. 1999. Predictive margins with survey data. *Biometrics*, 55: 652–659.

Kang, Joseph D. Y. and Joseph L. Schafer. 2007. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22: 523–539.

Kennedy, David M. 1997. Pulling levers: Chronic offenders, high-crime settings, and a theory of prevention. *Valparaiso University Law Review*, 31: 449–484.

Kennedy, David M. 1998. Pulling levers: Getting deterrence right. *National Institute of Justice Journal*, 236: 2–8. Retrieved April 22, 2010 from ojp.usdoj.gov/nij/journals/jr000236.htm.

Kennedy, David M. and Anthony A. Braga. 1998. Homicide in Minneapolis: Research for problem solving. *Homicide Studies*, 2: 263–290.

Kennedy, David M., Anne M. Piehl, and Anthony A. Braga. 1996. Youth violence in Boston: Gun markets, serious youthful offenders, and a use-reduction strategy. *Law and Contemporary Problems*, 59: 147–196.

Klein, Malcolm W. 1971. *Street Gangs and Street Workers*. Englewood Cliffs, NJ: Prentice Hall.

Klein, Malcolm W. 1995. *The American Street Gang: Its Nature, Prevalence, and Control*. New York: Oxford University Press.

Lauger, Timothy. 2010. *Known in "Naptown:" Legitimacy, Reputation, and Violence in the Intergang Environment*. Unpublished Ph.D. dissertation, Indiana University, Bloomington, IN.

McCaffrey, Daniel F., Greg Ridgeway, and Andrew R. Morral. 2004. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9: 403–425. Retrieved April 23, 2010 from rand.org/pubs/reprints/RP1164/.

McDevitt, Jack, Anthony A. Braga, and Shea Cronin, with Edmund F. McGarrell and Tim Bynum. 2007. *Project Safe Neighborhoods: Strategic Interventions—Lowell, District of Massachusetts: Case Study 6*. Washington, D.C.: Bureau of Justice Assistance. Retrieved April 26, 2010 from ojp.usdoj.gov/BJA/pdf/Lowell_MA.pdf.

McGarrell, Edmund, Steven Chermak, Jeremy Wilson, and Nicholas Corsaro, Reducing homicide through a "lever-pulling" strategy. *Justice Quarterly*, 23: 214–231.

McGarrell, Edmund F., Nicholas Corsaro, Natalie Kroovand Hipple, and Timothy Bynum. 2010. Project safe neighborhoods and violent crime trends in US cities: Assessing violent crime impact. *Journal of Quantitative Criminology*, 26: 165–190.

McGarrell, Edmund F., Natalie Kroovand Hipple, Nicholas Corsaro, Timothy S. Bynum, Heather Perez, Carol A. Zimmermann, and Melissa Garmo. 2009. *Project Safe Neighborhoods: A National Program to Reduce Gun Crime—Final Project Report*. Washington, D.C.: National Institute of Justice. Retrieved April 23, 2010 from ncjrs.gov/pdffiles1/nij/grants/226686.pdf.

McGarrell, Edmund F., Natalie Kroovand Hipple, and Nicholas Corsaro, with Ed Pappanastos, Ed Stevens, and James Albritton. 2007. *Project Safe Neighborhoods: Strategic Interventions—Middle District of Alabama: Case Study 5*. Washington, D.C.: Bureau of Justice Assistance. Retrieved April 26, 2010 from ojp.usdoj.gov/BJA/pdf/MD_Alabama.pdf.

National Institute of Justice. 1996. *Victim Costs and Consequences: A New Look*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved April 23, 2010 from ojp.usdoj.gov/nij/pubs-sum/155282.htm.

National Institute of Justice. 2001. *Reducing Gun Violence: The Boston Gun Project's Operation Ceasefire*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved April 23, 2010 from purl.access.gpo.gov/GPO/LPS27453154.

National Institute of Justice. 2002. *Responding to Gangs: Evaluation and Research*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved April 23, 2010 from ncjrs.gov/pdffiles1/nij/190351.pdf.

One Vision One Life. 2005. *Putting a New Face on Violence Prevention: A Strategy to Prevent Violence in Allegheny County*. Pittsburgh, PA: Author.

Papachristos, Andrew V., Tracey L. Meares, and Jeffrey Fagan. 2007. Attention felons: Evaluating Project Safe Neighborhoods in Chicago. *Journal of Empirical Legal Studies*, 4: 223–272. Project Safe Neighborhoods. n.d. Homepage. Retrieved June 10, 2009 from www.psn.gov/.

Ridgeway, Greg and Daniel F. McCaffrey. 2007. Comment: "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data." *Statistical Science*, 22: 540–543.

Robins, James M. and Andrea Rotnitzky. 2001. Comment: "Inference for semiparametric models: Some questions and an answer." *Statistica Sinica*, 11: 920–936.

Roehl, Jan, Dennis P. Rosenbaum, Sandra K. Costello, James R. "Chip" Coldren Jr., Amie M. Schuck, Laura Kunard, and David R. Ford. 2006. *Strategic Approaches to Community Safety Initiative (SACSI) in 10 U.S. Cities: The Building Blocks for Project Safe Neighborhoods*. Washington, D.C.: U.S. Department of Justice. Retrieved March 23, 2010 from ncjrs.gov/pdffiles1/nij/grants/212866.pdf.

Rosenbaum, Paul R. 2002. *Observational Studies*, 2nd Edition. New York: Springer.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70: 41–55.

Rosenfeld, Richard, Robert Fornango, and Eric Baumer. 2005a. Did *Ceasefire, Compstat,* and *Exile* reduce homicide? *Criminology & Public Policy*, 4: 419–449.

Rosenfeld, Richard, Robert Fornango, and Eric Baumer. 2005b. The straw man bluff: Reply to Berk. *Criminology & Public Policy*, 4: 467–469.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277: 918–924.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Silverman, Eli B. 1999. *NYPD Battles Crime: Innovative Strategies in Policing*. Boston, MA: Northeastern University Press.

Skogan, Wesley G. 1990. *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*. New York: Free Press.

Skogan, Wesley G., Susan M. Hartnett, Natalie Bump, and Jill Dubois. 2008. *Evaluation of CeaseFire-Chicago*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Spergel, I. A. 1976. Interactions between community structure, delinquency, and social policy in the inner city. In (Malcolm W. Klein, ed.), *The Juvenile Justice System*. Beverly Hills, CA: Sage.

Stewart, Eric Allen and Ronald L. Simons. 2009. *The Code of the Street and African-American Adolescent Violence*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved April 23, 2010 from purl.access.gpo.gov/GPO/LPS109682.

Tita, George K. and Greg Ridgeway. 2007. The impact of gang formation on local patterns of crime. *Journal of Research in Crime and Delinquency*, 44: 208–237.

Tita, George, K., Jack Riley, Greg Ridgeway, Clifford A. Grammich, Allan Abrahamse, and Peter W. Greenwood. 2003. *Reducing Gun Violence: Results from an Intervention in East Los Angeles*. Santa Monica, CA: RAND, MR-1764-NI. Retrieved April 23, 2010 from rand.org/pubs/monograph_reports/MR1764/.

Venkatesh, Sudhir Alladi. 1997. The social organization of street gang activity in an urban ghetto. *American Journal of Sociology*, 103: 82–111.

Webster, Daniel W., Jon S. Vernick, and Jennifer Mendel. 2009. *Interim Evaluation of Baltimore's Safe Streets Program*. Baltimore, MD: Johns Hopkins Bloomberg School of Public Health, Center for the Prevention of Youth Violence.

Weingart, Saul N., Francis X. Hartmann, and David Osborne. 1994. *Case Studies of Community Anti-Drug Efforts*. Washington, D.C.: National Institute of Justice. Retrieved April 23, 2010 from ncjrs.gov/pdffiles/anti.pdf.

Weisberg, Robert. 2005. Meeting consumer demand in modern criminology. *Criminology & Public Policy*, 4: 471–477.

Weisburd, David and Peter Neyround. 2011. Police science:Toward a new paradigm. In *New Perspectives in Policing*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Weisburd, David, 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.

Weisel, Deborah Lamm. 2002. The evolution of street gangs: An examination of form and variation. In (National Institute of Justice), *Responding to Gangs: Evaluation and Research*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, NCJ 190351. Retrieved April 22, 2010 from purl.access.gpo.gov/GPO/LPS37771.

Wellford, Charles F., John Pepper, and Carol Petrie. 2005. *Firearms and Violence: A Critical Review*. Washington, D.C.: National Academies Press.

Wilson, Jeremy M., Steven Chermak, and Edmund F. McGarrell. 2010. *Community-Based Violence Prevention: An Assessment of Pittsburgh's One Vision One Life Program*. Santa Monica, CA: RAND, MG-947-NIJ. Retrieved August 28, 2010 from rand.org/pubs/monographs/MG947/.

Wilson, William J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: University of Chicago Press.

Winship, Christopher and Jenny Berrien. 1999, June. Boston cops and black churches: New approaches to fighting crime. *Public Interest*.

Jeremy Wilson is an associate professor, research director, and director of the Anti-Counterfeiting and Product Protection Program in the School of Criminal Justice at Michigan State University. He has collaborated with police agencies, communities, task forces, and governments throughout the United States and the world on many salient public safety problems. He has written broadly in the areas of police administration, violence prevention, product counterfeiting, and internal security. His current projects focus on police staffing, resource allocation, consolidation, and the development of anticounterfeit strategy.

Steven Chermak is a professor in the School of Criminal Justice at Michigan State University. His research interests include domestic terrorism, identifying effective strategies for reducing crime/violence, policing, and media coverage of crime and justice. His current research projects include an examination of the life course of far-right extremist groups, state police estimates of the threat of terrorism in the United States, and a project to examine the intelligence practices of state, local, and tribal law enforcement agencies.

POLICY ESSAY

COMMUNITY-DRIVEN VIOLENCE REDUCTION PROGRAMS

Crime policy and informal social control

Megan Ferrier
Northwestern University

Jens Ludwig
University of Chicago

The article by Wilson and Chermak (2011, this issue) addresses an important topic: What policy levers outside of the criminal justice system can help reduce crime and violence in our nation's most distressed and dangerous urban areas? Wilson and Chermak try to answer that question by evaluating the effects of the One Vision One Life program in Pittsburgh, PA. One Vision involved a collaborative problem-solving process resulting in a multipart intervention that included continuing data analysis to guide program activities; the use of street workers to try to connect high-risk young people to social services such as job training and substance-abuse counseling; and some "violence-interruption" activity intended to de-escalate situations in which retaliatory and other types of violence were likely.

In what follows we first discuss whether these analyses are likely to have isolated the causal effects of the One Vision intervention, which is the key question that determines whether we can use these results to guide crime-policy decisions. Unfortunately we think the answer is "no," which is mostly a result of the nature of the intervention and how it was implemented rather than from any fault of the investigators themselves. We then offer some thoughts about the underlying logic behind One Vision, including the (usually implicit) theory behind these sorts of "kitchen sink" interventions that risk factors have more-than-additive effects on criminal behavior, as well as the practical challenges associated with asking real-world organizations to engage in complicated problem-solving and collaborative activities. We close with some discussion about the great value of learning more about how

Thanks to Wesley Skogan for helpful suggestions. This essay was supported in part by grants from the Joyce, MacArthur and McCormick foundations to the University of Chicago Crime Lab. All opinions are our own.

Direct correspondence to Megan Ferrier, Institute for Policy Research, Northwestern University, 2040 Sheridan Road, Evanston, IL, 60208–4100 (e-mail: m-ferrier@northwestern.edu).

DOI:10.1111/j.1745-9133.2011.00775.x

to stimulate more of what we hypothesize might be the most likely "active ingredient" behind programs like One Vision—informal social control.

The Threat of Omitted Variables Bias in Evaluating One Vision

One Vision was implemented in two target areas in Pittsburgh (Northside and the Hill District) in May 2004, and then it was expanded to another target area (on the Southside) in May 2005. The quasi-experimental research design takes advantage of the fact that the intervention was implemented in some places ("treatment neighborhoods") but not in others ("control areas"), to compare the trends in crime preintervention and postintervention in the treatment areas with crime trends in the control areas over the same time period.

The key assumption behind this standard "difference-in-differences" (DD) research design is that the crime trends in the comparison areas tell us something about what would have happened in the "treatment" (program) areas had there been no intervention. Put differently, the DD design assumes that crime trends in the comparison areas are informative about the counterfactual crime outcomes that the treatment areas would have experienced absent One Vision. Although this assumption is not directly testable, we can assess the assumption's plausibility by examining whether the treatment and control areas have similar crime trends during the period *before* the treatment is implemented (see Angrist and Pischke, 2009; Bassi, 1984; Heckman and Hotz, 1989; for an application to crime research, see Ludwig and Cook, 2000). Evidence for divergent crime trends between treatment and control areas during the "pretreatment" period reduces our confidence that any differences in crime trends between treatment and control areas after the intervention can be attributed solely to the intervention itself.

Based on this sort of model specification test, the raw data presented in Figures 1 through 3 in Wilson and Chermak (2011) raise concerns about the potential for omitted variables bias in the article's key empirical estimates. For example, Figure 3 shows that from 2001 to 2002 (i.e., during the pretreatment period), the number of gun assaults in the control areas roughly doubled (from approximately 200 to approximately 400). In contrast, the trend in gun assaults over the same period for two of the study's key treatment areas (the Hill district and the Southside areas) looks fairly flat in Figure 3. Examination of the other figures reveals similar patterns that suggest that the control areas may not provide useful estimates for the counterfactual crime trends we would have expected in the treatment areas absent One Vision.

Of course Figures 1 through 3 in Wilson and Chermak (2011) are just raw data, and the analysts do attempt to make the treatment and control areas more similar by using propensity score matching to adjust for a variety of census-tract–level measures of neighborhood sociodemographic composition, as well as for 1 year (2003) of pretreatment crime levels (homicide, aggravated assault, and gun assaults). Table 1 in Wilson and Chermak suggests that the propensity score matching seems to do reasonably well in adjusting for census tract characteristics because the average tract characteristics for the control group after reweighting the data with the estimated propensity scores look fairly similar to the average tract characteristics of the treatment group. But even after reweighting, the average pretreatment crime rates in the treatment and control areas continue to be quite different. For example, the homicide rate in 2003 (the "pretreatment" period) in the target areas was equal to 0.41; the homicide rate in the nontarget areas before the propensity score weighting was equal to 0.61 (approximately 50% higher than in the treatment areas), whereas even the reweighted comparison area average is still 0.63. A more formal test of similarity in pretrends between treatment and control areas would be to reestimate the full DD model with covariates using only data from the pretreatment period, selecting different pretreatment years to define placebo treatment indicators, which should be statistically indistinguishable from zero if the DD assumptions are met.

As the key to the credibility of the DD design is to have similar pretrends in the outcome of interest between the treatment and control areas, perhaps a better way to construct the analysis would have been to focus explicitly on matching treatment with candidate control areas using the entire history of pretreatment crime data that are available. This method would be more likely to yield a valid comparison group than Wilson and Chermak's (2011) approach, which seemed to match on just 1 year of pretreatment crime data together with sociodemographic tract attributes that are imperfect predictors of local-area crime rates. Moreover, because Figures 1–3 suggest the two different treatment areas seem to follow different pretreatment crime trends, there would be value in constructing different weighted groups of control areas—or "synthetic controls"—for each treatment area (see Abadie, Diamond, and Hainmueller, 2010, Abadie and Gardeazabal, 2003).

Our concern about the threat of omitted variables bias is relevant for thinking about the substantive questions Wilson and Chermak's (2011) article raises about why Pittsburgh's One Vision program was apparently less effective than similar interventions implemented elsewhere, such as Chicago CeaseFire or Baltimore's Safe Streets Project. Wilson and Chermak devote a great deal of discussion to differences across programs in staffing, emphasis, and implementation. But under Occam's razor, a simpler and more likely reconciliation of the different program results is that the true uncertainty band around the One Vision estimates is simply too large to rule out the null hypothesis that all of the programs have similar impacts.

Who knows whether alternative synthetic-control methods would have been able to create more suitable comparison areas and to narrow the uncertainty band around the One Vision impact estimates. The fundamental problem is that in Pittsburgh, as in so many places around the country, policy makers and practitioners want to know what their new pilot programs are achieving, but they do not design and implement these programs in advance in ways that will maximize the chances of a successful impact evaluation later on. Random assignment is one way to increase the odds of learning how the program works, but it is not the only way. For example, in some applications, candidate target areas could be systematically ranked on some set of criteria, which would allow for a regression discontinuity (RD) study (see Owens and Ludwig, 2011). Policy makers and practitioners who want to say that they have done a real evaluation too often just implement the program in some way that is convenient, and then they hope that researchers can come in afterward and work some econometric magic to uncover program impacts. But even methodological magic has its limits.

KISS?

Wilson and Chermak's (2011) discussion about how One Vision was implemented highlights the difficulty that government and nonprofit organizations often have in carrying out complex problem-solving activities and multipronged, collaborative interventions. Problem-oriented policing, for example, even in cities like San Diego, CA, that have made long-term efforts to support, facilitate, and encourage its use, is in practice rare at large scale despite evidence of its effectiveness, likely because it can be so challenging and demanding to implement (Cordner and Biebel, 2005). Implementation becomes even more complicated when more than one organization is involved, in part because even basic things like data sharing are so difficult. Agencies and organizations have strong incentives to protect information, partly to reduce the risks of revealing mistakes, making them reluctant to share and then analyze and use more integrated data. Note that the real-world consequences of these implementation challenges are, if anything, likely to be understated in the program evaluation literature. The reason is that many of the highest quality studies that use randomized experimental designs are analogous to what medical researchers call "efficacy trials," which test small "hot house" programs that are carried out with high fidelity under the watchful eye of bespectacled, laptop-toting researchers, rather than "effectiveness trials" that test interventions as they would actually be implemented at-scale in the real world.

Given the great difficulty of implementing complicated interventions in the real world, it is useful to reexamine the (often implicit) conceptual framework behind the sort of multicomponent "kitchen sink" or "synergistic" program that One Vision represents.[1] The underlying theory for many kitchen sink interventions is that risk or protective factors have interactive (more-than-additive) effects, which implies that programs should act simultaneously on multiple fronts in order to make the biggest possible difference with available resources. More formally, let Y_i be some measure of criminal involvement for individual (i), and let X_{1i} and X_{2i} be different candidate protective factors (say, human capital, and "health capital," which in this application could be thought of as "being drug-free"). The theory described by Equation 1 is that coefficient β_3 is negative and large (in

1. For a discussion of alternative ways of testing the basic logic behind kitchen-sink experiments without having to implement a kitchen-sink program, see Ludwig, Kling, and Mullainathan (2011).

absolute value) relative to the main effects β_1 and β_2.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + \upsilon_i \tag{1}$$

One empirically testable prediction of the "synergy" theory implied by equation (1) is that relatively lower-risk people should benefit relatively more from social policy interventions. The intuition is that if X_{1i} and X_{2i} exert an important interaction effect on the key outcome of interest, Y_i, then an intervention that tries to (say) improve human capital, X_{1i}, should have a more pronounced impact on those who already have a high level of health capital, X_{2i}.

Yet in real-world evaluations, evidence often shows us that the people who have the highest levels of baseline risk—and so are presumably more severely disadvantaged on multiple dimensions that affect the outcome—are the *most* responsive to social policy interventions, even to those interventions that try to influence just one or two relevant risk or protective factors. For example, cognitive-behavioral programs seem to have more pronounced impacts on those criminal-justice–involved people who have the most extensive prior criminal records (Landenberger and Lipsey, 2005). In the Tennessee STAR class-size reduction experiment, minority children and poor children (i.e., those at highest risk for educational failure) benefit the most from smaller elementary school class sizes (Krueger and Whitmore, 2001). In a study of the long-term effects of Head Start, arrest rates seem to decline relatively more among African Americans than among whites (Garces et al., 2002).

These findings are more consistent with a theory of diminishing marginal returns to social program investments than with the synergy theory described by Equation 1. The intuition behind diminishing marginal returns to behavioral change is quite straightforward: giving an extra thousand dollars to a low-income single-parent household on the South Side of Chicago should do more to change the schooling outcomes and delinquency risk of children in that home than would giving that thousand dollars to the parents of, say, Lindsay Lohan. (More formally the implication is that equation 1 should include squared terms for X_{1i} and X_{2i}, and that the coefficients on these squared terms are large in absolute value relative to the coefficient on the interaction of X_{1i} and X_{2i}).

Given the conceptual uncertainty about the value of trying to intervene on all fronts simultaneously, and the practical difficulties of implementing complicated multipart interventions in the real world, policy analysts and policy planners should think about whether there is value in prioritizing simplicity in policy design. We might heed the design principle articulated by aerospace engineer Kelly Johnson of Lockheed Skunk Works, who famously coined the term "Keep it simple stupid" (i.e., the KISS principle).[2]

2. For more information, go to en.wikipedia.org/wiki/KISS_principle.

Promoting Informal Social Control

Even if we were not concerned about the threat of omitted variables bias with Wilson and Chermak's (2011) estimates of One Vision's impacts, it would still be difficult to disentangle empirically the mechanisms through which the intervention reduced crime given the program's multiple components. With that caveat in mind, our own guess is that the most potentially important "active ingredient" with One Vision is the attempt to stimulate informal social control. This approach strikes us as more likely to be a key mechanism than, say, connecting youth to social services, which assumes that information (rather than motivation or the limited availability of high-quality social services) is the key barrier to service receipt by high-risk youth.

One reason to suspect that informal social control might be a key mechanism through which programs like One Vision might be able to reduce crime comes from the seminal work by Sampson, Raudenbush, and Earls (1997), showing that cross-neighborhood variation in violent crime rates is strongly predicted by neighborhood informal social control—or by what they call "collective efficacy" (see also Sampson, Morenoff, and Gannon-Rowley, 2002). What remains poorly understood are the specific ways in which public policy can causally intervene to strengthen the capacity of neighborhoods to carry out informal social control – particularly distressed, dangerous neighborhoods.

The Chicago Ceasefire program tries to strengthen local informal social control by hiring former gang-involved people to serve as "violence interrupters" and intervene to prevent retaliatory violence (Skogan, Hartnett, Bump, and Dubois, 2008). One practical challenge with the Chicago Ceasefire model is that keeping full- or even part-time people on the payroll is expensive, and government at every level in the United States is extremely budget constrained. Another potential challenge is political—risk-averse policy makers may worry about providing funding to an intervention that hires former gang members, given the risk for relapse is not zero and (fairly or not) the public relations downside from even a single relapse by a formerly gang-involved person on the public payroll could be considerable.

Are there alternative ways to stimulate informal social control? Our conversations with Columbia University sociologist Sudhir Venkatesh (personal communications, March 9, 2011) suggest the different roles that various neighborhood residents already play (at least in some parts of Chicago's high-crime South Side) in trying to defuse and de-escalate potentially violent events, particularly those involving young people. In our view, we would find great value in learning more about the types of residents who are involved with informal social control, what specific activities they carry out, and how public policy can help support them in their roles (and help overcome whatever barriers they face). Ethnographic research of the sort that Venkatesh carries out, which may help inform the design of new pilot programs and randomized intervention studies, would be an important complement to the sort of quantitative evaluation evidence presented in the One Vision study.

References

Abadie, Alberto and Javier Gardeazabal. 2003. The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93: 113–132.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105: 493–505.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bassi, Laurie J. 1984. Estimating the effect of training programs with non-random selection. *The Review of Economics and Statistics*, 66: 36–43.

Cordner, Gary and Elizabeth Perkins Biebel. 2005. Problem-oriented policing in practice. *Criminology & Public Policy*, 4: 155–180.

Garces, Eliana, Duncan Thomas, and Janet Currie 2002. Longer term effects of Head Start. *American Economic Review*, 92(4): 999–1012.

Heckman, James J. and Joseph V. Hotz. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, 84: 862–874.

Krueger, Alan B. and Diane M. Whitmore. 2001. The effect of attending small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*, 111: 1–28.

Landenberger, Nana A. and Mark W. Lipsey. 2005. The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1: 451–476.

Ludwig, Jens and Philip J. Cook. 2000. Homicide and suicide rates associated with implementation of the Brady Handgun Violence Prevention Act. *Journal of the American Medical Association*, 284: 585–591.

Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25: 17–38.

Owens, Emily and Jens Ludwig. 2011. *Regression Discontinuity Designs for Criminology Research*. Working Paper. Cornell University, Ithaca, NY.

Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology*, 28: 443–478.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277: 918–924.

Skogan, Wesley G., Susan M. Hartnett, Natalie Bump, and Jill Dubois. 2008. *Evaluation of CeaseFire-Chicago*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Wilson, Jeremy M. and Steven Chermak. 2011. Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life. *Criminology & Public Policy*. This issue.

Megan Ferrier is a research associate at Northwestern University's Institute for Policy Research and a research associate at the University of Chicago Crime Lab.

Jens Ludwig is the McCormick Foundation Professor of Social Service Administration, Law, and Public Policy at the University of Chicago, director of the University of Chicago Crime Lab, and research associate at the National Bureau of Economic Research.

POLICY ESSAY

COMMUNITY-DRIVEN VIOLENCE REDUCTION PROGRAMS

Comprehensive gang and violence reduction programs

Reinventing the square wheel

Malcolm W. Klein
University of Southern California

This essay has three components about factors that limit knowledge building for policy purposes. Several of these factors are directly described in the report on the One Vision program (Wilson and Chermak, 2011, this issue). I will comment first on several "lost issues," that is, issues that often are overlooked in the extensive literature on gang and violence programming. I will then follow with a list of other warnings about factors that limit what is learned (or learnable) from such programming. Finally, I will illustrate some of these concerns in a new, complex, and yet promising program now entering its third year.

Three Lost Issues

Gap Between Data and Policy

Wilson and Chermak (2011) note correctly that a considerable gap exists between "promising" strategies and demonstrably successful programs (demonstrable in scientific terms, one should add). Various project reports and—most notably—government summaries overstate promise in the face of mixed and negative results (Howell, 2007; OJJDP, 2010; Spergel, Wa, and Sosa, 2006). The collection of reports in enforcement-oriented programs edited by Decker (2003) and the comprehensive review of almost 60 approaches to gang reduction offered by Klein and Maxson (2006) document not just a gap but also a woeful absence of (a) well-implemented programs and (b) adequately evaluated programs. Indeed, one could infer from the Klein and Maxson summaries that almost everything is "promising" because

I am grateful to Margaret Gatz, Karen Hennigan, and Cheryl Maxson for their helpful comments on an earlier draft of this essay. Direct correspondence to Malcolm W. Klein, Department of Psychology, University of Southern California, Los Angeles, CA 90089 (e-mail: mklein@usc.edu).

DOI:10.1111/j.1745-9133.2011.00758.x

so little has been tested properly. The report by Wilson and Chermak represents a distinct departure from this pattern, although it unfortunately describes a program that actually increased levels of violence. A wise author noted decades ago (I have lost the reference) how remarkable it was that the most carefully evaluated social programs turn out to be the most likely failures.

Use of Former Gang Members in Gang Intervention

A second largely overlooked and seriously underresearched factor is the problem of using former street gang members, or former gang leaders, as instruments of change through outreach to potential and current gang members. Spergel's (2007) data from Chicago indicate that such an outreach can be beneficial, but these data almost stand alone. I want to suggest that there are both advantages and dangers in having former gang members serve as "interventionists" (to use the newest terminology).

The advantages are obvious and often noted by interventionists and their supporters. Gang-experienced outreach workers have "street cred" with almost instant opportunities to establish rapport with current or "wannabe" gang members. They also have access to gang intelligence about member relationships, rivalries, and impending conflicts that can be used to defuse hostilities.

The dangers may be less obvious; yet they can disrupt outreach efforts. The first of these dangers is that, in establishing rapport with targeted gang members, gang outreach workers often use their own prior gang status and exploits as legitimators of their credibility, thereby glorifying the attitudes and activities they are supposed to be discouraging among their gang clients. A second danger is the "walk in my shoes" assumption that only former gang members can be gang interventionists. One often hears this claim from former gang members in outreach programs, but it is nonsense. I have known and worked with effective gang workers who had no prior gang membership. They were social workers, recreation workers, and even probation officers!

A third danger—often cited by the police—is that former gang members in the outreach role can "turn off" agents from other programs, creating suspicion and antagonism that defeat collaborative efforts. Understandably, they are also reluctant to share information ("intelligence") about their gang member clients. The police, in particular, resent this failure (while being equally guilty of the failure to share information).

Finally, there is I believe a legitimate concern that turning ex-gang members or—worse yet—former gang leaders into activists for prosocial and antigang values places an unfair burden on them. They have been socialized by anywhere from 1 to 10 years or more in the gang culture. Now, we ask them to turn around 180° to eschew that culture and, in effect, work against it. How does this new role fit with their former and current gang peers? How does it fit with their habituated gang self-identities? Experienced observers of the gang programming arena can cite case after case of ex-gang members turned gang interventionists who fell off the wagon—committed new crimes or were assaulted, even killed, while in their new roles. You can take the member out of the gang, but sometimes you cannot take the gang out of the member.

Unintended Effects on Gang Cohesion

I want to emphasize a third lost issue raised in the article by Wilson and Chermak (2011). This is the problem of increasing gang cohesion inadvertently through the various practices involved in gang intervention. Studies as early as the 1970s (Klein, 1971) and as recent as the last decade (e.g., Thornberry, Krohn, Lizotte, Smith, and Tolin, 2003) have demonstrated the enormous effect that gang involvement has on individual illegal behavior among gang members. A strong case can be made, empirically as well as conceptually, from a social-psychological framework, in which programs that emphasize group processes already in place in gangs can (or will) increase gang cohesiveness and, therefore, gang crime (see Klein and Maxson, 2006, Ch. 5). The policy question then becomes, how can one intervene with groups like street gangs without unintentionally doing more harm than good? Whatever the answer suggested to gang interventionists, it is not one that has been welcomed by them.

Additional Issues in Gang Programming

This section of the essay expands the number of concerns raised by gang and violence reduction programs. I have chosen six of these to illustrate the potential for failing to mount demonstrable program successes.

The first of these is well known to scholars and almost totally ignored by public officials and program administrators: the regression effect. Gang and violence reduction programs are usually mounted as problems are increasing or have reached a tipping point of intolerance for violence levels. A quiet period does not generate much intervention activity. Peaks are followed by valleys. Programs initiated at peak times will normally play out after peaks are reached. The appearance of success may be nothing more than a demonstration of expected normal decrease. For example, the claimed success of Spergel's (2007) Little Village Project took place among the oldest gang members (who are more likely to desist from gang activity) and those with the higher levels of criminal activity (which are known to decrease over time). No such program "success" was demonstrated among the younger gang members. This regression problem could of course be obviated by the use of control designs—comparable areas, gangs, and gang members—but intervention programs almost never implement controlled research designs.

A second but related problem is that, in many instances, gang and violence reduction programs are initiated in locations that do not yield readily to broader generalizations. The One Vision program, for instance, took place in Pittsburgh, a city that was among the last major urban areas to develop a gang problem, and one in which the gang problem grew out of a drug distribution development rather than out of the usual obverse pattern. Gang problems in Los Angeles and Chicago take place in settings simply not typical of the several *thousand* jurisdictions with gang problems. These two cities are populated heavily by "traditional" street gangs (Klein and Maxson, 2006, Ch. 5), which are unlike the street gangs in most jurisdictions. Case studies in special circumstances do not lead readily to valid generalizations (see Spergel, 2006, for mixed results when six jurisdictions are involved).

A third problem is that of separating gang violence, specifically, from other sources of violence. They are not the same, although obviously some overlap occurs. From my viewpoint, the greatest concern here is that confusing the two obfuscates the special *group* nature of gang violence in favor of simpler efforts to understand and reduce individually based violent acts. If exaggerated group processes in gangs are overlooked—group norms, cohesiveness, structural variations, oppositional culture, and violence as a rallying cry for gang members—then the special impediments that groups present to much violence reduction programming will prevent much success.

A fourth, almost self-defeating character of many gang and violence reduction programs is their attempt to be comprehensive. From an evaluation viewpoint, this means that such programs will be multicomponent efforts. This sounds good and conceptually appropriate, but it raises the difficulty of disaggregating the program components to determine which are more and less successful, in what settings, with what types of program clients, and in what combination. The most assertive attempt at such component disaggregation is provided by Spergel's 2006 analysis of data from six comprehensive-model sites, an exercise in frustration at the very least.

A fifth problem of any program inserted into or imposed on a community is that of accounting for confounding effects of other program efforts. Police engage in special deterrent activities; social activists launch community-based responses to violence and gang suppression; ongoing welfare and youth agency programs not incorporated into the new comprehensive program renew or alter their character to maintain their positions; and political and philanthropic initiatives call out for attention and support. Essentially, this is the disaggregation problem once again, but this time with the task of separating out the effects of other efforts not under the control of the new program, and with little or no data collection to assess the confounds inherent in ongoing extra program activities.

Finally, in this discouraging list of problems, is that equivocal or negative results of large-scale intervention programs seldom lead to altered future programming. Such results more commonly lead to withdrawal of financial support, discouragement and apathy, and a search for "better ways" to spend our money and expend our energy. However, I can cite two prominent and contrasting exceptions to this pattern. The first is the effort of the Office of Juvenile Justice and Delinquency Prevention (OJJDP) in the U.S. Department of Justice to follow prematurely on the "Spergel Model" to develop variations in comprehensive gang programming. Here, the determination of OJJDP to forge ahead despite equivocal outcomes (including poor implementation as well as mixed results) has led to comprehensive programming as more of a social movement than a demonstrated successful initiative (Decker and Curry, 2003; Fearn, Decker, and Curry, 2006; OJJDP, 2010).

The other, far more encouraging exception is provided by the Gang Resistance, Education, and Training Project (G.R.E.A.T.). As described by Esbensen, Osgood, Taylor, Peterson, and Freng (2001), this gang prevention program, over 4 years of assessment, yielded slightly positive effects on intermediate outcomes but no effects on gang and criminal outcomes. However, as noted elsewhere (Klein and Maxson, 2006, Ch. 3), the result was that the outcome data were fed back to the program managers. These managers in turn produced a modified program that was in turn introduced into a large number of communities to test the effectiveness of the data-based new approach. Only tentative, first-year results have been reported, but the willingness of the research evaluators and the program managers to forge this new effort deserves special attention.

Square Wheels and Scholarly Intervention

Given all the foregoing discussion, it might seem surprising that I describe an effort by an informal consortium of criminologists to frame public policy in yet another large, comprehensive gang control program. I am one of a dozen experienced academic gang scholars and program evaluators recruited by the program's original director to help frame the program model and provide guidance in its evaluation.

Currently entering its third year of an anticipated 5 years, the program is designed to engage literally thousands of gang-prone youth and active street gang members in a three-pronged approach. The first is the prevention of gang activity by youth assessed as most at risk of joining gangs. The second is case management of known gang members aged 14 to 24 years to reduce their gang activity. The third is crisis intervention in gang confrontations, especially those reaching serious levels of violence. The program is run by the city with contracted services from private agencies. Additionally, there is a 5-year, heavily funded evaluation contracted to an independent, nationally respected corporation.

The independent group of scholars has been involved significantly in several of these areas. First they helped to develop the request for proposals (RFPs) for the prevention program, including the development of a gang-joining risk instrument to be used by the agencies to select prevention clients. Second, they helped to frame the RFP for the intervention program, including both case management and crisis intervention. However, although the intervention agencies were funded, it was not until the end of the second program year that the gang scholars were asked to devise a system for selecting case management clients from the huge pool of available gang members. Third, they were asked to frame the evaluation RFP, were involved in selecting the evaluation contract awardee, and have served as research advisors to the evaluation.

After 2 years, where does the program stand? There is now an agreed-on prevention model being applied to the youth most at risk of gang joining as assessed by the gang scholars' risk-assessment tool. Thousands of appropriate clients have been located. By way of contrast, no case management model is available for actual gang member selection and intervention, although new efforts in this direction are being mounted. The crisis intervention component is being implemented, but it has no input from the group of scholars. Finally, the model for the overall program evaluation has been under intense scrutiny by the group of gang scholars as well as by the city council and city controller. It has been delayed by the slow program development and by serious disputes about its design and data collection commitments. It is a work in progress.

As the reader may surmise, the program has been subject to many of the lost issues and other problems elucidated previously in this essay. The truly unique aspect is its early acceptance of the experience of the group of scholars. The use, indeed the value, of these people, might be in showing what data-based intrusion can accomplish. That, in any case, is what keeps these scholars involved despite the frustrations of policy involvement. I will mention just four issues that have been paramount to them in maintaining their activity. All I ask of the reader is to view these comments as relating to the use of criminologists in public policy, not as to the character of the program emanating from that policy.

Perhaps the most contentious issue raised between the group of scholars on the one hand and the program managers and contracted evaluators on the other was (and is) the development of control or comparison clients, gangs, and/or areas. The program sites were selected as the “worst” in the city, thus reducing the chances of random assignment to comparison and service areas. Program directors emphasized the immorality and political sensitivity of “denial of service” to control youth, gangs, or areas. Evaluators stressed the “impossible costs” of including control or comparison youth, as well as the limitations seemingly imposed by human consent requirements. The group of scholars is still pursuing this issue by offering alternative approaches for a comparative design; the jury is out.

Within the program administration, a series of obstacles (some temporary and some not) has emerged. These obstacles include changes in leadership and conceptual focus for the program models, absence of gang expertise in the administrative staff, reliance on non–gang-related standard operating procedures, and unfamiliarity with the major complexity of comprehensive programming. All this, slowly, is being surmounted as the third program year is approaching.

Third in this list is the built-in resistance to new service models and the requirements of the evaluators. Prevention service providers have slowly but steadily accepted the new necessities; the case management and crisis workers—generally self-labeled as “gang interventionists”—have proven far more resistant at the same time as program management has seemed somewhat cowed by the vehemence of this resistance. At this writing, however, there are signs of positive change and some accommodations. The group of scholars, until recently, has been held at arms’ length from that battle.

Finally, as a corollary of the preceding changes, is the set of decisions about how to select clients for the prevention component and those for the case management or intervention component. Here, the group of scholars has been involved heavily in establishing the criteria for selecting prevention clients. And after 2 years, the first serious attempt to involve the scholars similarly in case management client selection has been initiated. The outcome and its acceptance by service providers is unclear, but the effort is now underway.

In sum, this new and huge project (up to $100,000,000 over 5 years) shows many characteristics of past comprehensive approaches of dubious value (keep in mind the increased violence reported for the One Vision Program by Wilson and Chermak, 2011). It could be another square wheel. However, intrusion of the group of scholars has had a considerable positive impact on program models and evaluation design. Has this impact been enough? Probably not, is my guess. Can this be improved? Probably so, is also my guess. But as noted, it is all a work in progress, an interesting and occasionally exciting exercise in criminology and public policy.

References

Decker, Scott H. (ed.). 2003. *Policing Gangs and Youth Violence*. Belmont, CA: Wadsworth.

Decker, Scott H. and G. David Curry. 2003. Suppression with prevention, prevention without suppression: Gang intervention in St. Louis. In (Scott Decker, ed.), *Policing Gangs and Youth Violence*. Belmont, CA: Wadsworth.

Esbensen, Finn, D. Wayne Osgood, Terrance J. Taylor, Dana Peterson, and Adrienne Freng. 2001. How great is G.R.E.A.T.? Results from the longitudinal quasi-experimental design. *Criminology & Public Policy*, 1: 87–118.

Fearn, Noelle E., Scott H. Decker, and G. David Curry. 2006. In (Arlen Egley, Cheryl L., Maxson, Jody Miller, and Malcolm W. Klein, eds.), *The Modern Gang Reader*, 3rd Edition. Los Angeles, CA: Roxbury.

Howell, James C. 2007. Menacing or mimicking? Realities of youth gangs. *Juvenile & Family Court Journal*, 58: 39–50.

Klein, Malcolm W. 1971, *Street Gangs and Street Workers*. Englewood Cliffs, NJ: Prentice Hall.

Klein, Malcolm W. and Cheryl L. Maxson. 2006. *Street Gang Patterns and Policies*. New York: Oxford University Press.

Office of Juvenile Justice and Delinquency Prevention (OJJDP). 2010. *Best Practices to Address Community Gang Problems: OJJDP's Comprehensive Gang Model*. Tallahassee, FL: National Gang Center.

Spergel, Irving A. 2007. *Reducing Gang Violence: The Little Village Gang Project in Chicago*. Lanham, MD: Altamira Press.

Spergel, Irving A., Kwai Ming Wa, and Rolando Villarreal Sosa. 2006. The comprehensive, community-wide gang program model: Success and failure. In (James F. Short, Jr. and Lorine A. Hughes, eds.), *Studying Youth Gangs*. Lanham, MD: Altamira Press.

Thornberry, Terence P., Marvin D. Krohn, Alan J. Lizotte, Carolyn A. Smith, and Kimberly Tolin. 2003. *Gangs and Delinquency in Developmental Perspective*. Cambridge, U.K.: Cambridge University Press.

Wilson, Jeremy M. and Steven Chermak. 2011. Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life. *Criminology & Public Policy*. This issue.

Malcolm W. Klein, Ph.D., was trained in psychology; he spent his academic life in sociology and built his research career in criminology. His professional awards have come in all three fields, but his personal rewards have derived principally from the application of social-psychological concepts and methods to local, national, and international crime issues. His home base has been at the University of Southern California; however, his emotional roots lie in the Adirondack Mountains of New York, and his pleasures often have resided in European travels as revealed in his latest book, *The Street Gangs of Euroburg* (2009, iUniverse).

POLICY ESSAY

COMMUNITY-DRIVEN VIOLENCE REDUCTION PROGRAMS

Whither streetwork?

The place of outreach workers in community violence prevention

David M. Kennedy
Center for Crime Prevention and Control John Jay College of Criminal Justice

Wilson and Chermak's (2011, this issue) evaluation of Pittsburgh's One Vision One Life program adds to a growing body of evaluation research and field experience showing that outreach workers do not reliably reduce, and can even promote, street violence. More precisely, this study, and the broader record, shows this about outreach workers *operating in a particular way*. This result should probably not surprise us. Although this kind of streetwork has enjoyed a recent national vogue, it in fact has a history that goes back a couple of generations, and it is not particularly promising. At another point in that history, findings similar to Wilson and Chermak's and to other recent research effectively put an end to streetwork as an important element in community violence prevention. To do that again, I firmly believe, would be a great error. This policy essay will look at where we currently are, and what we currently know, about streetwork; what that should tell us about the kind of streetwork we should be pursuing; and how we should understand and act on the most recent body of research. In particular, it will look at the way streetwork organizations and streetworkers themselves do and do not coordinate and cooperate with law enforcement, particularly police departments. It will make a case that both research and field experience favor appropriately close relationships. It will then raise the idea that different sorts of relationships could have different meanings for community and offender notions of police legitimacy, with potentially powerful criminogenic or crime control implications.

I should say here that I have been actively involved in this debate for some time. In the relatively small world of community violence prevention, it has often been said that I

Direct correspondence to David M. Kennedy, Director, Center for Crime Prevention and Control, John Jay College of Criminal Justice, 555 West 57th Street, Room 601, New York, NY 10019 (e-mail: dakennedy@jjay.cuny.edu).

DOI:10.1111/j.1745-9133.2011.00770.x

am opposed to streetwork; this conversation has occasionally broken out into the public view (Ellison, 2010). That is absolutely incorrect. I believe entirely that streetwork can be a hugely important element in effective violence prevention, and that some research and field experience shows that to be true. I *have* been deeply concerned that one current version of streetwork, and one set of ideas about how streetwork should be deployed, is mistaken and that it has been broadly endorsed and adopted without due attention to actual impact on the ground. Wilson and Chermak (2011), and other relatively new research, are very important inputs into any such debate, and they will allow it to proceed in a far more concrete and empirically grounded fashion.

That debate has not been arid, dispassionate, and rooted in evidence and social science. It has, as one of my cocommentators writes, been about a rapidly growing and very appealing near-social movement centered around the "Chicago CeaseFire" model of streetwork (see Papachristos, 2011, this issue). That movement has grown, he also writes, until recently almost entirely without research and evaluation; has been characterized, promoted, and anointed as successful regardless; and has now assumed the character of what he describes as a "too big to fail" near-juggernaut. This so far advanced, he fears, that evidence will no longer matter. He is, I believe, absolutely correct about, at least, the former. We should understand that there is nothing new here. This is, in essence, how things like this happen. The history of crime prevention and control is in large part one of exactly this progression: plausible and appealing ideas that gain broad traction and support on the basis of little or no evidence and then turn out to be ineffective or deeply flawed. DARE, GREAT, mandatory arrest for domestic violence, Project Exile, boot camps, and a litany of others come from this mold.

So, however, do a range of ideas that *have* proved out: problem-oriented policing, to take one example. It is nearly always the case that getting a big idea out into practice in a meaningful way, winning professional and public support, generating interventions on the ground in a variety of settings, and creating the possibility for a useful range of impact assessments means stepping outside the neat boundaries of a social science research program. Most of those big ideas will not prove out in practice and that history will sometimes look political, irresponsible, and wasteful; some will, and their histories will look brave, insightful, and prescient. This is the way real public work happens: recognizing that reality does not mean that evidence does not matter. Rather, it is vitally important to get and act on evidence, particularly evidence of impact, when that becomes practical. We are reaching that moment with the Chicago CeaseFire version of streetwork.

I have extensive experience with streetworkers in cities across the United States, feel their appeal and that of their work, and share the enthusiasm for programs based on who they are and what they do. Streetworkers are, as a group, enormously powerful and impressive. Individually they tend to be people who come from the streets; have searing personal histories, have made very difficult personal journeys and have undergone profound transformations, and who live lives of extraordinary dedication to the most vulnerable and lost in their communities. Many are doing violence prevention and outreach work more or less on their own, often with little or no outside support or resources, before any programmatic structure comes along for them to join up with. They have enormous credibility and standing with hard-core street offenders; they are uniquely situated to draw those offenders off the streets, build bridges to social service programs, interrupt and mediate disputes, challenge toxic street norms, and the other elements of the core streetwork portfolio. And they all have critical success stories: this shooter diverted, this kid pulled out of the gang, and this beef squashed. Some of those successes are substantial and lasting, such as Tyrone Parker—a streetwork legend in Washington, DC—and his truce work with gangs in the Benning Terrace housing project there.[1] Streetworkers are, in formal terms, important elements in producing collective efficacy and informal social control. In ordinary language, they are saints.

The problem is that this has nowhere translated into routine, reproducible public safety outcomes. The evaluation and the broader field record are both clear, and consistent, on this. Tita and Papachristos (2010) report on 1960s-era evaluations of streetwork programs in Los Angeles, Chicago, and Boston that found little or no positive impact, and even iatrogenic effects. One of the most influential deployments of streetworkers, in conjunction with other efforts, was in Irving Spergel's famous Chicago "Little Village" project, which showed positive outcomes for program-involved gang youth but *increases* in area-level gang violence (these increases were smaller than in the most relevant comparison area, but they were increases nonetheless) (Spergel, Grossman, and Wa, 1998). Tyrone Parker in Washington, DC and those like him nationally were unable to produce enough of their individual-level or collective, Benning Terrace-type successes to add up to overall, meaningful community outcomes.

The experience in Boston, where I first encountered streetworkers as part of what became the Operation Ceasefire project, is typical. Violence in Boston peaked in 1990, with one core city response being to launch a streetwork program. It was robust and relatively large, well funded, and well managed. Streetworkers had standing: they were city employees and the initiative got personal attention from Boston Mayor Menino. It had close ties to social service agencies and had a relatively close working relationship with the Boston Police Department. It also did not work. This is not to say it had no impact, which is not provable and which I personally suspect was not the case. It is to say that whatever impact it had was not anything like sufficient: Lethal and near-lethal violence remained historically high in the city's active neighborhoods, and Boston streetworkers themselves were openly despairing about their ability to address it.

As is now well known, streetworkers in Boston became an important part of the Operation Ceasefire strategy and partnership, which included law enforcement, other

1. Laura Maggi, "Touch neighbourhood turns around as ex-gang members wise up." *Nation's Cities Weekly*, June, 1999.

community actors, and service providers working closely together in specific ways to address the city's violence problem, at which point violence went down dramatically. When Operation Ceasefire was dismantled in 2000 violence went up again, even though the streetworkers remained active. The same thing happened in the Stockton, CA, replication of Ceasefire, Operation Peacekeeper: A streetworker cadre was created as part of Peacekeeper, which was overall very effective, but when the larger Peacekeeper strategy ceased implementation, violence increased rapidly and dramatically.[2] Similar on-the-ground results have been seen in, to my personal knowledge, Rochester, NY; Cincinnati, OH; and Oakland, CA. In Cincinnati, for example, a streetwork project modeled explicitly on Chicago CeaseFire was deployed in one active neighborhood, without apparent effect. It was then incorporated as a central element of the larger, city-wide Cincinnati Initiative to Reduce Violence, modeled explicitly on the Operation Ceasefire framework and partnership, which has produced a 40%-plus reduction in gang homicide (Engel, Tillyer, and Corsaro, 2011).

The recent body of evaluations of the Chicago CeaseFire streetwork model, which is discussed in detail in the Wilson and Chermak (2011) article and by my cocommentators, paints the same basic picture. In Chicago itself the results are mixed; in Pittsburgh, Newark, and Baltimore, they find at best mixed, often negative, and sometime iatrogenic impact.

We should be very careful about the conclusion we draw from this record. It no more says that streetwork "doesn't work" than George Kelling's famous Kansas City preventive patrol experiment showed that patrol "doesn't work." Kelling showed, arguably, that *random* patrol doesn't work. Other forms of patrol, we know, *do* work. We have not given up on community economic development because the war on poverty didn't work; on engaging with kids because DARE and GREAT don't work; or on prosecuting dangerous offenders because Exile doesn't work. The research and field record, however, is now clear enough that we absolutely should, at this point, understand that the Chicago CeaseFire model cannot now be viewed, as its organizers say, as "a national public health strategy that has been scientifically proven to reduce shootings and killings."[3]

If that is true, but if there is still merit in streetwork, which I firmly believe, then the question is, how should effective streetwork be framed and deployed? I believe there is enough in the record to begin to answer this, as well.

An essential element of the Chicago CeaseFire model is a deliberate distance between law enforcement and outreach workers and organizations. This has been framed as critical to maintaining streetworkers' street credibility and personal safety. Some Chicago CeaseFire operations have conducted symbolic activity with law enforcement, such as marches, and/or shared information at high levels, as by conveying data about homicides and shootings to

2. Anthony A. Braga, "Pulling levers focused deterrence strategies and the prevention of gun homicide." *Journal of Criminal Justice* 36, 2008.

3. Chicago CeaseFire website, http://www.ceasefirechicago.org/.

the leadership of streetworker organizations. Some have rejected even that level of contact, as Wilson and Chermak (2011) document was true in Pittsburgh. Chicago CeaseFire and all or nearly all of those operating in its image have, as a matter of policy, not moved information from streetworkers to law enforcement; not engaged in coordinated street operations with law enforcement; and sent deliberate signals to their target street populations that they do not work with law enforcement.

This may be a critical matter at the tactical level and at a larger and very powerful strategic level. Tactically, it closes off the possibility of a range of important operational options. Even before Operation Ceasefire was framed in Boston, Boston streetworkers did such things as inform the Boston Police Department gang unit when they had information a drive-by would occur. The gang unit would flood the area with officers and thus prevent the drive-by. Part of what is critical about this story, and the operational relationship, is that the streetworkers were operating without being “snitches”—they gave up no names—and the gang unit operated without making arrests. The shooting was prevented in a way that worked for both parties. In extreme instances, streetworkers would make sure that gang members they thought were going to be killed on the street were taken off on parole violations or even new arrests. The streetworkers explicitly believed that that was better than risking a homicide, and the gang unit allowed them to trigger such action without, in turn, insisting that they give gang members up for their own reasons, for example, when the gang unit was conducting its own investigations.

Such relationships and understandings permit very potent tactics. It became a routine part of Boston’s Operation Ceasefire, for example, to have streetworkers carry messages to volatile gangs. When law enforcement intelligence indicated a brewing problem, a beef, or some other indication of imminent violence, law enforcement would ask a streetworker familiar with the gang or gangs involved to go to them and tell them that they were being watched and that violence would trigger a high-level law enforcement response. It was almost always effective. In Cincinnati, streetworkers organize voluntary community “call-ins” of gang members, which law enforcement attends. Law enforcement is making home visits on “impact player” gang members, warning them off violence; streetworkers follow up with the gang members and their families in the following days, reinforcing the message and offering mentoring and access to services. In Providence, RI, streetworkers routinely attend the Providence Police Department’s “CompStat” sessions and coordinate with law enforcement on gang and violent crime responses.

These relationships and operational possibilities sum to what is effectively a different model. It includes the best of traditional, stand-alone streetwork—street presence, dispute resolution, mentoring, service brokerage, relationship building, rumor control, and the like—*plus* a range of new possibilities. It is critical that such partnerships be framed carefully: the wrong perceptions **can** get a streetworker killed. It is clear, however, that such framings can be reached, and that streetwork in that context is fundamentally different than streetwork outside it.

It may also be that the street and community response to such partnerships is itself an important matter, with concrete violence- and crime-control implications. One of the most disturbing aspects of the Wilson and Chermak (2011) evaluation is the finding that violence in the Pittsburgh treatment areas *increased*. This is also sometimes a finding, as has been noted, in the previous generation of evaluations. Wilson and Chermak offer one plausible explanation, which in fact echoes the prior literature, in suggesting that streetwork could inadvertently foster gang cohesion. Another possibility is that the activities of streetworkers, and the structure of the relationship between streetworkers and law enforcement, actively undercut law enforcement and especially police *legitimacy*.

As we have long known, but are increasingly aware, legitimacy has direct links to crime and crime prevention. Where legitimacy is low, crime is more likely (Kane, 2005; Meares, 2009; Tyler, 2006). The troubled minority communities where streetworkers are deployed are, virtually by definition, low-legitimacy neighborhoods. The "outright hostility" between streetworkers and police Wilson and Chermak (2011) document could not have enhanced, and could have degraded, legitimacy. So too, however, might all sorts of signals conveyed by a lack of working relationship between streetworkers and police. Streetworkers who will not work with the police are saying, in a variety of ways, that the police are not legitimate. A violence prevention strategy that is explicitly based on a rejection of the police might well inadvertently undercut legitimacy. Such effects would not be assuaged by high-level working relationships between law enforcement and streetworker organizations when such relationships are, by design, invisible on the street. Opposite effects, however, might be produced by clearly visible working relationships—even bounded ones—between line streetworkers and law enforcement. Given how powerful legitimacy dynamics are being shown to be, such effects might well end up being among the most powerful streetworkers, and streetworker partnerships with others, could produce.

We have, then, an important and growing body of research pointing to the frequent ineffectiveness, or worse, of one model of streetwork. We have counterexamples of effective models of streetwork *in partnership with others* and as an element in a more robust strategy. We have some guidance as to the range of relationships, tactics, and larger dynamics that are likely at work in both instances. We have, I believe, one enormously powerful area to explore and address: how legitimacy is, and could be, addressed through such relationships and activities. If the "too big to fail" projection is to be avoided, streetwork, and our research around streetwork, needs to move in those directions.

References

Ellison, Jesse. 2010, January. Battle of the anti-violence gurus. *Newsweek*. Retrieved from http://www.newsweek.com/2010/01/05/the-battle-of-the-antiviolence-gurus.html.

Engel, Robin S., Marie S. Tillyer, and Nicholas Corsaro. 2011. *Reducing Gang Violence Using Focused Deterrence: Evaluating the Cincinnati Initiative to Reduce Violence (CIRV)*. Unpublished report, City of Cincinnati, Office of the Mayor, Cincinnati, OH.

Kane, Robert J. 2005. Compromised police legitimacy as a predictor of violent crime in structurally disadvantaged communities. *Criminology*, 43: 469–498.

Maggi, L. 1999, June. Tough neighborhood turns around as ex-gang members wise up. *Nation's Cities Weekly*. Retrieved from http://www.thefreelibrary.com/Tough+Neighborhood+Turns+Around+as+Ex-Gang+Members+Wise+Up.-a054902289.

Meares, Tracey. 2009. The legitimacy of police among young African-American men. *Marquette Law Review*, 92: 651.

Papachristos, Andrew V. 2011. Too big to fail: The science and politics of violence prevention. *Criminology & Public Policy*. This issue.

Spergel, Irving A., Susan F. Grossman, and Kwai Ming Wa. 1998. *Evaluation of the Little Village Gang Violence Reduction Project: The First Three Years*. Chicago, IL: University of Chicago.

Tita, George and Andrew V. Papachristos. 2010. The evolution of gang policy: Balancing intervention and suppression. In (Robert J. Chaskin, ed.), *Youth Gangs and Community Intervention: Research, Practice, and Evidence*, vol. 28. New York: Columbia University Press.

Tyler, Tom R. 2006. *Why People Obey the Law*. Princeton, NJ: Princeton University Press.

Wilson, Jeremy M. and Steven Chermak. 2011. Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life. *Criminology & Public Policy*. This issue.

David M. Kennedy is a professor of criminal justice and the director of the Center for Crime Prevention and Control at John Jay College of Criminal Justice.

POLICY ESSAY

COMMUNITY-DRIVEN VIOLENCE REDUCTION PROGRAMS

Too big to fail

The science and politics of violence prevention

Andrew V. Papachristos
Harvard University
The University of Massachusetts—Amherst

Criminologists and politicians walk to the beat of different drummers. The 4-year rhythm of political terms and the 24-hour buzz of the postmodern news cycle disrupt the slow and steady cadence of academic research. Criminologists strive for analytic rigor, sound research design, and objectivity, especially when trying to understand causal effects such as those demanded in most evaluation research. Politicians, however, pledge their allegiance not to the scientific method but to their constituents. Problems need fixing, lives need saving, and most nonacademics need solutions at a pace quicker than the processes of peer review. As a result, whereas the criminologist waits to make claims about causality and program efficacy until field experiments and mathematical models are complete, politicians and other denizens of the "real world" often rely on back-of-the-envelope calculations or simple cross-tabulations made on spreadsheets to discern whether a violence prevention program "works." Linking specific programs to decreases in crime becomes more of an art than a science.

Such divergent worldviews between science and politics have profound implications on violence prevention efforts. "Successful" violence prevention programs typically can secure better funding and resources, not to mention the attention of community leaders, politicians, and the press. Yet who determines whether a violence prevention effort is a success? The academic with his or her regression tables and field experiments? Program administrators and front-line workers with their on-the-ground knowledge and experience? Or the politicians addressing the problems of constituents? And what impact does a programmatic "failure" have on subsequent prevention efforts?

I would like to thank Christopher Wildeman and Anthony Braga for comments on earlier drafts of this essay.
Direct correspondence to Andrew V. Papachristos, Department of Sociology, University of Massachusetts, Amherst. (e-mail: andrewp@soc.umass.edu).

DOI:10.1111/j.1745-9133.2011.00774.x

I maintain that in the absence of consistent evidence and, more importantly, acknowledgment of *scientific* evidence, politics and political rhetoric will lead the charge in determining programmatic success. Many programs—such as the Chicago Project for Violence Prevention (hereafter, simply Chicago CeaseFire)—are often anointed as "successful" before any rigorous scientific evaluation has been commissioned, let alone completed.[1] In the case of Chicago CeaseFire, external evaluation of the program was not completed until nearly a decade after its inception (Skogan et al., 2009). By that time, the program was already dubbed a success by program staff, the media, and politicians and had even expanded within Chicago and into other cities. Although the formal evaluation of Chicago CeaseFire was mainly positive, results from other CeaseFire-like replications in Baltimore, MD, Newark, NJ, and now Pittsburgh, PA, are less promising. Even with such modest and, at times, conflicting scientific evidence, CeaseFire is still packaged as a blueprint for national and even international violence prevention efforts.

The advancement of programmatic and political agendas in the face of limited scientific evidence fundamentally alters the ecology of the violence prevention world. As the title of this essay suggests, I argue that promoting programs as "models" without sufficient evidence can create a situation where violence prevention initiatives become "too big to fail." Massive amounts of political and economic capital have been diverted to programs such as Chicago CeaseFire, perhaps at the expense of smaller and equally successful programs. When such programs work—and at least some evidence suggests that CeaseFire has positive effects—they advance both science and practice, and hopefully, they are associated with significant reductions in violence. But, when they fail—or rather when negative results may develop—no one pays attention. More often than not, negative or null results of juggernaut programs like Chicago CeaseFire do not do what they should do: raise even a modest alarm about the direction of violence prevention efforts. Because programs such as these have been preordained to succeed, failure would spell the end of political and programmatic careers, not to mention it would represent a waste of taxpayer and foundation dollars. More than that, the failure of such programs would leave a massive void in our approach to violence prevention, in large part because we have put all of our eggs into one basket. When programs are too big to fail, good alternatives are simply too difficult to locate or else too obscure to sell to political audiences—without the mantel of "success," people will be less likely to invest. Programs such as these must succeed. And they do, regardless of what good science has to say about it.

The evaluation of Pittsburgh's One Vision One Life program by Wilson and Chermak (2011, this issue) provides an excellent example of a thoughtful scientific analysis of a

1. For the sake of disclosure, I have worked on the evaluations of both Project Safe Neighborhoods (PSN) and CeaseFire in Chicago (see Papachristos, Meares and Fagan 2007; Skogan, Hartnett, Bump and Dubois, 2009).

program with massive political momentum. The One Vision program is partially modeled after Chicago CeaseFire (Skogan et al., 2009). Strikingly, Wilson and Chermak's findings run *counter* to the Chicago CeaseFire evaluation as well as to the political assessments of CeaseFire-like programs more generally: Not only do the authors fail to find evidence of a violence reduction effect, but also their analysis finds that One Vision actually is associated with *increases* in violence. Wilson and Chermak provide several important discussion points about how their findings build on and call into question CeaseFire-style programs, all with an eye toward improving future strategies. Sadly, although the analysis is thorough and convincing, I suspect that it will not even cause a hiccup in the political machinery promoting CeaseFire-like programs.

In the remainder of this essay, I extrapolate a bit more on the science and politics of Chicago CeaseFire. Then, I focus on the Chicago evaluation results as well as on the results from replications in Baltimore, Newark, and Pittsburgh. I conclude by discussing the implications such a debate has for the idea of "evidence-based practice."

The Science and Politics of Chicago CeaseFire

Wilson and Chermak (2011) analyze a replication of one of the biggest and most politically vibrant violence programs today—a version of the Chicago CeaseFire program.[2] At its core, the Chicago CeaseFire initiative is based on an old approach commonly referred to as "street work": the use of outreach workers to work directly with gangs, gang members, and troubled youth to provide direct services and mediate disputes before they become violent.[3] Like previous street work efforts, CeaseFire relies not necessarily on professionally trained social workers but on street-oriented individuals (in particular, ex-gang members and ex-offenders) who have local knowledge of the neighborhood and gangs targeted for intervention. To be sure, CeaseFire has made significant advancements in the street work approach. First, CeaseFire reframed violence prevention in a public health framework; specifically, the program believes that changing attitudes toward gun violence requires changing norms and behaviors in the same way other public health efforts have tried to alter behaviors like cigarette smoking, drunk driving, and risky sexual activity.[4] Second, CeaseFire has expanded previous street work models by differentiating more fully the roles and functions of its staff. For example, the current CeaseFire model incorporates "violence interrupters," whose job it is to mediate gang/neighborhood disputes, and more traditional

2. It is important to note that the Chicago CeaseFire program is operationally and organizational distinct from the Operation Ceasefire that was part of the Boston "pulling levers" strategies, although several significant similarities exist, including the use of street workers.

3. The origins of street work programs such as these can be found as early as World War II when sociologists Clifford Shaw and Henry McKay initiated the Chicago Area Project. See Tita and Papachristos (2010) for a recent review of street work programs (also see Klein, 1971; Spergel, 1968).

4. The "crime as a public health problem" is also a much older idea that gets revisited in the CeaseFire model (see Hemenway, 2006, for a review).

caseworkers, who are engaged in direct service provision. Third, CeaseFire's street work efforts are integrated into larger community-level activism such as media campaigns, rallies, protests, town hall meetings, and so on. In a sense, CeaseFire has become a *social movement*, not simply an intervention program.

CeaseFire started in 1999 in several of Chicago's highest crime neighborhoods. Since its inception, researchers on the CeaseFire staff have monitored neighborhood-level crime indices, boasting of dramatic programmatic effects as early as the first year after implementation (CeaseFire, 2011). CeaseFire expanded rapidly based, in part, on such internal assessments and, in part, on political posturing and rhetoric: By 2005, the program expanded into approximately 12 operational sites in Chicago (Skogan et al., 2009: 2–19). The program became so politically popular that in June 2005, First Lady Laura Bush visited with street workers and lauded CeaseFire as a model of how to help at-risk youth. CeaseFire's activities and success has been spotlighted in dozens of media outlets—including an article by Alex Kotlowitz (2008) in the *New York Times Magazine*—and recently, it was the subject of a documentary movie (also by Kotlowitz) that premiered at the 2011 Sundance Film Festival.

External evaluation of CeaseFire did not commence until 2005 and was not complete until 2008—nearly a decade after the program began. Like many other programs before it, the evaluation of CeaseFire and the fact that the program's initial design did not include any formal experimental design components forced the National Institute of Justice (NIJ)–funded research team to try and reconstruct "comparison" neighborhoods to match with "treatment" neighborhoods by statistically matching on similar sociodemographic and crime characteristics.[5] Therefore, any true experimental design was stymied by the politics of funding, state legislation, and the nonrandom assignment of treatment clients and gangs—a common casualty, but not necessarily a fatal one, in the evaluation of criminal justice and violence prevention programs (Skogan et al., 2009: 2–20). The results of this external evaluation were by and large *positive* and have since been published in a NIJ report (Skogan et al., 2009). However, to the best of my knowledge, the Chicago CeaseFire findings have not been subject to peer-review publication (outside of the NIJ report) nor have they been replicated successfully.[6]

5. The lack of research design and external evaluation in the original CeaseFire model is surprising given the founder's medical and public health background as well as the housing of the program in a school of public health—a field that stresses the need for randomized trials and experiments in determining program efficacy.

6. An early analysis of PSN in Chicago conducted by my colleagues and myself (Papachristos et al., 2007) included a brief analysis of the presence of CeaseFire in several neighborhoods; using a quasi-experimental design and propensity score matching to create statistical counterfactual groups, we found no discernable effects of CeaseFire above and beyond PSN. Although our analysis was *not* intended to evaluate CeaseFire specifically, it does suggest that perhaps some findings of the NIJ report might be sensitive to methodology, timing, and the selection of comparison groups.

Is CeaseFire a "Success"?

Skogan et al.'s (2009) evaluation of Chicago CeaseFire represents a detailed and massive evaluation of both outcomes and outputs. In fact, one of the greatest contributions of this evaluation is the documentation and assessment of CeaseFire activities, clients, and logic models. The results from a longitudinal analysis of crime trends in CeaseFire neighborhoods and comparison areas find significant drops in shootings with decreases ranging from 16% to 34%. Furthermore, Skogan et al. (2009) also (a) considered the concurrent occurrence of the PSN program that was present in some (but not all) of the CeaseFire neighborhoods, (b) discussed the limitations of the models and data, and (c) described the difficulties of politics and funding in determining the selection of treatment neighborhoods. Overall, Skogan et al. worked with the available data, and their evaluation represents a good faith effort at assessing programmatic effects.

The media (not the researchers) called these Chicago results "beyond dispute" ("Seeking safe passage," 2009), thus providing the impetus for expansion and replication. However, a closer examination of the 238-page report and 212-page technical appendix may perhaps leave some room to question the political designation of these findings as "indisputable." An interesting aspect of the report by Skogan et al. (2009) is the inclusion of two independent analyses by Richard Block (technical appendix, pp. B1–B34), who analyzed the changes in shooting "hot-spots" in CeaseFire and treatment districts, and myself, who analyzed changes in network patterns of gang homicide (Skogan et al., 2009: technical appendix, pp. C1–C35). Both of these efforts were conducted independently from the main evaluation, and each of us used our own data. In his analysis, Block reviewed seven CeaseFire program neighborhoods and found positive results in three CeaseFire areas, one "probably" positive result, and three inconclusive or null results. Strikingly, my findings mirrored Block's almost exactly. Of the eight CeaseFire locations I analyzed, I found positive programmatic effects in two target areas (which were identical to two of Block's positive CeaseFire areas) and potentially a third positive result. The remaining five areas exhibited either no effect or the results were inconclusive.

Considering these full results, one might conclude more accurately that the Chicago program was successful in some of the program's targeted areas. In fact, such an interpretation would be more consistent with the findings of other CeaseFire replications, including those in Baltimore, Newark, and Pittsburgh.

In 2005—even before the NIJ-funded evaluation was underway on Chicago CeaseFire—a replication was already beginning in Newark (see Boyle, Lanterman, Pascarella and Cheng, 2010). Whereas the Chicago evaluation of CeaseFire's influence on violence relied on data provided mainly by the police department, the Newark study analyzed changes in gunshot wounds reported at trauma centers. The analyses reported by Boyle et al. found no evidence of a statistically significant decrease in gunshot injuries in the CeaseFire areas compared with similar neighborhoods.

With the assistance of a $1.6 million grant from the U.S. Department of Justice and technical assistance from the Chicago CeaseFire team, another CeaseFire replication was implemented in Baltimore in 2007—once again, before the Chicago external evaluation was complete (Webster, Vernick and Mendel, 2009). The results from the Baltimore CeaseFire model also produce mixed findings, but the overall findings lend little support of a convincing programmatic effect. On the positive side, one intervention neighborhood experienced a significant decrease in homicides postintervention. However, this finding is somewhat offset by the fact that two target neighborhoods experienced *increases* in nonfatal gunshot injuries (Webster et al., 2009, especially Tables 6 and 7).[7]

Taken in this context, the results presented by Wilson and Chermak (2011) are perhaps not surprising either analytically or politically. Analytically, the results coincide with those of Newark and Baltimore but diverge from the way the Chicago study has been portrayed. Politically, Wilson and Chermak are taking on a behemoth that was self-christened as a success and validated by politicians and the media well before the positive results of external evaluators. More than that, the Pittsburgh study illustrates how, once again, cities were willing to partake in a program that was deemed a "success" without really considering the scientific evidence.

Is CeaseFire Too Big To Fail? The Implications for Evidence-based Practice

Thus far we have observed that (a) Chicago CeaseFire was anointed a success well before external evaluation; (b) the NIJ evaluation provides some support for programmatic success, but a more detailed examination of the technical reports might suggest that such a positive assessment is not infallible; (c) subsequent replications of the CeaseFire model have by and large produced negative results; and (d) cities continue to look toward CeaseFire despite the aforementioned results.

The continued proliferation of CeaseFire-like programs in the face of largely negative replication results may seem scientifically odd. How many other public health initiatives or clinical trials would continue to receive funding in the face of similar results? Politically, however, this is an indication of the "too big to fail" phenomena: Programs draw a lot of attention and resources without consistent empirical evidence, and consequentially, stakeholders have a lot riding on their perceived success. CeaseFire programs like the ones in Baltimore, Newark, and Pittsburgh invested in a "successful" model and, therefore, potentially had much more to lose in the face of failure than if they had invested in a smaller program without the mantel of success. Negative results are dismissed all together or else flatly ignored by the media and, no doubt, by program administrators. When the First Lady and Alex Kotlowitz say a program is a success, who is a criminologist to say otherwise?

7. This interim report showed some significant changes in attitudes toward gun violence, although these did not seem to translate into changes in gun use or aggregate patterns of gun violence in intervention areas.

To be sure, public safety cannot always wait for academic findings, and often, practitioners will accept "best guesses" as evidence of programmatic success. If held to their word, politicians must be better equipped to engage and respect scientific findings. For our part, we criminologists must be able to work with practitioners and politicians in an intelligible and jargon-free manner outside of our thought experiments and counterfactuals (see Skogan, 2010). Regardless, as scientific evidence begins to mount, a reassessment of our violence prevention strategies is clearly in order. In the case of CeaseFire, the evidence demonstrates some positive results in Chicago followed by replications with null effects.[8]

How *should* we interpret these results? And, what affect should they have on subsequent proliferation of the CeaseFire model? "Too big to fail" would suggest that we cannot do anything—that stopping the momentum of massive programs would derail violence prevention efforts and leave a void in programmatic efforts. A more balanced approach suggests that both practitioners and academics would reassess our efforts to improve the focus and results of violence prevention strategies.

The good news is that the violence prevention world is perhaps ready to make serious progress on both scientific and political fronts. Politics, especially since the Obama administration, has experienced considerable advancements in the application of "evidence-based practices"—a preferential use of programs that are demonstrated to have positive effects on reducing violence (Robinson, 2010). The mounting evidence on CeaseFire-like programs provides a moment for us to understand *how* the CeaseFire model holds under the criteria of evidence-based practice. For instance, compare the idea and use of the idea of "evidence-based practice" as applied to Chicago CeaseFire with the evidence-based ethos employed by place-based policing strategies (Braga, Papachristos and Hureau, 2011). Place-based policing strategies have been and continue to be subject to rigorous evaluation. More importantly, the results of such evaluation are systemized and subjected to additional scrutiny by the scholars and policing experiments. For example, a recently updated Campbell Collaboration Systematic Review and meta-analysis of 18 randomized controlled trials and quasi-experiments reported significant crime control gains associated with hot spots policing programs (Braga et al., 2011). Negative results are not discarded but are gleaned for how to proceed in future policing interventions.

If we truly wish to subject the CeaseFire model of violence prevention to the rigors of evidence-based practice, then, like the approach taken in place-based policing, we must learn from the positive *and* negative evaluations alike. One area that comes to mind for expanding the use of evidence-based practice as applied to CeaseFire-like programs is how to measure what actually is going on. To date, *all* the evaluations examined only the changes in aggregate rates of crime or violence—that is, crime either increased or decreased after

8. The Minneapolis Domestic Violence Experiment follows a similar trajectory in which initial success was followed by failures in replication. See Sherman (1992) for a summary of this policy issue.

the CeaseFire intervention. We have very little statistical assessment of *group effects* directly linking reductions in mediated conflicts to declines in violence or of *individual effects* linking the participation of CeaseFire clients to increased participation in prosocial activities and to decreased criminal activities.

A particularly important aspect in need of serious consideration is the way we evaluate street work. The press and program administrators most often attribute such positive effects of CeaseFire to the work of outreach workers or "violence interrupters" (indeed, the documentary on Chicago CeaseFire spotlights these street workers). Yet, *none* of the evaluations of CeaseFire-like programs have measured the "treatment" given by street workers—we simply have not devised a way to measure what it is street workers do, how often they do it, and how such efforts are linked to changes in violence.[9] If street work is a model of violence prevention that we should continue to pursue—and I think it *is*—then we need to understand its effects more clearly. Once we can pinpoint the effects of some of CeaseFire's moving parts, we can then potentially assess the model more effectively as well as assess which aspects of the model might be breaking down in replication. In sum, we should seize the moment to make our programs better, not simply bigger. Evidence-based practices can help us in this quest but only if we have clear expectations of what constitutes evidence and what constitutes success.

References

Boyle, Douglas J., Jennifer L. Lanterman, Joseph E. Pascarella, and Chia Cheng. 2010. *Impact of Newark's Operation CeaseFire*. Newark: Violence Institute of New Jersey, University of Medicine and Dentistry of New Jersey.

Braga, Anthony A., Andrew V. Papachristos, and David Hureau. 2011. *Effects of Hot Spots Policing on Crime* (revised systematic review report submitted to Campbell Collaboration Crime and Justice Group; manuscript available upon request from authors).

CeaseFire. 2011. Organizational website. Retrieved July 9, 2011 from ceasefirechicago.org/results.shtml.

Hemenway, David. 2006. *Private Guns, Public Health*. Ann Arbor: The University of Michigan Press.

Kennedy, David. M. 1997. Pulling levers: Chronic offenders, high-crime settings, and a theory of prevention. *Valparaiso University Law Review*, 31: 449–484.

Klein, Malcolm. 1971. *Street Gangs and Street Workers*. Englewood Cliffs, NJ: Prentice Hall.

9. For example, in my own report in the CeaseFire evaluation, I tracked changes in the exchange of homicides between gangs. However, because of the lack of data on actual interventions and mediations of street workers, I could not link street worker efforts explicitly to the actions of specific gangs or specific violent exchanges (see Skogan et al., 2009, technical Appendix, pp. C-1–C-35). That is, although I reported that retaliatory shootings decreased in a particular area or among a particular gang, I could not identify whether those were the actual gangs that were clients of outreach workers.

Kotlowitz, Alex. 2008. Blocking the transmission of violence. *New York Times Magazine*. May 4.

Papachristos, Andrew V., Tracey L. Meares, and Jeffrey A. Fagan. 2007. Attention felons: Evaluating Project Safe Neighborhoods in Chicago. *Journal of Empirical Legal Research*, 4: 223–272.

Robinson, Laurie. 2010. The federal role in promoting evidence based practice. *The Academy of Experimental Criminology Newsletter*, April.

Seeking Safe Passage: Right under our noses? 2009. *Chicago Tribune*. October 18.

Sherman, Larry. 1992. *Policing Domestic Violence*. New York: The Free Press.

Skogan, Wesley G., Susan M. Hartnett, Natalie Bump, and Jill Dubois. 2009. *Evaluation of CeaseFire-Chicago*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Skogan, Wesley. 2010. The challenge of timeliness and utility in research and evaluation. In (John Koflas, Natalie Kroovand Hipple, and Edmund McGarrell, eds.), *The New Criminal Justice: American Communities and the Changing World of Crime Control*. New York: Routledge.

Spergel, Irving A. 1968. *Street Gang Work: Theory and Practice*. Reading, MA: Addison-Wesley.

Tita, George and Andrew V. Papachristos 2010. The evolution of gang policy: Balancing intervention and suppression. In (Robert J. Chaskin, ed.), *Youth Gangs and Community Intervention: Research, Practice, and Evidence*. New York: Columbia University Press.

Webster, Daniel W., Jon S. Vernick, and Jennifer Mendel. 2009. *Interim Evaluation of Baltimore's Safe Streets Programs*. Baltimore, MD: John Hopkins Bloomberg School of Public Health, Center for the Prevention of Youth Violence.

Wilson, Jeremy M. and Steven Chermak. 2011. Community-driven violence reduction programs: Examining Pittsburgh's One Vision One Life. *Criminology & Public Policy*. This issue.

Andrew V. Papachristos is currently a Robert Wood Johnson Health and Society Scholar at Harvard University and an assistant professor of sociology at the University of Massachusetts, Amherst. His research examines neighborhood social organization, street gangs, interpersonal violence, and social networks.

RACIAL DISPARITY IN WAKE OF THE *BOOKER/FANFAN* DECISION

Racial disparity under the federal sentencing guidelines pre- and post-*Booker*

Lessons not learned from research on the death penalty

Raymond Paternoster
University of Maryland

The article by Ulmer, Light, and Kramer (2011, this issue) and the corresponding policy essays by Albonetti (2011, this issue), Engen (2011, this issue), Scott (2011, this issue), and Spohn (2011, this issue) in this section of *Criminology & Public Policy* examine the effect of several U.S. Supreme Court decisions on sentencing disparity under the federal sentencing guidelines. In 1984, Congress enacted the Sentencing Reform Act, which created the United States Sentencing Commission (USSC). One motivation for the Act was the belief that too much discretion was provided to judges in the federal system and that as a result there was great disparity in sentencing White and minority defendants.[1] The USSC was given the task of developing and implementing sentencing guidelines for federal judges as a means of controlling judicial discretion, with the goal of achieving greater "uniformity" in sentencing. Prior to the guidelines, federal judges had virtually unlimited discretion to impose sentences so long as they met broad statutory requirements. Under the guidelines, however, the judge had to calculate a defendant's criminal history and offense level score under strict rules, the result of which was the placement of the defendant on a sentencing grid. The sentence found in the grid was the presumptive sentence, and although departures could be made, the reason for the departure had to be given either in open court or in a written judicial opinion. Furthermore, to monitor and ensure compliance with the guidelines, the Reform Act also provided for appellate review of any departures from the

I am very grateful to Shawn Bushway, Tom Loughran, and Bobby Brame for reading a draft or two of this and for providing helpful conversation. Direct correspondence to Raymond Paternoster, Department of Criminology & Criminal Justice, University of Maryland, College Park, MD 20742–8235 (e-mail: rpaternoster@crim.umd.edu).

1. Other, more conservative, critics also argued that there was too much discretion given to federal judges, but the unwanted product it produced was leniency or "softness" in sentences rather than disparity.

DOI:10.1111/j.1745-9133.2011.00778.x

guidelines, with sentences imposed within the calculated guideline range reviewable only under a "clear error" standard (a *de novo* standard of appellate review). In sum, recognizing that "unwarranted disparity caused by broad judicial discretion is the ill that the Sentencing Reform Act seeks to cure,"[2] Congress attempted to reduce discretion and disparity by trying to make federal sentencing practices comply with strict legal formula. Mandatory guidelines and an appellate review that strictly scrutinizes for compliance with those standards seem like a pretty good way to reduce disparity in sentencing, a suspicion supported by research that includes previous empirical work by the authors of the policy essays (Albonetti, 2011; Engen, 2011; Scott, 2011; Spohn, 2011) in this section (Bushway and Forst, 2011; Bushway and Piehl, 2001, 2007; Engen and Gainey, 2000; Frase, 2005; Reitz, 2005; Spohn, 2000; Tonry, 1996; Ulmer, 2000).[3,4]

How do we know whether mandatory guidelines were successful in reducing sentencing disparity in federal courts? According to a 2010 study conducted by the USSC, a study that serves as the backdrop for the article by Ulmer et al. (2011) in this section, when the mandatory nature of the guidelines was removed and made advisory by the Supreme Court in *United States v. Booker* (2005) and when the review standards were relaxed and made more deferential by *Rita v. United States* (2007) and *Gall v. United States* (2007), sentencing disparity dramatically increased. This finding was not surprising to me for two reasons.

First, the guidelines that were created sharply reduced judicial discretion at sentencing by providing a reasonably narrow range within which the judge's sentence was supposed to fit, and any departure from the presumptive sentence was subject to strict appellate review. It was clear that judges (although importantly not prosecutors) had much less discretion under mandatory guidelines and much more sentencing discretion after *Booker*. We should not be too terribly surprised, then, that when sentencing guidelines that were carefully crafted to reduce judges' sentencing discretion, that were mandatorily imposed, and that were ultimately subject to a rigorous review are now only advisory and the review more deferential that greater disparity in sentencing would result.

Second, as someone who has done state capital punishment but not federal sentencing research, I was not at all surprised by the USSC's findings of greater racial disparity in sentencing after *Booker*. What the Court did in *Booker, Rita*, and *Gall* was to insist that federal sentences be both uniform, that is, consistent across persons, and reflect the unique

2. Stephen S. Trott, Letter to Hon. William W. Wilkins (April 7, 1987); the letter is reprinted at *Federal Sentencing Reporter,* 8: 196 (1995).

3. In their recent review of this area, Bushway and Piehl (2007: 464–465) noted that, "Studies of four jurisdictions using this type of empirical model have found little evidence of racial disparity on the part of judges in strict guideline systems . . . but evidence of substantial disparity in the voluntary guideline systems . . . . These results are largely consistent with the general claim that presumptive sentencing guidelines reduce judicial discretion and racial disparity."

4. To be completely fair, in her excellent review of the literature on sentencing guidelines Spohn (2000) is a little more ambiguous as to whether or not they reduce disparity.

culpability of individual offenders. Although sentencing guidelines were supposed to create consistency and greater uniformity in federal sentences, *Gall* admonished that reviewing courts cannot simply assume that the guideline ranges are reasonable but must "make an *individualized assessment* based on the facts presented" (pp. 596–597).The problem, of course, is that sentences can become more uniform and consistent only at the expense of individual uniqueness and that sentences can be tailored to an individual's unique culpability only by reducing consistency. The requirement that both principles must be honored is precisely what the Supreme Court did with absolutely no success in *Gregg v. Georgia* (1976) and its companion cases. In *Gregg* (and *Proffitt v. Florida* [1976] and *Jurek v. Texas* [1976]), the Court approved guided discretion statutes because the enumeration of statutory aggravating and mitigating circumstances together with appellate review promised to make death sentences more even handed and less capricious. At the same time, they pointedly struck down mandatory death sentencing schemes because they treated persons "not as uniquely individual human beings, but as members of a faceless, undifferentiated mass" (*Woodson v. North Carolina* [1976: 281]). As a result, death sentences had to be both consistent and individualized. The consequence of trying to conform to both consistency and individualized sentencing was that death sentences under post-*Gregg* procedural reforms were as disparate as those under the denounced standardless system condemned in *Furman v. Georgia* (1972). The comparable Scylla and Charybdis that federal judges have to maneuver through post-*Booker* is the need to adhere simultaneously to the principle of consistency in sentencing in order to ensure uniformity in sentences and the need to tailor the sentence to the unique culpability of individual offenders in the review of such sentences. The USSC's 2010 report simply confirms what capital punishment researchers had found, that it is difficult in practice to reconcile these two principles and that increased sentencing disparity is the likely result.

The USSC's 2010 report showing greater racial disparity in federal sentences after *Booker* and *Gall* seems to me, then, to be both believable and predictable. In their article that anchors this issue, Ulmer et al. (2011) offer "an alternative analysis to the USSC's 2010 report" based on "different analytical and modeling choices." In their article, they show that racial disparity is more prominent in the prison/probation decision than in the sentence length decision, greater among immigration offenses than among nonimmigration crimes, and greater when the post-*Booker* period is compared with the sentencing regime covered under the PROTECT Act (2003) than the immediate pre-*Booker* period. As the results of the two studies, by USSC and Ulmer et al., seem to conflict with one another, with the perhaps impression that one must be "right" and the other "wrong," they bear brief discussion.

First, what the USSC (2010) did in it its analysis, although involving different "methodological choices" than Ulmer et al. (2011), seems eminently reasonable and their results are due a great deal of respect. Consistent with sound regression discontinuity designs, the USSC analysis compared post-*Booker* racial disparity with that which existed in the most proximate time period before *Booker*. Furthermore, in modeling the prison/probation decision along with the length of sentence decision, the USSC was assuming that the two decisions were related to one another and to that extent they deserved to be treated jointly. Ulmer et al.'s different methodological choices included a decision to compare different prior time periods with the post-*Booker* period than that used by the USSC. They find that the amount of disparity post-*Booker* depends on which prior period it is compared with. This is interesting, although not surprising, and although it may put bounds on what might be happening in the postguideline period, it does not invalidate the USSC finding that disparity has increased relative to the most recent period. Ulmer et al. also made the choice of separately analyzing subsets of the population that the USSC had, excluding immigration cases from one of their analyses. What they found was greater disparity among immigration than among nonimmigration cases, confirming the frequent finding that a general "treatment" effect may not hold for all strata within a population. It seems to me that what they have done here, however, is to elaborate and extend the USSC's findings rather than raise a suspicion about them.

Ulmer et al. (2011) also preferred to conduct separate analyses of the prison/probation and the length of sentence decision, finding greater disparity in the former than in the latter. Without making any judgment about the USSC's (2010) choice to analyze both in the same equation, it is worth thinking about the implications of Ulmer et al.'s strategy to conduct a separate analysis. Ulmer et al. are asserting that judges first make a decision whether to sentence a defendant to prison or probation, and then completely independent of this decision, they decide how much time the person should be sentenced for. This assertion seems to me to be a very strong claim about how sentencing is actually done by judges. Whatever disagreements one may have with the USSC's decision to treat the two decisions, the model of sentencing presumed by Ulmer et al. should be completely understood.

In sum, what does an outsider make of the USSC's (2010) findings and those of Ulmer et al. (2011)? Because of my experience with the U.S. Supreme Court's attempt to regulate state death sentences, I was not surprised by the fact that the USSC found greater disparity after the *Booker* and *Gall* decisions. Although I now have a more nuanced understanding[5] of this thanks to Ulmer et al.'s extensive analyses, nothing in their report would lead me to change my opinion that a consequence of moving from mandatory to advisory federal sentencing guidelines with a more relaxed standard of review is greater racial disparity in sentencing. What both sets of studies would seem to agree on is that when the federal sentencing regime became advisory, disparity in sentencing increased, with a disagreement mainly about the magnitude of that increase. But this is not the whole story.

5. Spohn (2011) concluded that what we have here with Ulmer et al. and the USSC's report are "two major studies [that] have reached different conclusions." I do not necessarily see it quite that way. I see two analytically/methodologically very different studies that have come to slightly different conclusions that complement one another.

Many practitioners and scholars did not like the mandatory federal sentencing standards because they were too harsh and retributive, and they may fear that empirical findings of greater disparity under an advisory regime such as those reported by the USSC may hasten their return. That battle will have to be fought on its own ground, however, and not on the grounds that advisory standards will not result in greater sentencing disparity.

So then, "Why deny the obvious child?" Although I was not surprised that the USSC's 2010 report found an increase in disparity in federal sentences after *Booker* and *Gall*, I was surprised by the denial that permeates the policy essays that join Ulmer et al.'s (2011) article. I have noted that these policy essay authors (Albonetti, 2011; Engen, 2011; Scott, 2011; Spohn, 2011) are part of a sentencing literature in criminology that has concluded that sentencing guidelines, when they are mandatory and strictly monitored, have reduced disparity. Based on their own previous excellent work on guidelines and the observations they presented in their essays, it is not clear why these authors would be surprised by, or why they would deny, the USSC's findings. It is difficult to tell whether this reaction is based on genuine doubt that *Booker* and *Gall* resulted in greater disparity, or more on the aforementioned fear that these findings might be used to bring back some variation of the old mandatory/strict review regime.

For example, both Albonetti's (2011) and Spohn's (2011) policy essays provide excellent historical context to the creation of the U.S. Sentencing Commission, and like any good history, they help us understand the present. They show that there was great dissatisfaction with what was perceived to be an unduly large amount of disparity in federal sentences, and consensus that this disparity resulted from two features of the existing sentencing regime: (a) judges had extensive sentencing discretion, and (b) sentences were not generally reviewable. From another direction, however, there was equal angst at the time that federal sentences, particularly with respect to drug and gun crimes, were too lenient. As Albonetti correctly points out, the 1984 Sentencing Reform Act was not the only sentencing-related piece of legislation passed by Congress at that time; there was also the Anti-Drug Abuse Act of 1986, the Comprehensive Crime Control Act of 1984, and the Omnibus Anti-Drug Act of 1988, all of which were expressions of a desire to get tougher on crime.

Both the desire to minimize sentencing disparity and the desire to punish more certainly and harshly could be satisfied by a sentencing regime that provided judges with fairly narrow sentencing ranges that increased when the offense was more serious and the offender's criminal history was more extensive, and which required appellate review when there were departures from this—ta da . . . mandatory sentencing guidelines with limited ability for a judge to go outside those guidelines. One agreed upon consequence of the 1984 Sentencing Reform Act is exactly what Albonetti (2011) noted in her essay, it "virtually transformed sentencing practices. Policy priorities were aimed at severely limiting judicial discretion in an attempt to eliminate unwarranted sentence disparity." In her essay, Spohn (2011) made precisely the same observation, that the Sentencing Reform Act was born from both liberal and conservative political sentiments and that the mandatory regime with appellate review devised and implemented by the USSC was intended to curtail disparity sharply. She says, "[i]t is also clear that the guidelines were designed to eliminate discrimination based on legally irrelevant characteristics of the offender. . . . the potential for racial and ethnic discrimination was limited by the fact that the guidelines were mandatory and that judge-initiated departures were regulated closely." The first point I want to make about this is that because the mandatory guidelines were created to "curtail disparity sharply" and that they "virtually transformed sentencing practices," why should anyone be surprised that when they were both made advisory and less stringently scrutinized after *Booker* and *Gall* that greater disparity would result? The second point is that as liberals, who wanted to get rid of disparity, entered into a Faustian bargain with conservatives, who wanted to get tougher on crime, why would anyone be surprised by the fact that mandatory guidelines based on offense seriousness and offender criminal history scores tended to be punitive?

It would seem to an outsider like me that given the USSC's (2010) findings about greater disparity immediately after *Booker*, and the fact that all parties seem to agree that the least amount of sentencing disparity was found during the PROTECT regime when sentencing guidelines were both mandatory and stringently reviewed, that if one really wanted to minimize discretion, they would want to return to that kind of regime. However, as we will see in a moment, *no one* among the current set of authors wants to do that, primarily it seems because they do not wish to return to a regime that they believe is unnecessarily harsh and severe. Before getting to that issue, however, a few more observations need to be made.

I found myself in agreement with certain of Engen's (2011) conclusions. His policy essay is very much a painstaking dissection of the different analytical and methodological choices that the authors of the two studies made (Ulmer et al., 2011; USSC, 2010), and he concluded that "certain methodological choices probably affected the conclusions" reached by both studies. I agree. In addition, after reading Engen's comments, I was convinced that there was likely *more* disparity in the post-*Booker* period than what Ulmer et al. reported. He raised the issue that the decision of Ulmer et al. to include in their model both the defendant's criminal history and the presumptive sentence might be suppressing a larger race effect, a possibility that is often discussed in this literature and that even Ulmer et al. acknowledged: "Furthermore, the criminal records of Black males may themselves be the product of discriminatory processes."

An additional conclusion reached by Engen (2011) that bears repeating is that "[t]he general unavailability of data on charging and plea-bargaining remains . . . *the greatest challenge to the validity of sentencing research* " (emphasis added). As all parties to this discussion acknowledge, what is lacking from both the USSC's (2010) and Ulmer et al.'s (2011) analysis is information about prosecutors' decisions. Another lesson not learned from capital punishment research was that the lion's share of the disparity in how criminal cases are handled often occurs long before sentencing takes place—in the hands of the prosecutor. Countless empirical studies of how states impose the death penalty have shown that there is substantial disparity at the decision of the prosecutor to charge a crime as capital, and there is minimal disparity at sentencing. State capital punishment research has shown, therefore, that most of the "action" is at the level of the charging decision. Although there seems to be general agreement about the need to consider prosecutorial decision making in federal sentencing research (see Bushway and Piehl, 2007), there seems to be no urgency in doing something about it. There is instead great lamentation about the lack of this data and its implications, and an internecine quibble over the results of analyses of conviction data sets. I would suggest that because both parties to this issue (USSC and Ulmer et al.) have only conviction data, the argument here is about a small part of the story of disparity in federal sentencing either with or without mandatory sentencing guidelines.

Furthermore, it is not as if either of the two sets of authors can blithely dismiss this omission of prosecutorial information by making the claim that they are only speaking to disparity in *sentencing* as (again, Engen [2011] nicely points this out) discrimination at earlier points in the system distorts the types of cases that are considered at sentencing. It seems a bit odd to have such sharp arguments and disagreements over the results of sentencing studies that fail to consider prosecutorial charging decisions. Hopefully, at some point, the hand wringing and the "weeping, wailing, and gnashing of teeth" will stop and something will be done to collect good prosecutorial decision-making data at the federal level.

Scott's (2011) policy essay is somewhat of the dissident in this series of essays because he chose to focus mainly on the common ground between the USSC (2010) report and the Ulmer et al. (2011) article rather than on the differences, but I think a reminder of the great deal of common ground is very well served: "in several of their key findings, the Commission's research and the new analysis by Ulmer et al. (2011) reach similar results." The common ground identified by Scott is that in both studies: (a) the imprisonment decision disadvantages Black males in comparison with White males, (b) evidence of racial disparity under the mandatory guidelines before 2003 was fragile, and (c) race disparity against Black males was at its lowest level "when the Guidelines were at their most mandatory and inflexible" and when there was strict appellate review, under the PROTECT Act regime. This last point is very interesting because how to respond to it was something that all four policy essays had in common.

As I suggested a few paragraphs earlier, if one is interested in driving sentencing disparity down to its minimal level, the evidence would seem to indicate that you have mandatory sentencing guidelines and a very stringent appellate review—the regime under the PROTECT Act. It makes complete sense. However, *none* of the authors of these essays settles on that conclusion. Albonetti (2011), for example, after acknowledging that mandatory guidelines and strict review, "virtually transformed sentencing practices" nevertheless refuses to believe that making them advisory and the review more deferential would result in substantially more disparity and that "[t]here is no need to institute statutory remedies for sentences that do not greatly differ from those imposed under pre-*Booker* mandatory guidelines structure." Essentially, she dismisses the USSC (2011) and accepts Ulmer et al. (2011), thinking that judges essentially behave the same pre and post-*Booker,* even though she acknowledged that mandatory guidelines "transformed sentencing practices." Despite making virtually a similar argument about the power of mandatory guidelines and strict review in sharply reducing disparity, Spohn (2011) too does not argue for their return. We start to get a sense of what the real opposition is in Engen's (2011) essay where he refers to the fact that "many observers, including federal judges, believe [the guidelines] are unjust." He goes on to wonder what should be made of strict consistency or disparity in sentencing under "laws [that] are indeed unfair." In the very next paragraph, Oz's curtain is lifted when he implies that greater disparity may not be such a bad thing if "with increased discretion, judges will hand down sentences that most observers agree are more 'appropriate,' on average, than if they had followed the guidelines closely, and that all races, ethnicities, and genders will benefit from this discretion." More appropriate in what sense? Because sentences are less severe when guidelines are advisory? Is the real problem with mandatory guidelines and stringent review, then, not disparity but substantive injustice as Engen argues? If it is, then (a) that is not an empirical question that can be addressed by scientific studies, and (b) perhaps we should bring it out into the open and address it directly. Scott (2011) also saw a return to the regime under the PROTECT Act as a logical policy alternative, noting that it was the policy choice recommended by Justice Stevens in his *Booker* dissent, and Scott provides sound reasons as to why this route makes sense if you want to diminish disparity. Like the others, however, Scott dismisses this policy remedy for myriad reasons.

An outsider to this debate about lesser or greater disparity after the *Booker* decision gets the sense that at least to some degree we are not really talking about how much more or less disparity exists under mandatory versus advisory sentencing guidelines. The uniformity and resoluteness with which a return to mandatory guidelines and strict review are rejected by the policy essay authors in this section, and the only slightly veiled appreciation for greater substantive justice under advisory guidelines that permits greater discretion to judges, both lead me to suspect that at least part of the issue among federal sentencing scholars is that the mandatory guidelines were too severe for their taste and they fear any attempt to return to them. Research findings such as those in the USSC 2010 report, it is suspected, would naturally lead in that direction. In my language, given a choice between the greater consistency of *Gregg* or the greater discretion and ability to individualize sentences of *Woodson*, they definitely prefer the latter. That, however, seems a preference that empirical research cannot address.

References

Albonetti, Celesta A. 2011. Judicial discretion in federal sentencing: An intersection of policy priorities and law. *Criminology & Public Policy*. This issue.

Bushway, Shawn and Brian Forst. 2011. Studying discretion in sentencing. Working paper.

Bushway, Shawn D. and Anne Morrison Piehl. 2001. Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law & Society Review*, 35: 733–764.

Bushway, Shawn D. and Anne Morrison Piehl. 2007. Social science research and the legal threat to presumptive sentencing guidelines. *Criminology & Public Policy*, 6: 461–482.

Engen, Rodney L. 2011. Racial disparity in the wake of *Booker/Fanfan*: Making sense of "messy" results. *Criminology & Public Policy*. This issue.

Engen, Rodney L. and Randy Gainey. 2000. Conceptualizing the role of legal and extra-legal factors under sentencing guidelines: Reply to Ulmer. *Criminology*, 38: 1245–1252.

Frase, Richard S. 2005. Sentencing guidelines in Minnesota, 1978–2003. In (Michael H. Tonry, ed.,) *Crime and Justice: A Review of Research*, vol. 32. Chicago, IL: University of Chicago Press.

Reitz, Kevin R. 2005. The enforceability of sentencing guidelines. *Stanford Law Review*, 58: 155–174.

Scott, Ryan W. 2011. Race disparity under advisory guidelines: Dueling assessments and potential responses. *Criminology & Public Policy*. This issue.

Spohn, Cassia. 2000. Thirty years of sentencing reform: The quest for a racially neutral sentencing process. In *Policies, Processes and Decisions of the Criminal Justice System*. Vol. 3, *Criminal Justice 2000*. Washington, DC: U.S. Department of Justice.

Spohn, Cassia. 2011. Unwarranted disparity in the wake of the *Booker/Fanfan* decision: Implications for research and policy. *Criminology & Public Policy*. This issue.

Tonry, Michael H. 1996. *Sentencing Matters.* New York: Oxford University Press.

Ulmer, Jeffery T. 2000. The rules have changed—so proceed with caution: A comment on Engen and Gainey's method for modeling sentencing outcomes under guidelines. *Criminology*, 38: 1231–1243.

Ulmer, Jeffery T., Michael T. Light, and John H. Kramer. 2011. Racial disparity in the wake of the *Booker/Fanfan* decision: An alternative analysis to the USSC's 2010 report. *Criminology & Public Policy*. This issue.

U.S. Sentencing Commission. 2010. Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis. Washington, DC: Author.

Court Cases Cited

Furman v. Georgia, 408 U.S. 238 (1972).

Gall v. United States, 552 U.S. 38 (2007).

Gregg v. Georgia, 428 U.S. 153 (1976).

Jurek v. Texas, 428 U.S. 262 (1976).

Proffitt v. Florida, 428 U.S. 242 (1976).

Rita v. United States, 551 U.S. 338 (2007).

United States v. Booker, 543 U.S. 220 (2005).

Woodson v. North Carolina, 428 U.S. 280 (1976).

Statutes Cited

Anti-Drug Abuse Act of 1986, Pub.L. No. 99–570, § 1105(a), 100 Stat.3207–11 (1986).

Comprehensive Crime Control Act of 1984, S. Rep. No. 98–225, at 1312 (1983); H.R. Rep. No. 98–1017, at 55–56 (1984).

Omnibus Anti-Drug Act of 1988, Pub.L. No. 100–690, § 6371, 102 Stat. 4181, 4370 (1988).

PROTECT Act of 2003, Pub.L. 108–21, 117 Stat. 650, S. 151, enacted April 30, 2003; amendment Stat. 667.

Sentencing Reform Act of 1984, 18 U.S.C. §§ 3551–3626 and 28 U.S.C. §§ 991–998, as amended 1985–1988, 1990, 1992, 1994, and 1996.

Raymond Paternoster is professor in the Department of Criminology & Criminal Justice at the University of Maryland. His research interests include offender decision making, capital punishment, and the application of quantitative methods to criminological research.

Overview of: "Racial disparity in the wake of the *Booker/Fanfan* decision

An alternative analysis to the USSC's 2010 report"

Jeffery T. Ulmer
Michael T. Light
John H. Kramer
The Pennsylvania State University

Research Summary

The U.S. Sentencing Commission (USSC) released a report in March 2010 concluding that disparity in federal sentencing has increased in the wake of the U.S. Supreme Court decisions in United States v. Booker *(2005) and* Gall v. United States *(2007). In light of this USSC report, we provide an alternative set of analyses that we believe provides a more complete and informative picture of racial, ethnic, and gender disparity in federal sentencing outcomes post-*Booker *and* Gall*. We attempt first to replicate the USSC's models. Then, making different modeling assumptions, we present alternative models of sentencing outcomes across four time periods spanning fiscal years (FY) 2000 to 2009. We find that post-*Booker/Gall*:*

1. *Race/ethnic/gender disparity in sentence length decisions is generally comparable with pre-2003 levels;*
2. *African American males' odds of imprisonment have increased significantly post-*Gall*;*
3. *Immigration cases account for a significant proportion of sentence length disparity affecting Black males;*
4. *"Government-sponsored" below* Federal Sentencing Guidelines *sentences are a greater source of racial disparities than judge-initiated deviations.*

Finally, because much of the debate surrounding the Booker *decision involves questions of whether the guidelines must be mandatory to be effective, we also present analyses of federal sentencing disparities prior to the 1996* Koon v.

DOI:10.1111/j.1745-9133.2011.00760.x

United States *decision, a period when the guidelines were arguably most constraining of judicial decision making. Comparing post-*Booker *and* Gall *to pre-Koon sentencing practices, we find:*

5. *With the exception of incarceration disparities for black males, all race/ethnic/gender groups are sentenced either the same or less harshly (compared to whites) under the new advisory system.*

Policy Implications

One of the chief policy aims of the sentencing guidelines, and the sentencing reform movement more generally, is the reduction of racial, ethnic, and gender disparities in punishment. In the aftermath of the Booker *and* Gall *decisions, which made the* Federal Sentencing Guidelines *effectively advisory, many commentators have feared that federal judges, as a result of their increased discretion, may use offender characteristics at sentencing, and thus disparities based on race, ethnicity, and gender would increase. A recent USSC report, which found that sentencing disparities have increased in the wake of* Booker *and* Gall*, suggests that such fears are warranted and has further strengthened calls for a policy fix to post-*Booker *sentencing. One such proposed policy solution was House bill H.R. 1528 which would have, among other things, transformed the* Sentencing Guidelines *into a complex system of mandatory minimums in order to curb judicial discretion.*

*In response to the USSC report as well as policy solutions to introduce a more rigid sentencing system, we suggest that the federal sentencing data, as yet, provide insufficient evidence of increased post-*Booker *disparity to warrant renewed restrictions on judicial discretion. Our analysis, generally, shows that sentencing disparities post-*Booker *and* Gall *are comparable to those just prior to these decisions, and are actually considerably less than in earlier time periods when the guidelines were more rigid and constraining. This latter finding raises serious questions about whether the guidelines must be mandatory in order to limit racial/ethnic/gender disparities.*

*However, we do find that disparity in the imprisonment decision for black males increased post-*Gall*, which is ground for concern. Yet our results do not lend support to policy solutions that would seek to simply "re-mandatorize" the guidelines. Because our analysis shows that the bulk of extralegal disparities are observed in the incarceration decision and not the sentence-length decision (a distinction that is not made in the USSC report), reintroducing a rigid sentencing scheme may actually exacerbate incarceration disparities while having limited impact on sentence length disparities. Overall, rather than "blanket" solutions such as broad reductions in judicial discretion, we think that any policy changes to the* Sentencing Guidelines *should be focused on areas that are shown to be associated with sentencing disparities. According to our results,*

such areas include incarceration decisions, immigration offenses, and government-sponsored downward departures. Ultimately, we think careful consideration of the most problematic areas of sentencing, including those decisions made by prosecutors as well as judges, will be most effective at curbing extralegal disparities and increasing fairness at sentencing, both of which were original goals of the sentencing reform movement.

Keywords

sentencing disparity, *judicial discretion*, *federal courts*, *United States v. Booker decision*

Racial disparity in the wake of the *Booker/Fanfan* decision

An alternative analysis to the USSC's 2010 report

Jeffery T. Ulmer
Michael T. Light
John H. Kramer
The Pennsylvania State University

The U.S. Supreme Court ruled in *United States v. Booker* and a joint case *United States v. Fanfan* (2005, hereafter *Booker* and *Fanfan*, respectively) that the Federal Sentencing Guidelines (hereafter, Guidelines) would henceforth be advisory rather than presumptive in federal sentencing decisions. Many fear that the wake of *Booker/Fanfan* might have brought increased unwarranted disparity based on the social status characteristics of defendants (see reviews by Frase, 2007; Hofer, 2007). Hofer (2007) argued that if a primary goal of federal sentencing reform was a reduction of unwarranted disparity, the impact of *Booker/Fanfan* on disparity is among the most important questions facing sentencing policy makers. Chief among these concerns is the degree of disparity connected to race and ethnicity, the reduction of which was a key reason for the Guidelines' creation.

The U.S. Sentencing Commission (USSC) released a report in March 2010 concluding that racial disparity in federal sentence lengths has indeed increased in the wake of the *Booker* and *Gall v. Unietd States* (2007, hereafter *Gall*) decisions. Specifically, the report's "refined models" found that Black males had approximately 5% greater sentence lengths than White males in 2003–2004, 15% greater sentence lengths after the *Booker* decision, and approximately 21% greater sentence lengths post-*Gall*. Thus, from the report, it seems that racial disparity affecting Black males (and Black defendants in general) has become

We thank D. Wayne Osgood and Shawn Bushway for their helpful input on earlier drafts of this article. Direct correspondence to Jeffery T. Ulmer, Department of Sociology, The Pennsylvania State University, 211 Oswald Tower, University Park, PA 16802 (e-mail: jtu100@psu.edu).

DOI:10.1111/j.1745-9133.2011.00761.x

worse in the years since *Booker*, and especially since *Gall*. This is an alarming development for those who are rightly concerned with the racial fairness of federal justice.

Yet, the USSC 2010 report's analyses made some methodological choices that differ from those of several federal sentencing studies in the literature, and we detail these choices in the subsequent discussion. It is, therefore, important to examine whether the USSC 2010 racial disparity findings are apparent when different analytical and modeling choices commonly found in the sentencing literature are made. In addition, the USSC research staff was not directed in their 2010 report to present an analysis of whether disparity has increased post-*Booker* in sentences that depart/deviate from the Guidelines, and they did not compare their refined model findings with time periods earlier than the years when the PROTECT Act was in force (2003–2004). We, therefore, present such analyses because judicial discretion to deviate from the Guidelines has increased post-*Booker*, and Guidelines departures have been found to be the locus of extralegal disparity in research on pre-*Booker* sentencing (Albonetti, 1997; Hartley, Maddan, and Spohn, 2007; Johnson, Ulmer, and Kramer, 2008; Kempf-Leonard and Sample, 2001; Maxfield and Kramer, 1998; Mustard, 2001; Steffensmeier and Demuth, 2000).

Our analysis may present a fuller picture of the nature of racial disparity in the wake of the *Booker* and *Gall* decisions, as well as the relaxation of constraints on judicial discretion that they brought. According to Attorney General Holder (2009) as well as sentencing policy observers such as Paul Hofer (2007), this is one of the most pressing and timely questions faced by the federal sentencing community. If unwarranted disparity has increased in the post-*Booker/Gall* years, some argue that policy remedies are necessary to return the Guidelines somehow to a mandatory status and to attempt to roll back the judicial discretion granted by *Booker* and subsequent decisions.

In a recent essay in *Criminology & Public Policy*, Engen (2009) also noted the paucity of research on what happens in the wake of the repeal or relaxation of presumptive sentencing schemes. By examining sentencing in the aftermath of *Booker*, which loosened constraints dramatically on federal judicial discretion, we are helping to address the agenda Engen (2009) proposed. In sum, we provide a timely alternative analysis that we believe provides more specificity and guidance regarding questions vital to federal sentencing policy: (a) whether and how much racial disparity in federal sentencing has increased in the aftermath of *Booker* and *Gall*; (b) whether disparity has increased in particular kinds of sentencing decisions (i.e., sentence lengths, imprisonment, and Guidelines departures/deviations) or for particular offenses; and (c) whether the levels of racial disparity post-*Booker* are significantly greater compared with longer term federal sentencing patterns.

The Return of Federal Judicial Discretion: The *Booker* and *Gall* Decisions

From 1996 to 2005, legal developments moderately expanded judicial sentencing discretion, then sharply restricted it, and finally, culminating in *Booker*, dramatically expanded it again. From 1987 to 1996, discretion historically resting with the judiciary was tightly constrained and shifted to the prosecutor (Stith and Cabranes, 1998). Congress continued to restrict judges' sentencing discretion during this period, sending directives to the Commission, and passing mandatory minimums to be incorporated into the Guidelines. Then, in *Koon v. United States* (1996, hereafter *Koon*), the Supreme Court restored some discretion to judges by establishing an "abuse of discretion" standard for appellate review of departures from the Guidelines. Congress later sought to counter *Koon* with the Feeney Amendment to the PROTECT Act of 2003, which replaced the "abuse of discretion" standard for departures with a "de novo" appellate review of sentences, gave prosecutors control over the third point of the "acceptance of responsibility" Guidelines reduction, and directed the Commission to reduce departure mechanisms.

Then, the *Booker* decision in 2005 ruled that the mandatory Guidelines could not constitutionally assess "real offense" conduct that increased sentences on factors not considered at trial by a jury. The Court's solution was that the Guidelines would become *advisory*. Judges must consider the Guidelines, but their discretion was returned to at least pre-PROTECT Act, although not to pre-Guidelines, levels. Also, in the wake of *Booker,* the standard of review now relies on the "reasonableness" of the sentence and on an "abuse of discretion" standard rather than on correct application of the Guidelines. Stith (2008: 1,427) stated: "*Booker*, the Sentencing Commission and Main Justice may still be calling signals but the decision makers on the playing field—judges and prosecutors—need not follow them."

Subsequently, the Court enhanced the judges' discretion restored in *Booker* by clarifying the meaning of "advisory" in *Rita v. United States* (2007), where it ruled that federal appellate courts *may* but are not *required* to presume Guidelines sentences to be reasonable. Consequently, sentences outside the Guidelines cannot be automatically regarded as unreasonable. In *Gall*, the Court went further and held that district judges *may not* automatically presume the Guidelines range to be reasonable and must "*make an individualized assessment based on the facts presented*" (*Gall*, pp. 596–597, emphasis added). *Gall* thus implies that district courts should make an individualized assessment of whether a Guidelines sentence is reasonable or whether a sentence outside the Guidelines is more reasonable. In *Kimbrough v. United States* (2007), the Court ruled that in cases involving crack cocaine, judges could reasonably conclude that Guidelines sentences were not reasonable in an individual case.

Policy observers have had different reactions to these developments. U.S. Attorney General Eric Holder (2009: 1) noted that uniformity and the control of judicial discretion per se do not guarantee justice: "The desire to have an almost mechanical system of sentencing has led us away from individualized, fact-based determinations that I believe, within reason, should be our goal." Some, including the USSC, have adopted a "wait-and-see" approach to post-*Booker* sentencing. For example, in 2005, an American Bar Association (ABA) Task Force Report recommended that sufficient time be allowed to evaluate the efficacy of the new "advisory" guidelines, asserting that "the advisory remedy crafted in *Booker* may well prove as good or even better than the mandatory guidelines" (ABA, 2005: 339).

However, prominent U.S. Attorney John Richter (2008: 340), presenting a view held by many other federal prosecutors, argued that "[p]ost-*Booker* sentencing threatens equal justice under law." The dissenting opinions of Justices Stevens, Thomas, and Scalia in *Booker* each noted that Congress clearly intended to restrict judicial discretion to curb unwarranted disparity, and they argued that the Court majority's remedy of making the Guidelines advisory would jeopardize that goal.

Along these lines, then-Attorney General Alberto Gonzales (2005: 325) claimed that, since *Booker*, there has been "increasing disparity in sentences," and therefore, the *Guidelines* were in need of a legislative fix. Specifically, Gonzales (along with others) supported the proposed Consumer Privacy Protection Act,of 2005, which would have (a) transformed the Guidelines into a complex system of mandatory minimums (Bowman, 2005), (b) essentially forbidden the consideration of mitigating factors at sentencing (Berman, 2005), and (c) restricted severely the use of nonprosecutorial downward departures. Other sentencing scholars have proposed "hybrid" solutions. For example, legal scholar and former Special Counsel to the USSC Frank Bowman proposed simplifying the sentencing table to only nine base offense levels (down from the current 43) where no upward departures from the base sentencing range would be permissible (so as not to run afoul with *Booker*), although downward departures based on "acceptance of responsibility," motions by the prosecutor, or other relevant mitigating factors would be allowed (Bowman, 2005).

Research on Federal Sentencing Disparity

Much scholarly research on federal courts has assessed unwarranted disparity under the pre-*Booker* Guidelines. These studies often found small-to-moderate racial and ethnic sentencing differences benefitting Whites, although Guidelines-relevant factors exert much larger effects than offender status characteristics (e.g., Albonetti, 1997, 1998; Johnson et al., 2008; Kautt, 2002; Mitchell and MacKenzie, 2004; Mustard, 2001; Steffensmeier and Demuth, 2000; USSC, 2004). Evidence also suggests that extralegal differences in punishment are tied to departure sentences (Albonetti, 1997; Hartley et al., 2007; Johnson et al., 2008; Kempf-Leonard and Sample, 2001; Maxfield and Kramer, 1998; Mustard, 2001; USSC, 2004). Research using pre-*Booker* data showed that young minority males in particular were disadvantaged in incarceration decisions and sentence lengths (Doerner and Demuth, 2009); that defendant race, age, and gender influenced prosecutorial charge reductions, which in turn influence sentencing outcomes (Shermer and Johnson, 2010); that the degree to which race/ethnicity and gender influence sentencing varies significantly by judge (Anderson and Spohn, 2010); and that Hispanic defendants are most disadvantaged in sentencing in federal districts where Hispanics are least numerous, but not at all disadvantaged in districts with large Hispanic populations (Feldmeyer and Ulmer, 2011).

However, with the exception of two USSC reports published in 2006 and 2010, all the published research on federal sentencing disparity is based on pre-*Booker* data and most is based on pre-PROTECT Act data. The 2006 report showed that most federal cases continued to be sentenced in conformity with the Guidelines but that the rate of above-range, government-sponsored below-range, and other below-range sentences increased. Multivariate analyses showed that social status factors were associated moderately with sentence length but that their effects pre- and post-*Booker* were similar, and that race actually had more influence on sentence lengths in 1999–2000 than in the early post-*Booker* period (USSC, 2006). The report also examined conformity and departures by circuit and district from 2001 to January 2006 and concluded that regional sentencing differences have been relatively stable. A commentary on this report stated: "With a little over a year's experience under *Booker*'s new 'advisory' guidelines regime, the cumulative results can be summarized as 'much ado about nothing, or at least much ado about not very much'" (Thompson, 2006: 269). Overall, the 2006 report notes that disparity decreased in the PROTECT era, but after *Booker,* it returned to levels comparable with those of the pre-PROTECT act era.

However, the USSC 2010 report, which included data up to FY 2009, found that race disparity had increased in the post-*Gall* period compared with the PROTECT Act period. Their models first replicated the analyses in the 2006 report with the newer data included, and then they estimated a "refined model." Their "*Booker* report" model showed that Blacks received approximately 2% longer sentences than Whites (not significant) during the PROTECT era but that Blacks received 7% and 10% longer sentences than Whites in the post-*Booker* and post-*Gall* periods, respectively. Notably, Black–White sentence-length differences ranged from a high of 14% to a low of 8% in the pre-PROTECT years FY 1999–2002, as stated in the USSC 2006 report.

Their "refined model," which did not control for criminal history, but did differentiate Black, White, and Hispanic defendants by gender, found that Black males received 5.5% longer sentences than White males in the PROTECT period, 15% longer sentences post-*Booker,* and 21% longer sentences post-*Gall*. The 2010 report also found that noncitizens were increasingly sentenced more harshly than U.S. citizens and that gender disparity fluctuated across time periods. To be clear, the USSC 2010 report did not claim that *Booker* and *Gall* caused increases in racial disparity and recognized that other factors not related to the two decisions could be driving these increases. Nonetheless, the report's findings would seem to provide support for critics of the two decisions who call for remedies to reconstrain judicial discretion.

For our purposes, the USSC 2010 report's analyses have four notable methodological features: (a) the sentence-length models included nonimprisonment cases as sentence lengths of "0," thus combining the incarceration and length decision into one analysis (and used ordinary least-squares [OLS] regression, rather than tobit regression); (b) the report included immigration offenses in the analyses; (c) sentence-length models equated periods of alternative confinement with periods of imprisonment; and (d) the refined model did not control for criminal history because of concerns about multicollinearity. Yet, previously published USSC studies of disparity in federal sentencing, such as the USSC's 2004 report, along with several studies published in the criminology literature, often made methodological choices that differed from these four features. Given these differing methodological choices, it is therefore important to examine whether the USSC 2010 racial disparity findings hold in the face of different analytical and modeling choices commonly found in sentencing studies.

To begin, the USSC's 2010 sentence length models included nonimprisonment cases as sentence lengths of "0," thus combining the incarceration and length decision into one OLS analysis. This strategy is relatively uncommon in the sentencing literature because it (a) assumes that there is no selection in the imprisonment decision relative to the length decision, (b) creates problematic distributional issues for standard OLS regression, and (c) offers opaque results regarding policy recommendations for the Guidelines. Although some might argue that nonincarceration sentences should be included as zeros because these offenders' "true" sentence lengths are not unobserved or censored, but are actually 0 months, we argue that this approach would be analogous to conducting research on wage disparity and including the unemployed, claiming that unemployed people actually receive wages of $0 (see Bushway, Johnson, and Slocum, 2007). We prefer to view only those *selected for incarceration* as eligible to receive sentence lengths. In other words, we view offenders' sentence lengths as conditional on whether they were sentenced to prison (and we will consider only imprisonment cases as sentence lengths, as we explain later).

This issue also raises the problem of the potential for selection bias, which is endemic to research on criminal justice decision making, and there is no definitive "right" way to handle it (Bushway et al., 2007; Stolzenberg and Relles, 1997). Most state and federal sentencing studies treat the imprisonment decision and the length decision as two related but distinct decisions (what Bushway et al. [2007] call the "two-part model"), and then these studies consider the issue of selection by including or not including a Heckman two-step correction for selection bias stemming from the imprisonment decision in sentence-length models (for some among many examples, see Anderson and Spohn, 2010; Doerner and Demuth, 2009; Johnson et al., 2008; Kautt, 2002; Peterson and Hagan, 1984; Spohn and Holleran, 2000; Steffensmeier and Demuth, 2000; Steffensmeier, Kramer, and Streifel, 1993; Ulmer, 1997; Ulmer, Eisenstein, and Johnson, 2010). The purpose of doing the Heckman correction is to generate estimates that refer to the *potential* population of everyone who could have been selected (Bushway et al., 2007). This strategy is in contrast to analyses that include only those who were sentenced to prison (i.e., second part of the two-part model) because these estimates refer only to the *actual* incarcerated population. Although the uncorrected two-part model may not capture potential selection *bias* because it focuses only on the effects on imprisonment length conditional on being imprisoned (Bushway et al., 2007), the approach used in the *Booker* reports does not assume any selection in the imprisonment decision by treating nonincarcerated offenders as incarcerated offenders for 0 months.

In other words, the latter approach assumes there is no distinction between those actually selected and those potentially selected, and it treats everyone as having a sentence length. It is for the preceding reasons that scholars attempted to use the Heckman correction; however, because the application of the selection procedure is problematic for several reasons in sentencing research, we opted to analyze sentence lengths for only those who were actually incarcerated.[1] That is, we employ a two-part model (although we do report the results of our "corrected" models in the text).

A second solution for problems with selection is the use of tobit regression (see Bushway et al., 2007), which treats the sentence-length variable as an instance of censoring. Because thousands of convicted offenders do not receive prison each year, these "zeros," when left in the distribution, create a problem with left censoring (i.e., there is a large category of individuals at the bottom of the distribution, which violates OLS assumptions regarding a normally distributed dependent variable). In this case, a tobit model treats the sentence-length distribution as a normal one but explicitly treats the zero sentence lengths as a point of censoring, and it assumes that the likelihood function would be normally distributed were it observed fully (for examples, see Albonetti, 1997, 1998; Bushway and Piehl, 2001; Kurlychek and Johnson, 2004). As an illustration of this censoring problem, we display the distributions of the sentence-length–dependent variables used by the USSC (panel A) and in our analysis (panel B) in Figure 1.

Both variables are displayed along a logarithmic scale. From a purely statistical standpoint, the distribution in panel A is highly problematic for an OLS regression equation. First, the variable does not approximate a normal or even near-normal distribution. Indeed, the modal category of this distribution is –4.61 [$\ln(0.01) = -4.61$], which is the furthest left tail of the distribution.[2] As we demonstrate subsequently, this modeling choice has dramatic effects on how sentence-length results are interpreted. Panel B, in contrast, displays the distribution of the dependent variable used in our analysis. By analyzing only those who

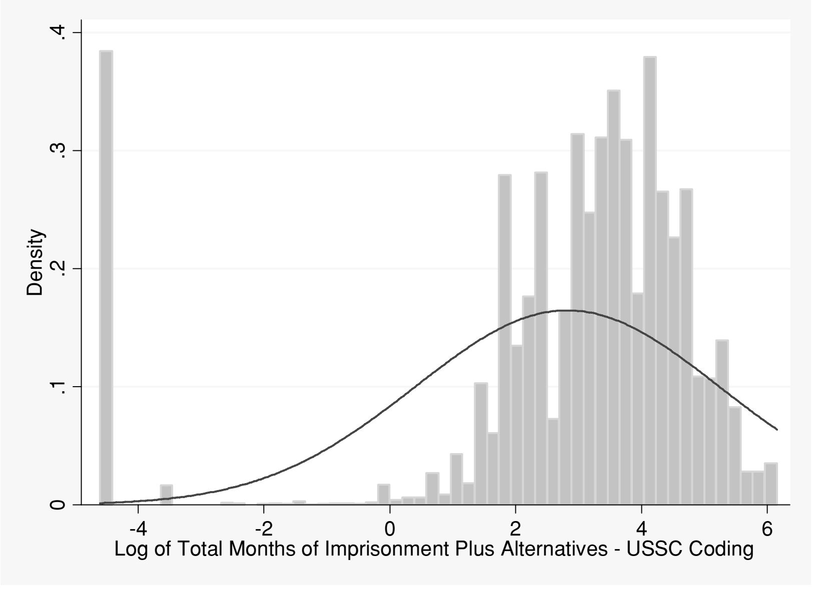
1. Rather than including "zeros" in standard OLS models, scholars have adopted several strategies to account for this possible selection bias, often using the Heckman two-step procedure, in which an individual's probability of being selected into the population of interest (in this case those receiving sentence lengths) is first calculated (using the inverse mills ratio), and then this conditional probability is entered into the OLS model. Although this selection correction may be justified theoretically, as Bushway et al. (2007) demonstrated, often its application is complicated and problematic because sentencing data usually do not include proper selection instruments that affect only an offender's likelihood of incarceration but not his or her length of imprisonment. As a result, the selection equation often includes many of the same predictors (i.e., criminal history, offense severity, race, etc.) as the substantive equation (sentence length) that introduces problems with multicollinearity, and model identification. Using similar procedures as Bushway et al. (2007), we find that using the Heckman procedure produced substantively similar results as those we present (discussed subsequently) but did in fact introduce problematic multicollinearity into our models of sentence length.

2. The USSC gave all zeros a value of 0.01 prior to logging because the log of 0 is not mathematically possible. Hence, all offenders who received probation or who were not incarcerated make up this category of –4.61 on the logarithmic scale.

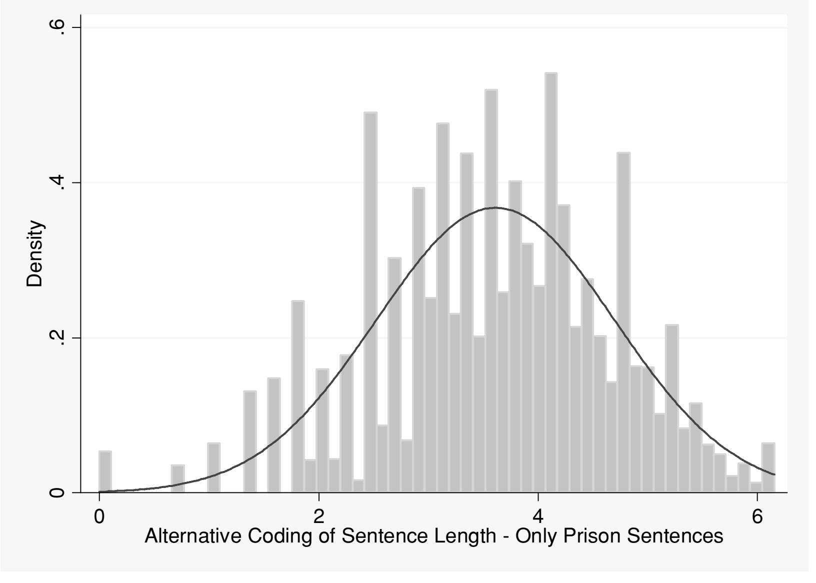
FIGURE 1

Histograms of Dependent Variables: Panel A - USSC Coding; Panel B - Authors' Alternative Coding

Panel A



Panel B



were actually sentenced to prison, logging approximates a normal distribution and does not require any correction for censoring. Again, although this modeling choice may be susceptible to selection bias, to ensure the robustness of our results, we also ran tobit regression models (discussed in text subsequently).

Perhaps the strongest argument against modeling sentencing decisions similar to the *Booker* report is the inability to separate out disparities occurring at either the incarceration stage or the sentence-length stage. As stated, although we acknowledge that our models of sentence length may ignore potentially problematic selection bias, from a policy perspective, we argue that our analytical approach is more appropriate because it does not conflate the incarceration and sentence-length decisions. By combining both decisions into one model (aside from the distributional and statistical issues discussed previously), the USSC model does not allow for the possibility that predictors might have different effects on imprisonment and sentence-length. However, this situation has often been found to be the case in sentencing research—and in fact, it is common to find that extralegal variables such as race/ethnicity have stronger impacts on incarceration than on sentence-length in sentencing research (see reviews by Spohn, 2000; Zatz, 2000; see also Doerner and Demuth, 2009; Johnson, 2006; Kramer and Ulmer, 2009; Steffensmeier et al., 1993; Ulmer and Johnson, 2004). In fact, the strategy adopted in the 2010 report differs from modeling choices made in previous USSC publications, in which incarceration and sentence-length analyses were modeled separately (see USSC, 2004: ch. 4). By combining the incarceration and length decisions into one model, the USSC report may be overstating the amount of sentence-length disparity, yet failing to pinpoint disparity in the incarceration decision.

The second methodological issue is that much of the previous research on federal sentencing has either excluded immigration or noncitizen cases from the analysis (Doerner and Demuth, 2009; Steffensmeier and Demuth, 2000; Ulmer et al., 2010) or analyzed these cases separately (Feldmeyer and Ulmer, 2011) for several reasons. First, often immigration offenses are handled differently than other federal crimes because of the intersection of immigration and criminal law, the possible involvement of foreign governments (Steffensmeier and Demuth, 2000), and the use of deportation as a sentencing option (only for non-U.S. citizens, who make up the overwhelming majority of immigration offenders). In fact, in the USSC's 15-year assessment of how well the Guidelines have accomplished the goals set out by Congress, the Commission excluded noncitizens from their analysis of racial, ethnic, and gender disparity because "inclusion of non-citizens, who are often non-White, confounds race and ethnicity effects of those with citizenship" (USSC, 2004: 120). Second, districts with comparatively large numbers of immigration cases commonly employ "fast-track" programs designed to expedite such cases (Bowman, 2003), whereas others do not have such fast-track programs. Fast-track programs present problems with uniformity in the system because the affected sentences are dependent not just on an offender's criminal conduct but on the district in which the offender is prosecuted (Maxfield and Burchfield, 2002). In the absence of controls for district variation (such as fixed-effects models) or fast-track departures, this presents a potential omitted variable bias. Also, U.S. citizens would seldom be convicted of offenses involving "unlawfully entering or remaining in the U.S." (see §2L1.2 in the *U.S. Guidelines Manual*), which represent more than 70% of all immigration crimes. Again, this is not to say that immigration offenses should not be evaluated in sentencing outcomes. On the contrary, given the dramatic growth of such offenses in federal courts, we think this especially important issue deserves critical attention. However, we do argue that there are good reasons to suspect that immigration offenses are handled in distinct ways from most other offenses, and any analysis of federal data should be attuned to their distinctiveness, perhaps analyzing them separately.

The third methodological difference also involves the dependent variable. Whereas much previous research has examined sentences of incarceration to *prison* (which is the method we employ in our analysis; see also Bushway and Piehl, 2001; Doerner and Demuth, 2009, for similar analyses), the USSC uses a dependent variable that captures the months of *confinement* to either prison, home detention, community confinement, and intermittent confinement.[3] In other words, the racial disparities in "sentence lengths" reported by the Commission could be a result of different prison sentences or could be a result of different terms of community confinement or home detention. Although it is certainly important to research racial disparities in these other forms of confinement, we argue that sentences of home detention (and other forms of confinement) are qualitatively different from time in prison to the point where such sentences should not be analyzed as equivalent forms of incarceration.

Finally, the USSC 2010 report did not include controls for criminal history in their "refined" models because of issues of multicollinearity and because criminal history is one of the components of the presumptive sentence measure (see the Data section). However, criminal history has been shown to be an important independent predictor of sentencing outcomes beyond that captured by the presumptive sentence in published research on state and federal sentencing (Albonetti, 1998; Doerner and Demuth, 2009; Feldmeyer and Ulmer, 2011; Johnson and Betsinger, 2009; Johnson et al., 2008; Ulmer, 2005). These studies, along with our analysis, did not report severe multicollinearity with these two measures; however, criminal history was notably correlated with race (Black defendants tend to have higher mean criminal history scores).[4] Thus, any increase in racial disparity could possibly be because judges (or prosecutors) put more weight on criminal history in the wake of *Booker* and *Gall*.

Why would researchers want to control for criminal history in sentencing models above and beyond its influence through the presumptive Guidelines sentence? One answer is that, even if criminal history influences sentencing over and above the effect of Guidelines

3. This variable is SENSPLT0 in the USSC data files.

4. The bivariate correlation between criminal history and presumptive sentence is approximately 0.35 in all time periods.

minimums and is therefore a discretionary rather than a Guidelines-driven consideration of criminal history, sentencing variation explained by criminal history is not variation explained by race or ethnicity (or other defendant social statuses). That criminal history may *mediate* part of the effect or race/ethnicity or other characteristics indicates to us the importance of controlling for it when we try to identify the sentencing effect of race/ethnicity that is not attributable to other factors. Also, as we note, the USSC in their "*Booker*" models and several other federal sentencing studies include criminal history in sentencing models. However, as we discuss in the Results and Conclusion sections, there is legitimate debate as to the proper method for accounting for offender criminal history in sentencing studies that deserves additional attention.

It is also important to put the racial disparity findings from the USSC's refined model in broader temporal context. That is, how do post-*Booker* levels of sentencing disadvantage for Black males, for example, compare with Black male sentencing patterns in the pre-PROTECT Act era, or even before the important 1996 *Koon* decision? Perhaps the relatively low levels of racial disparity during the PROTECT Act era were atypical in the history of the Guidelines, and post-*Booker* racial disparity levels are comparable with earlier periods when the Guidelines were mandatory, but the PROTECT Act restrictions were not in effect. If this were the case, then it would not support arguments that the *Booker* and *Gall* decisions, and the increased judicial discretion they brought, produced a new trend of racial disparity in federal sentencing.

We attempt first to replicate the USSC's refined sentence length model (and also extend this analyses to the pre-PROTECT Act era) and then present alternative models that (a) examine disparity in the incarceration and length decisions separately, (b) control for criminal history, (c) do not equate alternative confinement with imprisonment, and (d) show levels of disparity with immigration offenses included in the models versus when they are excluded. We also extend the time period comparisons of racial disparity to the pre-PROTECT and the pre-*Koon* eras.

In addition, the 2010 report did not present an analysis of whether disparity has increased in sentences that depart/deviate from the Guidelines in the post-*Booker* periods, and it did not compare their refined model findings to time periods earlier than the years when the PROTECT Act was in force (2003–2004). We, therefore, present an analysis of whether disparity in departures (and which kinds of departures) has increased post-*Booker* and post-*Gall*, since judicial discretion to deviate from the Guidelines has increased post-*Booker*, and Guidelines departures have been found to be the locus of extralegal disparity in research on pre-*Booker* sentencing.

Data

The data come from the U.S. Sentencing Commission's Standardized Research Files, which are the same data used by the USSC for its reports. Consistent with the USSC's reports on the effects of *Booker*, we use the four time periods noted previously to assess the impact of *Booker*: (a) cases sentenced in the pre-PROTECT Act period, which includes fiscal year 2002 (October 1, 2001–September 20, 2002) and fiscal year 2003 through April 2003; (b) cases sentenced in the PROTECT Act period which includes the second part of fiscal year 2003 (see footnote 1) and fiscal year 2004 through June 2004, which corresponds with the decision by the Supreme Court in *Blakely v. Washington* being handed down on June 24, 2004 (hereafter *Blakely*); (c) cases sentenced in the post-*Booker* period (January 2005 through November 2007); and (d) cases sentenced post-*Gall* (December 2007 through September 2009).[5,6,7] The unit of analysis is each sentenced case.

Dependent Variables

Our analysis examines the following three dependent variables for each of the four time periods: (a) length of sentence, (b) the imprisonment decision, and (c) the likelihood of receiving downward departures from the Guidelines, where substantial assistance and nonsubstantial assistance ("other" departures) are analyzed separately. Coefficients from the four separate time periods (pre-PROTECT, PROTECT, early post-*Booker*, and later post-*Booker*) are compared using z tests (Clogg, Petkova, and Haritou, 1995; Paternoster, Brame, Mazzerolle, and Piquero, 1998). The first dependent variable is the sentence length ordered for each offender (capped at 470 months). For the analyses after Figure 2, our dependent variable differs from that used in the USSC 2010 report in that we only use terms of imprisonment in our analysis, whereas their analysis includes months of alternative confinement including home detention, community confinement, and intermittent confinement. The USSC's sentence-length models also contain those who did not receive confinement sentences (e.g., probation) as sentence lengths of "0" (or 0.01, because 0 cannot be logged), whereas our analyses after Figure 2 do not. Because the sentence-length variable is skewed positively and regression diagnostics indicated problematic standard errors, we use the natural log transformation (as did the USSC in its reports). Our other dependent variables are dichotomies: (a) incarceration = 1, 0 if not; and (b) downward departures (of particular kinds) = 1, 0 if not.

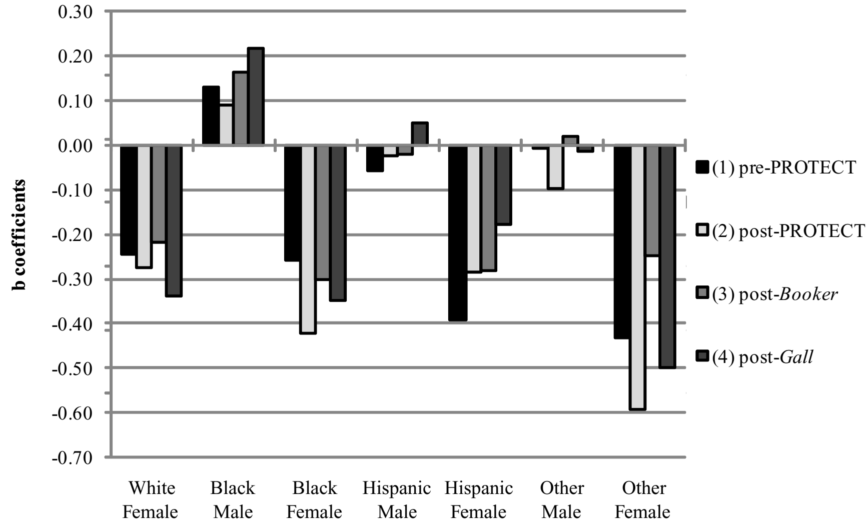
Independent Variables

Consistent with prior research, we control for the *Guidelines-recommended sentence* by including a measure of the presumptive sentence equal to the minimum months of incarceration recommended by the sentencing guidelines after adjusting for any mandatory

5. Seven months of fiscal year 2003 were prior to the effective date of the PROTECT Act (October 1, 2002–April 30, 2003), and 5 months were after (May 1, 2003–September 30, 2003).
6. We remove the period between the *Blakely* and *Booker* decisions to remove any potential *Blakely* effects.
7. The *Booker* decision was handed down on January 12, 2005, and *Gall* was decided on December 10, 2007.

FIGURE 2

U.S. Sentencing Commission "Refined" OLS Models of Sentence Length



Race-Gender Combinations	**(1) pre-PROTECT** *Ln Length*[a]	**(2) post-PROTECT** *Ln Length*	**(3) post-*Booker*** *Ln Length*	**(4) post-*Gall*** *Ln Length*
	b	*b*	*b*	*b*
White Male (reference)				
White Female	-0.244[*]	-0.275[*]	-0.216[2*]	-0.340[1,2,3*]
Black Male	0.130[*]	0.089[*]	0.164[1,2*]	0.217[1,2,3*]
Black Female	-0.258[*]	-0.422[1*]	-0.300[2*]	-0.349[1*]
Hispanic Male	-0.058[*]	-0.023	-0.019[1*]	0.049[1,2,3*]
Hispanic Female	-0.391[*]	-0.286[1*]	-0.280[1*]	-0.177[1,2,3*]
Other Male	-0.008	-0.097[1*]	0.020[2]	-0.014[2]
Other Female	-0.432[*]	-0.592[1*]	-0.247[1,2*]	-0.499[3*]

[1] Coefficient is significantly different from pre-PROTECT Act estimate based on two-tailed z test ($p < .05$).
[2] Coefficient is significantly different from post-PROTECT Act estimate based on two-tailed z test ($p < .05$).
[3] Coefficient is significantly different from post-*Booker* estimate based on two-tailed z test ($p < .05$).
[a] Models include controls for all variables in Appendix A.
$p < .01$.

minimum trumps (Albonetti, 1998; Engen and Gainey, 2000; Johnson and Betsinger, 2009; USSC, 2004b). This measure incorporates the offense severity level and the criminal history, and it accounts for statutory sentencing provisions (i.e., mandatory minimum penalties) that affect the final presumptive sentence. As with sentence length, we cap the presumptive sentence variable at 470 months and take the natural log to reduce positive skewness.[8] Although criminal history is included in the presumptive sentence measure, we follow previous research (e.g., Albonetti, 1998; Doerner and Demuth, 2009; Johnson and Betsinger, 2009; Johnson et al., 2008; Ulmer, 2005; Ulmer et al., 2010) and include an additional control for the offender's criminal history score.

We also control for the type of offense with a set of dummy variables (drug, violent, fraud, firearms, and other offenses, with property offenses as the reference category). We control for two case characteristics: whether the offender was detained prior to sentencing, coded 1 if the offender was detained and 0 otherwise; and whether the individual was convicted by trial, coded 1 for a trial conviction and 0 otherwise. Our sentence-length analyses include as predictors dummy variables for whether the defendant received an upward, downward, or substantial assistance (5K1) departure (coded 1 for these departures and 0 otherwise).

As in the USSC 2010 report's refined models, race/ethnicity and gender are combined into a set of dichotomous categories, a practice sometimes found in other sentencing studies as well (e.g., Doerner and Demuth, 2009; Kramer and Ulmer, 2009; Steffensmeier, Ulmer, and Kramer, 1998). In all analyses, we include dummy variables for Black males, Hispanic males, Black females, Hispanic females, White females, other race/ethnicity males, and other race/ethnicity females, with White males as the reference category. We also include a dummy variable for citizenship, with noncitizens coded as 1. Education is captured with four separate dummy variables: less than high school, high school graduates, some college, and college graduates as the reference.

Results

First, we present our replication of the USSC's refined model, adopting their sentence-length variable (with nonconfinement sentences-included and with alternative confinement counted as equivalent to imprisonment) as well as their coding of all independent variables, but we extend the time period of comparison to the pre-PROTECT Act era.[9] Second, we present our alternative sentence-length models across the four time periods. We then present similar models of the incarceration decision to compare racial/gender disparity across the different decision types. Fourth, incarceration and sentence-length decisions are reanalyzed without immigration offenses to evaluate the influence of these cases on demographic

8. A constant of 0.1 is added to all zero values for the presumptive sentence variable but not for the sentence-length–dependent variable. Taking the log of zero would exclude these values from the analysis. This is appropriate for the dependent variable because we want to analyze only those offenders who actually received a sentence length. The zeros are retained in the presumptive sentence variable (by adding 0.1 to all 0 values) because we want to retain those cases where an offender's minimum sentence was 0 months but he/she still received a prison sentence.

9. See the Appendix in *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis* (USSC, 2010) for a description of all coding procedures used in USSC analyses.

disparities in sentencing outcomes. Fifth, we compare post-*Booker* sentencing to sentencing practices prior to *Koon v. United States* (1996) to test the validity that a return to more mandatory guidelines will "correct" the problems wrought by *Booker* and *Gall*. Finally, we examine the effects of race/ethnicity–gender categories on the likelihood of receiving different kinds of downward departures/deviations from Guidelines across the time periods. Our primary focus is on comparing and contrasting our findings with those of the USSC with regard to disparity connected to the race/ethnicity–gender categories across the various time periods, and on extending the analysis of post-*Booker* race/ethnicity–gender disparity to decisions that depart/deviate below the Guidelines.

Replication and an Alternative to the USSC 2010 Report

Figure 2 shows the results from our replication of the models run in the USSC 2010 report.[10]

The results in Figure 2 display the USSC models of sentence length (logged) regressed on offender characteristics, case processing factors, offense categories, and Guidelines factors for each of the four time periods.[11] Consistent with previous research, race/ethnicity and gender exert significant effects on sentence lengths in all time periods. Moreover, these results display similar patterns reported by the USSC, where certain forms of disparity have increased since *Booker* and *Gall*. For example, the Black male effect decreased from 0.130 in the pre-PROTECT era to 0.089 after the passing of the PROTECT Act, but then it increased to 0.164 in the wake of *Booker* and then again to 0.217 after *Gall*. Moreover, z tests show that these increases are statistically significant (see Appendix A). Figure 2 also shows that White female disparity has increased slightly since *Booker* and *Gall*, as has the Hispanic male effect.

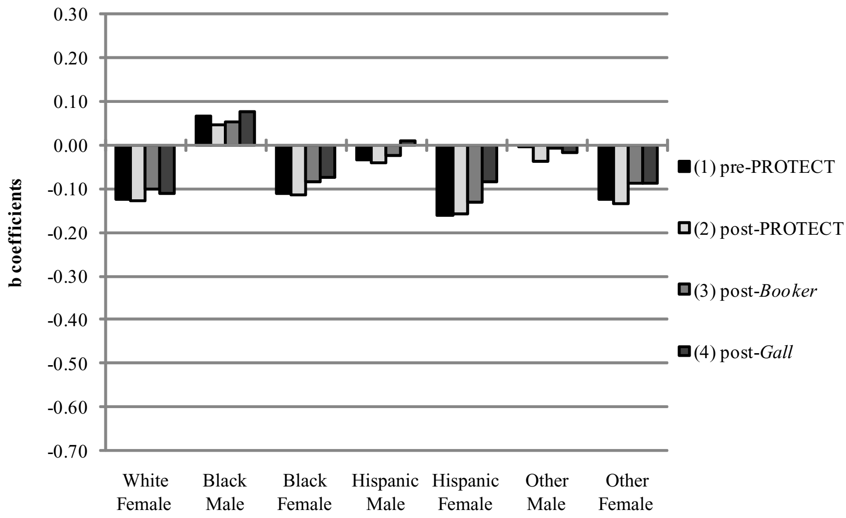
Not only do some forms of disparity show increasing trends over time, but also these effects are large compared with the results published in previous sentencing research. For example, the Black female effect for the post-PROTECT era is –0.422, which corresponds to 34% lesser sentence lengths ($\exp[-0.422] - 1 \times 100 = -34$) compared with White males, net of controls. Interpreted substantively, this effect indicates that at the mean sentence length (62.6 months), Black female offenders receive sentences that are nearly 2 years (21.0 months) less on average compared with their White male counterparts. Given the relatively narrow sentencing ranges in the Guidelines, this is perhaps a shocking result.

10. For parsimony, we display only the results for the race-gender combinations. The full models are shown in Appendix A.

11. The results shown here are not identical to those published by the USSC. We attempted several different coding procedures to try to obtain the exact results of the USSC without success. However, the patterns of results are generally consistent with those published by the USSC, and these differences do not account for the different results we display based on modeling choice. In fact, our results in places display greater disparity than the USSC report. For example, the Black male effects in Figure 2 are slightly greater in the post-PROTECT and post-*Booker* periods than reported by the USSC.

FIGURE 3

Alternative OLS Models of Sentence Length for Incarcerated Offenders



Race-Gender Combinations	(1) pre-PROTECT *Ln Length*[a]	(2) post-PROTECT *Ln Length*	(3) post-*Booker* *Ln Length*	(4) post-*Gall* *Ln Length*
	b	*b*	*b*	*b*
White Male (reference)				
White Female	-0.125[*]	-0.128[*]	-0.099[1,2*]	-0.109[*]
Black Male	0.066[*]	0.045[1*]	0.053[1*]	0.077[2,3*]
Black Female	-0.112[*]	-0.115[*]	-0.084[1,2*]	-0.075[1,2*]
Hispanic Male	-0.034[*]	-0.039[*]	-0.025[*]	0.011[1,2,3]
Hispanic Female	-0.162[*]	-0.156[*]	-0.132[1*]	-0.085[1,2,3*]
Other Male	-0.004	-0.036[*]	-0.005[2]	-0.018
Other Female	-0.123[*]	-0.134[*]	-0.086[*]	-0.087[*]

[1] Coefficient is significantly different from pre-PROTECT Act estimate based on two-tailed *z* test ($p < .05$).
[2] Coefficient is significantly different from post-PROTECT Act estimate based on two-tailed *z* test ($p < .05$).
[3] Coefficient is significantly different from post-*Booker* estimate based on two-tailed *z* test ($p < .05$).
[a] Models include controls for all variables in Appendix B.
$p < .01$.

It is thus important to test the robustness of such findings against reasonable and common alternative modeling strategies.

Figure 3 reports the results from our alternative models of sentence length for all four time periods.[12]

12. Full results are available in Appendix B.

In separate models (not shown), we reran the analysis in Figure 3 without controlling for criminal history to assess the independent effect of excluding criminal history from models of sentence length. We find that criminal history has significant and substantial effects above and beyond the presumptive sentence. A one-unit increase in criminal history results in approximately 4% longer sentences, above that which is already captured by the presumptive sentence measure. Moreover, consistent with our predictions, including criminal history explains a significant portion of the race/ethnicity/gender effects. Put differently, criminal history seems to *mediate* a notable portion of the Black male effect. However, this was similarly true across time periods, and the inclusion or exclusion of criminal history does not change our conclusions about whether racial or ethnic disparity increased post-*Booker*/*Gall*. Across each time period, the racial/ethnic and gender disparities are approximately 20% larger when criminal history is not controlled for, and these effects vary across different racial–gender measures. For example, although criminal history accounts for virtually none of the Black female disparity, Black male disparity is more than 30% larger when a measure of criminal history is not included in the analysis. On the one hand, one could argue that by excluding criminal history from their "refined" models, the USSC may be overestimating racial and gender disparities in all time periods because part of the Black male effect in particular is explained by criminal history. On the other hand, one could argue that the true sentencing disadvantage of Black males is captured by not including criminal history because its Guidelines-based influence should occur through the presumptive sentence. Regardless, criminal history similarly mediates the Black male effect (in particular) across time periods.[13]

Although there are several differences in variable selection compared with Figure 2 (see the Data section for description), the most important difference between Figures 2 and 3 is the choice of dependent variable. In Figure 3, we include only those offenders who actually received a term of incarceration, whereas the USSC models included offenders who did not. The results reported in Figure 3 present a very different view of racial and gender disparity in the wake of *Booker* and *Gall*. First, the sizes of the disparity effects are substantially smaller. For example, whereas the Black male effect was 0.130, 0.089, 0.164, and 0.217 across the four time periods in Figure 2, they are 0.066, 0.045, 0.053, and 0.077 across the time periods in Figure 3. In other words, removing sentences of nonincarceration reduces the effect sizes by approximately 40% in each time period, and this pattern of results is generally consistent for the other racial–gender effects sizes. This reduction is almost entirely a result of removing the nonimprisonment cases—our omission of the alternative confinement cases as sentence lengths does not change the results notably. The USSC's

13. We also examined whether criminal history *moderates* the Black male effect by running supplemental analyses (available on request) that interacted criminal history by each race–gender dummy variable. We found a small moderation whereby the effect of criminal history was slightly greater for Black males (the Black male × Criminal history interaction term coefficients were as follows: pre-PROTECT = .005, PROTECT = .006, post-*Booker* = .004, post-*Gall* = .007. However, the differences in this interaction term across time periods were not statistically significant.

decision to include alternative confinement cases as sentence lengths therefore seems to be of negligible importance.

The second notable change in the pattern of results is the trends in effects over time. Although the Black male effect shows a similar pattern as Figure 2, where the effect dipped in the post-PROTECT era only to increase post-*Booker* and *Gall*, *z* tests show that the post-*Booker* effect is actually significantly *less than* the pre-PROTECT era, and there is no significant difference between the effects post-*Gall* and pre-PROTECT. In short, Black male disparity returned to the pre-PROTECT state in the wake of *Gall*. For other racial–gender effects, it seems that there is actually less disparity in sentence lengths after *Booker* and *Gall*. For example, compared with White males, the effects for Black females, Hispanic males, and Hispanic females are actually significantly *less* in the *Booker* and *Gall* periods than in the PROTECT era. In no case does it seem that racial–gender length disparities have substantially increased since *Booker* and *Gall*. Our findings, however, do raise serious questions about why our results differ from those of the USSC 2010 report. Because our dependent variable includes only terms of imprisonment, does this mean that disparity in the incarceration decision has increased after *Booker* and *Gall*? We explore this question in the next section.

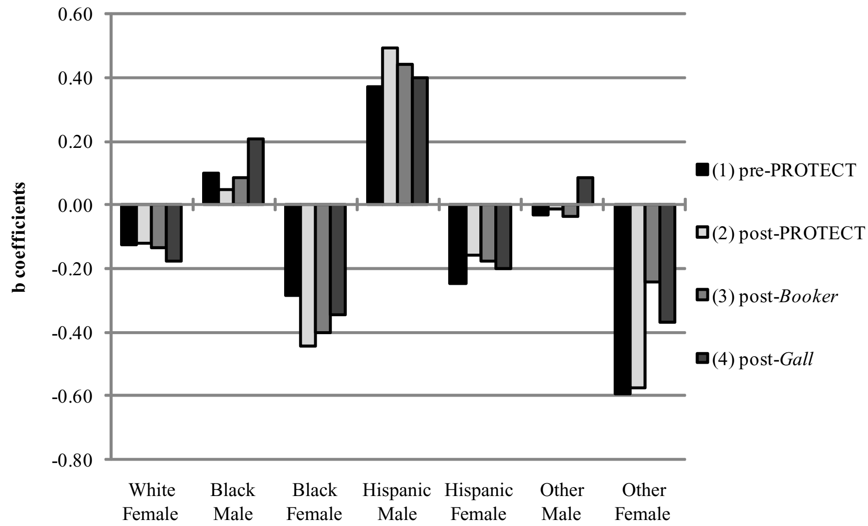
Incarceration Decisions

Figure 4 reports the results of logistic regression models of the whether the offender was sentenced to prison regressed on the same independent variables reported in Appendix B.[14] The results offer mixed support for whether incarceration disparities have increased over time. Although the White female effect seems to have increased slightly, going from –0.125 in the pre-PROTECT era to –0.175 post-*Gall*, *z* tests show that none of the time period effects is significantly different from each other. This pattern of nonsignificant (or marginally significant) differences is generally true for nearly all the other race–gender effects as well but with one exception, the Black male effect. Consistent with the pattern of results for the sentence-length decision, Black male disparity decreased in the PROTECT era but then increased in the *Booker* and especially the *Gall* time periods. *z* tests confirm that post-*Gall* Black male *imprisonment* disparity is greater than in the previous time periods. These results explain the difference in sentence-length disparity between the USSC report and our models. Whereas they interpret their findings as increases in sentence-length disparity, we show that some differences in effects are actually caused by increased disparity in the incarceration decision. By including imprisonment and length decisions into the USSC's dependent variable, these two distinct patterns of results become conflated. Such results raise questions about the extent to which our sentence-length models in Figure 3 are biased by selection.

14. The only difference between the predictors in Figure 3 and 4 is the presumptive sentence variable is unlogged in Figure 4. The full table of incarceration results is available from the authors on request.

FIGURE 4

Logistic Regression Models of Incarceration



	(1) pre-PROTECT	(2) post-PROTECT	(3) post-*Booker*	(4) post-*Gall*
Race-Gender Combinations	*Ln Length*[a]	*Ln Length*	*Ln Length*	*Ln Length*
	b	*b*	*b*	*b*
White Male (reference)				
White Female	-0.125*	-0.122	-0.136*	-0.175*
Black Male	0.101	0.046	0.084	0.209[2,3,*]
Black Female	-0.283*	-0.444*	-0.401*	-0.346*
Hispanic Male	0.369*	0.492*	0.441*	0.400*
Hispanic Female	-0.245*	-0.158	-0.177*	-0.199*
Other Male	-0.032	-0.013	-0.038	0.084
Other Female	-0.592*	-0.573*	-0.240[1,2*]	-0.368*

[1] Coefficient is significantly different from pre-PROTECT Act estimate based on two-tailed z test ($p < .05$).

[2] Coefficient is significantly different from post-PROTECT Act estimate based on two-tailed z test ($p < .05$).

[3] Coefficient is significantly different from post-*Booker* estimate based on two-tailed z test ($p < .05$).

[a] Models include controls for all variables in Appendix B.

$p < .01$.

We ran our sentence length models both with and without a Heckman two-step correction factor, as discussed. For the purposes of the Heckman correction, we attempted to find exclusion restrictions and to estimate an incarceration model that was substantively different from the sentence-length model. This was difficult because most variables that significantly predict imprisonment also predict length, although the strength of the effects is sometimes different. Nonetheless, our selection model included a dummy variable for presumptive disposition (despite whether the Guidelines-recommended imprisonment), instead of Guidelines minimum. The selection model also omitted defendant education because this did not exert significant effects on imprisonment in the selection probit model. The Black male length effects in the Heckman corrected model are as follows (all are significant at $p < .001$): pre-PROTECT = .063, PROTECT = .041, post-*Booker* = .048, and post-*Gall* = .08). In the Heckman corrected models, the Hispanic male effect was –.03 pre-PROTECT and –.02 PROTECT, and –.016 post-*Booker* and .02 post-*Gall*. The other race/gender effects are comparable with those in the results we present.

In supplemental analyses, we also included all cases for the four time periods together in models of incarceration and length, and we included interaction terms for each race/ethnicity/gender variable times each time period (with pre-PROTECT left out as a reference category). These terms then allow us to include all cases together in one model with the same error structure and to examine differences in Black male effects, for example, across time periods in the same model.[15] The results generally corroborate what we present in our alternative analyses previously in that sentence-length disparity for Black or Hispanic males is not significantly greater post-*Booker* or post-*Gall* than the pre-PROTECT or PROTECT eras. In fact, in the full time period interaction models, the increased incarceration odds post-*Gall* for Black males do not attain statistical significance. Furthermore, these models show that the Black male sentence-length effects post-*Booker* and post-*Gall* are slightly but significantly *less* than that in the pre-PROTECT era, whereas our models in Figure 3 show no significant differences between the pre-PROTECT and post-*Booker*/post-*Gall* Black male effects. The Black male × Post-*Booker* coefficient is –.023, and the Black male × Post-*Gall* coefficient is –0.017. Both indicate relatively small differences and are likely significant primarily because of the much larger number of cases in our combined-years model. Thus, the safest thing to say from our analyses is that the levels of sentence-length disparity affecting Black males seems to be nearly identical pre-PROTECT Act, post-*Booker*, and post-*Gall*. However, we present the separate models here as our main analysis for comparability with the USSC 2010 report, and these separate models make it easier to compare each time period with one another.

We also estimated tobit models that combined nonimprisonment and imprisonment sentences, treating 0 as a censoring point. The Black male effects are as follows (all are significant at $p < .001$): pre-PROTECT = .064, PROTECT = .041, post-*Booker* = .05, and post-*Gall* = .07. In the tobit models, the Hispanic male effect changed from –.04 pre-PROTECT and PROTECT, to –.03 post-*Booker* and .008 (not significant) post-*Gall*.

15. We would like to thank several helpful reviewers for suggesting this alternative modeling approach.

TABLE 1

Alte rnative Mode ls of Sentence Length and Incarce ration WITHOUT Immigration Offenses

	(1) pre-PROTECT	(2) post-PROTECT	(3) post-*Booker*	(4) post-*Gall*
Sentence Length [a]	**b**	**b**	**b**	**b**
White Male (reference)				
White Female	−0.107[*]	−0.113[*]	−0.077[1,2,*]	−0.067[1,2,*]
Black Male	0.055[*]	0.028[1,*]	0.035[1,*]	0.040[1,*]
Black Female	−0.097[*]	−0.105[*]	−0.068[1,2,*]	−0.059[1,2,*]
Hispanic Male	−0.014	−0.033[1,*]	−0.013[2,*]	0.005[1,2,3]
Hispanic Female	−0.148[*]	−0.140[*]	−0.103[1,2,*]	−0.043[1,2,3,*]
Other Male	−0.005	−0.027	0.006[2]	−0.012
Other Female	−0.123[*]	−0.114[*]	−0.052[1,2,*]	−0.050[1,2,*]
N	60,226	46,400	121,625	75,209
Adjusted R2	0.820	0.832	0.847	0.849
	(1) pre-PROTECT	**(2) post-PROTECT**	**(3) post-*Booker***	**(4) post-*Gall***
Incarceration [a]	**b**	**b**	**b**	**b**
White Male (reference)				
White Female	−0.119	−0.127	−0.136[*]	−0.166[*]
Black Male	0.126[*]	0.082	0.113[*]	0.228[2,3,*]
Black Female	−0.240[*]	−0.407[1,*]	−0.399[1,*]	−0.338[*]
Hispanic Male	0.321[*]	0.407[*]	0.405[*]	0.389[*]
Hispanic Female	−0.157	−0.074	−0.107	−0.122
Other Male	0.035	0.041	0.009	0.153
Other Female	−0.497[*]	−0.514[*]	−0.186[1,2]	−0.279
N	73,897	56,578	146,620	91,080
−2 log likelihood	31368.3	21318.6	50975.2	33249.8

[1] Coefficient is significantly different from pre-PROTECT Act estimate bas ed on two-tailed z-test ($p < .05$).
[2] Coefficient is significantly different from post-PROTECT Act estimate based on two-tailed z-test ($p < .05$).
[3] Coefficient is significantly different from post-Booker estimate based on two-tailed z-test ($p < .05$).
[a] Models include controls for all variables in Appendix B.
$p < .01$.

Immigration Offenses

With immigration crimes accounting for more than 25% of all federal sentences in 2009, immigration offenses are an important component of federal sentencing. However, as stated, these offenses offer unique challenges to researchers interested in comparability with other crimes, across time, and across federal courts. In Table 1, we evaluate whether immigration offenses have played a role in changing racial-gender disparity since *Booker* and *Gall* in both the incarceration and sentence-length decisions.

The results in Table 1 show the racial–gender effects for all four time periods across the incarceration and sentence-length decisions, excluding immigration offenses. For parsimony, we report only the race/ethnicity–gender effects (the full tables are available on request).

The results for the trends in disparity in sentence-length decisions, compared with those reported in Figure 2, show that a substantial amount of racial–gender disparity can be attributed to immigration offenses. For each racial–gender effect across the four time periods, immigration offenses alone account for roughly 25% of the effect size. However, the impact of immigration offenses varies substantially across groups, accounting for roughly 40% of the Hispanic male and Black male effects but for only 10% of the other male effect. These results show clearly that immigration offenses offer unique challenges to the federal criminal justice system. Even though Hispanics comprise the overwhelming majority of immigration offenders, the inclusion of these offenses without properly accounting for the degree of interdistrict variation that goes along with the unique district policies (i.e., use of fast-track departures) used to deal with them results in greater estimates of racial–gender length disparity than would be the case if immigration offenses were excluded.

The impact of immigration offenses seems to have only a modest effect on incarceration disparity. Whereas excluding immigration offenses actually shows slightly (although non-significant) increases in Black male disparity, immigration crimes account for approximately 10% of the Hispanic male and more than 40% of the Hispanic female disparity effects.

Advisory versus Mandatory Guidelines: A Broader Time Comparison

So far, our models have found little substantive change in sentence-length disparities based on race and gender when comparing the pre-PROTECT era with the post-*Booker* and *Gall* eras, but there has been an increase in Black male incarceration disparity. Also, roughly a quarter of all racial–gender disparities can be attributed to immigration offenses, likely resulting from the distinct methods certain federal districts use to handle the dramatic increase in immigration crimes.

As discussed, some commentators have claimed that increasing disparities post-*Booker* are caused by the increased discretion afforded judges (see, e.g., Gonzales, 2005; Richter, 2008), and to prevent such disparity, the Guidelines need to be made mandatory once again. However, none of the critics of the new advisory system have demonstrated that there was actually less disparity during the many years when the Guidelines were mandatory. More to the point, are the PROTECT Act period and pre-PROTECT years since 2000 the only relevant comparisons? What about the many years prior to the PROTECT Act when the Guidelines were also mandatory?

Since 1996, considerable "back-and-forth" struggling has occurred between the Supreme Court and Congress about the proper amount of judicial discretion at sentencing (see Stith [2008] for a detailed discussion). The Supreme Court decision in *Koon v. United States* (1996) was a watershed in this struggle, and the aftermath of this decision eventually led to Congress's attempts to restrict judicial sentencing discretion even more strongly with the PROTECT Act (see Stith, 2008). Recall that in *Koon*, the Supreme Court held that departure decisions made by district judges should be given due deference by appellate courts

TABLE 2

Sentence Length and Incarce ration Models Comparing Pre–Koon with post-Booker and post-Gall

Sentence Length[a]	(1) pre-*Koon* *Ln Length* *b*	(2) post-*Booker* *Ln Length* *b*	(3) post-*Gall* *Ln Length* *b*
White Male (reference)			
White Female	−0.200[*]	−0.110[1,*]	−0.109[1,*]
Black Male	0.122[*]	0.057[1,*]	0.077[1,*]
Black Female	−0.107[*]	−0.102[*]	−0.075[*]
Hispanic Male	0.012	−0.015[1,*]	0.011
Hispanic Female	−0.125[*]	−0.130[*]	−0.085[1,*]
Other Male	0.017	−0.007	−0.018 1
Other Female	−0.109[*]	−0.097	−0.087[*]
Incarceration[a]	**(1) pre-*Koon* *Incarceration* *b***	**(2) post-*Booker* *Incarceration* *b***	**(3) post-*Gall* *Incarceration* *b***
White Male (reference)			
White Female	−0.208[*]	−0.156[*]	−0.207[*]
Black Male	−0.008	0.078	0.175[1,*]
Black Female	−0.558[*]	−0.500[*]	−0.500[*]
Hispanic Male	0.489[*]	0.588[*]	0.437[*]
Hispanic Female	−0.419[*]	−0.200[1,*]	−0.282[*]
Other Male	−0.131	−0.053	−0.018
Other Female	−0.734[*]	−0.332[1,*]	−0.498

[1] Coefficient is significantly different from pre-*Koon* estimate based on two-tailed z-test ($p < .05$).
[a] Models include controls for all variables in Appendix B with the exception of "Pre-Sentence. Detention" because this information was not collected in the pre-*Koon* data.
$p < .01$.

and established that departures by judges should be examined by an "abuse of discretion" standard.

Thus, prior to *Koon* and its modest relaxation of restrictions on judges' ability to depart from Guidelines, the Guidelines were arguably more "mandatory" than at any other point in their history except perhaps the PROTECT era. We, therefore, use federal sentencing data from fiscal years 1994 and 1995 as a comparison time period versus post-*Booker* and post-*Gall*. If there is less disparity in the pre-*Koon* time period compared with *Booker* and *Gall*, this might mean that the post-*Booker* environment of advisory Guidelines has fostered greater disparity, and it would support calls for renewed restrictions on judicial discretion.

Table 2 shows the results for sentence length and incarceration decisions in the pre-*Koon*, post-*Booker*, and post-*Gall* time periods.

For parsimony, we display only the results for the racial–gender effects (full tables are available on request; note that the post-*Booker/Gall* effects are not identical to Figures 3 and 4 because we had to omit presentencing detention as a predictor.)[16] Beginning with the sentence-length results, it seems that the post-*Booker* and post-*Gall* disparities are considerably *less than* those found prior to *Koon*. Indeed, the White female, Black male, and Hispanic female effects are significantly less in either the *Booker* or *Gall* periods than prior to *Koon*. In no instance has there been a significant increase in sentence-length disparities since *Koon*. Put simply, racial and gender sentence-length disparities are less today, under advisory Guidelines, than they were when the Guidelines were arguably their most rigid and constraining.

However, disparities in the incarceration decision show considerably more stability among the three time periods. Of the seven racial–gender effects shown, three of them (Black female, Hispanic male, and other male) show no significant changes, two effects display significant reductions in disparity (Hispanic and other females), and only one effect shows a significant increase in disparity (Black male); this latter finding is specific to the post-*Gall* period.

These findings call into question the notion that mandatory guidelines, per se, result in reduced racial and gender disparities. Although we do find that Black male incarceration disparity has increased post-*Gall* compared with the pre-*Koon* period, we also find that Black male sentence-length disparity has been reduced considerably. Moreover, of the 14 racial–gender effects shown in Table 2 (seven effects across two sentencing decisions), seven show that post-*Booker* or post-*Gall* disparities are significantly *less than* those found prior to *Koon*, and the other six effects in general display slight (although nonsignificant) reductions in disparity.

It should be noted that incarceration and sentence-length decisions are not the only punishment decisions that judges make. In fact, much of the political and legal controversies surrounding the Guidelines have pertained to departures, and previous research has showed that departures from the Guidelines are a locus of disparity in federal courts (Albonetti, 1998; Johnson et al., 2008; Mustard, 2001). As our final test of whether *Booker* and *Gall* have resulted in greater disparities, Figure 5 reports the results from logistic regression models of whether an offender received a nonsubstantial assistance downward departure from the Guidelines.[17]

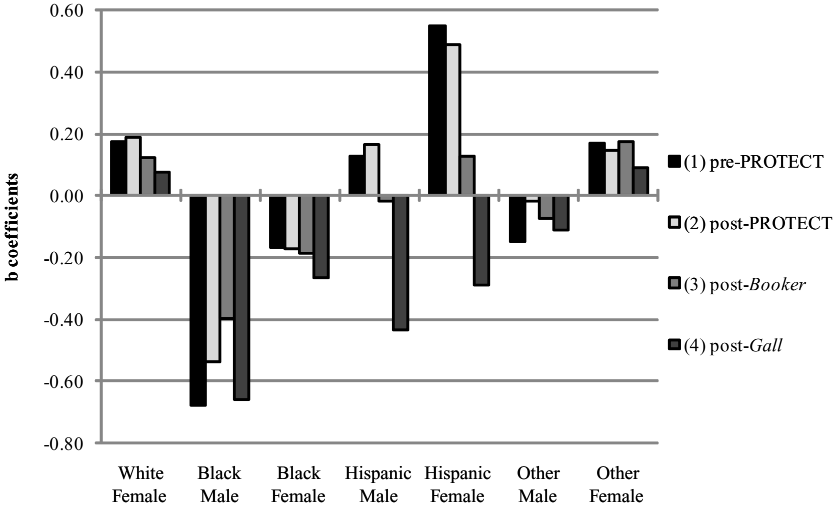
In all time periods, there is evidence of racial and gender disparity. Black males and females are less likely to receive an "other" downward departure compared with their White

16. All models in Table 2 include all variables from Figures 3 and 4 except whether the offender was detained pending sentencing. Although this variable is shown to have an effect on racial and gender disparities, information on this measure is not available in USSC data in the pre-*Koon* period. Thus, to compare across time periods directly, this measure was removed.

17. Full models are shown in Appendix C.

FIGURE 5

Logistic Regression Models of "Other" Downward Departures



	(1) pre-PROTECT	(2) post-PROTECT	(3) post-*Booker*	(4) post-*Gall*
Race-Gender Combinations	*Downward Dep*[a]	*Downward Dep*	*Downward Dep*	*Downward Dep*
	b	*b*	*b*	*b*
White Male (reference)				
White Female	0.173[*]	0.189[*]	0.122[*]	0.077
Black Male	-0.678[*]	-0.538[1,*]	-0.397[1,2,*]	-0.661[2,3,*]
Black Female	-0.167[*]	-0.172	-0.185[*]	-0.266[*]
Hispanic Male	0.130[*]	0.167[*]	-0.019[1,2,]	-0.432[1,2,3,*]
Hispanic Female	0.551[*]	0.486[*]	0.129[1,2*]	-0.290[1,2,3,*]
Other Male	-0.147	-0.017	-0.073	-0.111[*]
Other Female	0.168	0.145	0.177[*]	0.089

[1] Coefficient is significantly different from pre-PROTECT Act estimate based on two-tailed *z* test ($p < .05$).

[2] Coefficient is significantly different from post-PROTECT Act estimate based on two-tailed *z* test ($p < .05$).

[3] Coefficient is significantly different from post-*Booker* estimate based on two-tailed *z* test ($p < .05$).

[a] Models include controls for all variables in Appendix C.

$p < .01$.

male counterparts, net of controls, whereas White females are more likely to receive this form of sentencing discount. The trends in the effects, however, do not show that *Booker* and *Gall* have increased disparity substantially. *z* tests show that none of the Black female effects are significantly different across time, and this is true for the White female effects as well. The likelihood of a Black male receiving this sentencing discount actually improved significantly in the post-*Booker* period compared with the pre-PROTECT and post-PROTECT eras. However, since *Gall*, this form of sentencing disparity has returned to the effect found in the pre-PROTECT time period, as evidenced by the nonsignificant z score between times (4) and (1). Interestingly, the disparity against Hispanic males and females seems to have increased in the post-*Gall* time period only. Whereas Hispanic offenders (male and female) were slightly more likely to receive an "other" downward departure in the pre-PROTECT and post-PROTECT periods, they are significantly less likely to receive this sentencing discount post-*Gall*.

In all, these results do not show that *Booker* and *Gall* produced greater disparity in the likelihood of minority offenders to receive nonsubstantial assistance departures. In most cases, the disparities returned to the pre-PROTECT effect sizes, although Hispanic disparity does seem to have increased since *Gall*. It should be noted, however, that this pattern of increased disparity against Hispanics is also true for substantial assistance departures. Figure 6 reports the results for the likelihood of receiving a substantial assistance departure across the four time periods.[18]

Again, there is clear disparity in the application of these departures in all time periods, specifically for Black and Hispanic males. Similar to the results in Figure 5, we find that Black male disparity increased in the post-*Gall* period, but this effect is not significantly different than in the pre-PROTECT period. This is not the case for Hispanic males, who have witnessed a significant increase in disparity post-*Gall* compared with all other time periods. These results are substantively important because, whereas most commentary on the effects of *Booker* and *Gall* has focused on how judges have reacted to their newfound discretion, little has been mentioned about how these cases may affect prosecutorial behavior. The results in Figure 6 suggest that disparity against Hispanic males in the prosecutorial use of substantial assistance departures has increased considerably since *Gall*.

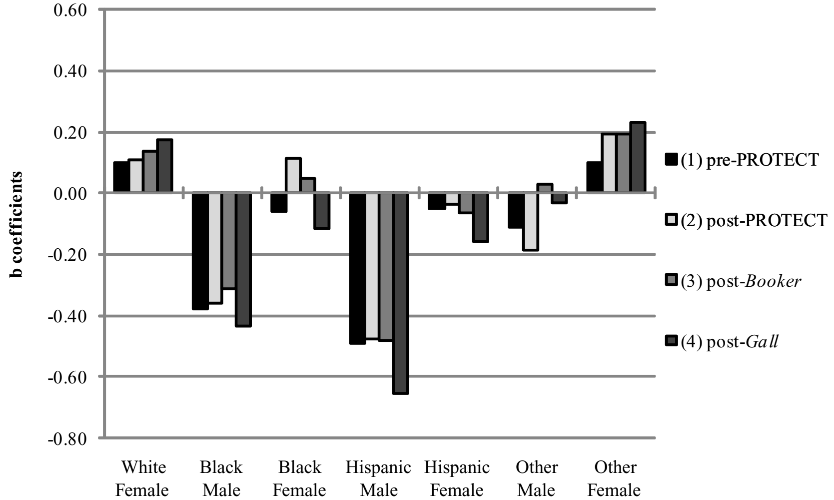
The idea that prosecutorial discretion, as opposed to judicial discretion, has been more of a locus of disparity in the wake of *Booker* and *Gall* receives additional support in Table 3.

Prior to *Booker*, the USSC did not keep detailed information on different types of downward departures except to indicate substantial assistance departures. However, government-sponsored downward departures and even fast-track departures were around long before *United States v. Booker*. After *Booker*, the USSC began keeping more detailed information on these specific departures types. Thus, in Table 3, we model the likelihood of receiving a judge-initiated downward departure, a government-sponsored downward departure, or a fast-track departure in the post-*Booker* and *Gall* time periods. The first part of the table shows that Black and Hispanic males are particularly disadvantaged in their likelihood of receiving a judge-initiated departure, and both forms of disparity have become significantly worse post-*Gall* compared with post-*Booker*. These results lend some support to those who claim that judges have used their newfound discretion in discriminatory ways. However, it is important to note that these effect sizes for "true" judge-initiated departures

18. Figure 6 includes all controls shown in Appendix C. Full tables are available from authors on request.

FIGURE 6

Logistic Regression Models of Substantial Assistance (5K1.1) Departures



	(1) pre-PROTECT	(2) post-PROTECT	(3) post-*Booker*	(4) post-*Gall*
Race-Gender Combinations	*5K1.1 Dep*	*5K1.1 Dep*	*5K1.1 Dep*	*5K1.1 Dep*
	b	*b*	*b*	*b*
White Male (reference)				
White Female	0.101*	0.109	0.135*	0.175*
Black Male	-0.378*	-0.361*	-0.313*	-0.436[3,*]
Black Female	-0.057	0.112[1]	0.050	-0.116[2,3]
Hispanic Male	-0.489*	-0.474*	-0.482*	-0.656[1,2,3,*]
Hispanic Female	-0.050	-0.037	-0.066	-0.157*
Other Male	-0.113	-0.186*	0.029[2]	-0.030
Other Female	0.102	0.194	0.195*	0.233

[1] Coefficient is significantly different from pre-PROTECT Act estimate based on two-tailed z test ($p < .05$).

[2] Coefficient is significantly different from post-PROTECT Act estimate based on two-tailed z test ($p < .05$).

[3] Coefficient is significantly different from post-*Booker* estimate based on two-tailed z test ($p < .05$).

[a] Models include controls for all variables in Appendix C.

$p < .01$.

are considerably less than those found in Figure 4, which contained all three of the different nonsubstantial assistance departure types combined into the "other" downward departure category. For example, in both the post-*Booker* and post-*Gall* periods, the effect for Black males in judge-initiated departures is only half the size of the effect for "other" downward departures, which suggest that prosecutor-sponsored departures are responsible for the other half. Moreover, the disparity against African Americans is considerably greater in both forms of government-sponsored departures, and it has increased to a greater extent since *Gall*.

TABLE 3

Logistic Regression Models of Different Departure Types in the post-Booker Era

	(1) post-*Booker*	(2) post-*Gall*
Judge Initiated Departure[a]	**b**	**b**
White Male (reference)		
White Female	0.124*	0.077
Black Male	−0.200*	−0.341^{1,*}
Black Female	−0.020	−0.050
Hispanic Male	−0.213*	−0.308^{1,*}
Hispanic Female	0.119*	−0.014^{1}
Other Male	−0.024	0.015
Other Female	0.044	0.091
	(1) post-*Booker*	**(2) post-*Gall***
Government Sponsored Dep.[a]	**b**	**b**
White Male (reference)		
White Female	0.045	−0.061
Black Male	−0.299*	−0.599^{1,*}
Black Female	−0.499*	−0.438*
Hispanic Male	0.264*	0.110^{1}
Hispanic Female	0.561*	0.317^{1,*}
Other Male	0.353*	0.474*
Other Female	0.473*	0.689*
	(1) post-*Booker*	**(2) post-*Gall***
Fast-Track Departures[a]	**b**	**b**
White Male (reference)		
White Female	0.274*	0.420*
Black Male	−2.580*	−3.595^{1,*}
Black Female	−2.200*	−2.902*
Hispanic Male	0.059	−1.118^{1,*}
Hispanic Female	0.148*	−1.148^{1*}
Other Male	−0.511*	−1.184^{1,*}
Other Female	0.784*	0.104^{1}

[1]Coefficient is s ignificantly different from post-*Booker* estimate based on two-tailed z-test ($p < .05$)
[a]Models include controls for all variables in Appendix C.
$p < .01$.

Taken together, although much scholarly attention has been devoted to the changes in judges' discretion, the results in Figure 5, Figure 6, and Table 3 suggest that the post-*Booker* and post-*Gall* eras have observed equal or greater changes in prosecutorial behavior.

Conclusions

If a primary goal of federal sentencing reform was a reduction of unwarranted disparity, the impact of the *Booker/Fanfan* decision on disparity is among the most important empirical questions facing sentencing policy makers (Hofer, 2007). Indeed, U.S. Attorney General Eric Holder (2009) emphasized the need for such research in recent remarks to Congress. We have provided an alternative to and extension of the USSC's 2010 report, which found that sentence-length disparity affecting Black males has increased relative to the PROTECT era. Our sentence-length findings differ in important respects with the USSC 2010 report, and our analyses go beyond theirs to provide a more extensive and fine-grained analysis of different sentencing decisions where disparity affecting Black males (and others) may occur.

First, in analytically separating the imprisonment decision from the length decision, we find that a considerable part of the USSC's Black male disparity findings are attributable to their analyses' combining of the imprisonment and length decisions into one model (and using OLS regression, rather than other options such as tobit models). We find that Black male incarceration odds have stayed relatively stable from pre-PROTECT up through post-*Booker*. Interestingly, the pre-PROTECT, PROTECT, and post-*Booker* periods show greater imprisonment decision disparity affecting Black males than the period before the 1996 *Koon* decision. However, Black male imprisonment odds do increase significantly post-*Gall*. This post-*Gall* increase in Black males' odds of imprisonment plays a big part in driving the USSC's findings of greater Black male sentence-length disparity. We have in fact shown that *post-Gall increases in Black male disparity are specific to the imprisonment decision* and not to sentence-lengths. This indicates that it matters a great deal for questions of disparity how one defines sentence outcome variables, and how one deals with selection into imprisonment, as well as the issue of censoring.

Second, we find that post-*Gall* sentence-length disparity disadvantaging Black males has increased significantly only with respect to the PROTECT era, and *not* in comparison with earlier periods. Notably, the post-*Booker*/*Gall* levels of Black male sentence-length disparity are lower than in the pre-*Koon* period, in addition to the pre-PROTECT era. Thus, one concludes that the post-*Booker* era has brought greater sentence-length racial disparity disadvantaging Black males *only* when one's basis of comparison is the PROTECT era.

Regarding racial disparity, the truly unusual period in the history of the Guidelines may be the PROTECT Act era, rather than the post-*Booker*/*Gall* eras. Taking the long view, the relatively low levels of disparity in the PROTECT period were an anomaly compared with the earlier years when the Guidelines were also mandatory (particularly the pre-*Koon* period), as well as the post-*Booker* years. If post-*Booker*/*Gall* racial/gender length disparity levels were comparable with or lower than levels in previous periods when the Guidelines were also mandatory, this calls into question the notion that the post-*Booker/Gall* eras of advisory Guidelines have produced *uniquely high* levels of racial disparity in sentence lengths. In our view, this also calls into question the need for blanket policy remedies that would attempt to curtail overall judicial sentencing discretion in the name of reducing disparity in sentence lengths.

Third, criminal history has an effect on sentences independent of Guidelines presumptive sentences, and criminal history *mediates* a notable portion of the Black male effect. The implications of this mediation should be considered further. On the one hand, Black males on average have higher criminal histories compared with White males (the Black male–criminal history correlation across all years is 0.25, whereas the White male–criminal history correlation is –.29). Thus, it could be argued that criminal history, even its discretionary consideration beyond its influence in establishing the presumptive sentence, captures "real" legally relevant differences—differences that are not attributable to race/gender categories. Therefore, the "true" degree of disparity is that which is left over after criminal history is taken into account. However, one could argue that the consideration of criminal history itself disadvantages Black male defendants and that the "true" degree of sentencing disadvantage is that which is produced by courts' discretionary consideration of criminal history beyond its influence on presumptive sentence because the consequences of such consideration fall harder on Black males.[19] Furthermore, the criminal records of Black males may themselves be the product of discriminatory processes and may be viewed subjectively and counted as more serious than those of other offenders. Although answering these questions is beyond the scope of this current article, it is safe to say that controlling for presumptive sentence and criminal history likely produces lower bound estimates of racial disparity, and our results suggest a need for future research and policy discussions about the consideration of criminal history, how it should be modeled properly, and its differential impact on Black males. It should be noted, however, that criminal history mediates the Black male effect similarly across each time period and, thus, does not account for the different trends in disparity between our analysis and the USSC 2010 report.

Fourth, a substantial portion of the sentence length disparity affecting Black males across time is attributable to *immigration offenses*, especially for the post-*Booker/Gall* period. When immigration offenses are removed from the models, Black male sentence-length disadvantage is notably less than when immigration offenses are included. What is more, when immigration offenses are removed, there is actually significantly *less* length disparity affecting Black males in the post-*Booker/Gall* periods than in the pre-PROTECT era. Incidentally, this is also true when we use the USSC model specification as in Table 1 without immigration offenses (available on request).

Booker and especially *Gall* gave judges more freedom to deviate from the Guidelines. Thus, if judges sentence Black males increasingly more severely compared with others post-*Booker/Gall*, logically we should view greater disparity affecting Black males (or others) in downward departures/deviations. We observe no such increase for Black males compared with the pre-PROTECT era. Black females are less likely to receive overall downward departures post-*Gall* (although not significantly), but there is no significant Black female

19. As one reviewer noted, "we need to avoid kitchen sink models when looking for racial disparity."

disadvantage in *judge-initiated* deviations (whereas there is for government-sponsored deviations). We also observe a greater Black male disadvantage in substantial assistance and government-sponsored departures than in judge-initiated departures. Furthermore, Hispanic males are significantly less likely to receive overall downward departures post-*Gall* compared with the pre-PROTECT and PROTECT eras. However, a substantial portion of this post-*Gall* disparity in overall downward departures affecting Hispanic males seems to be caused by dramatic post-*Gall* declines in the likelihood of Hispanic males receiving government-sponsored and fast-track departures, two decisions influenced heavily by prosecutors. Overall, the departure findings do not point to unique and comparatively large post-*Booker*/*Gall* racial/ethnic disparities in *judge-initiated* Guidelines deviations. In fact, greater disparity affecting Black and Hispanic males characterizes departures decisions heavily influenced by prosecutors more than judge-initiated departures.

Must the Guidelines be mandatory to be influential and to constrain disparity? Furthermore, why might *Booker* and *Gall not* have resulted in increased disparity? Perhaps the Guidelines serve a norm-setting function (Kramer, 2009) and have become embedded in the organizational and legal culture of federal courts. As Reitz (2005) observed, the Guidelines continue to structure federal sentencing in the aftermath of *Booker*—courts must continue to calculate and consider them and must provide legally defensible reasons for deviating from them. Furthermore, state court sentencing guidelines, such as Pennsylvania, Minnesota, Washington, Florida, and others have never been mandatory, and the federal Guidelines now have a legal status similar to such state sentencing guidelines (Kramer and Ulmer, 2009). Evidence exists that a major reason Pennsylvania's guidelines were influential was their norm-setting function: They became embedded in local court communities as taken-for-granted decision tools (Kramer and Ulmer, 2009; Ulmer, 1997). Although sentencing disparities affecting Black and especially Hispanic males, particularly in incarceration decisions, still exist under Pennsylvania's guidelines, these disparities have been reduced over time (Kramer and Ulmer, 2009).

Our study is certainly not the last word on the impact(s) of *Booker* and its aftermath on federal sentencing. We need to monitor levels of disparity continually, and our analysis raises some troubling questions. What accounts for the increase in the imprisonment odds of Black males post-*Gall*? What is responsible for the greater racial disparity among immigration cases, which are clustered in certain districts and processed in distinctive ways? Additional research on the role of race in immigration cases is needed. Why have Hispanic males become so much less likely to receive government-sponsored and fast-track departures post-*Gall*? We cannot answer these questions, but future research should continue to monitor more nuanced effects of *Booker* and *Gall* by evaluating sentencing outcomes for specific types of offenders, offenses, and specific decisions.

A chief limitation in our study is our inability to address disparities that might occur in earlier stages of case processing, such as charging and conviction processes. Our major goal in this article was to address important implications raised by the USSC 2010 report, which focused on disparity in sentence lengths, and thus sentencing-stage discretion. However, we are acutely aware that prosecutors have always played a crucial role in federal sentencing, especially under the Guidelines. Our departure analyses differentiated substantial-assistance, government-sponsored, and fast-track departures from judge-initiated departures. However, offenders' exposure to Guidelines punishments is to a great extent a product of prosecutors' charging decisions and the plea agreement process, in which negotiated stipulations about Guidelines-relevant conduct and offense-specific behavior (which raise or lower the final offense level) are commonplace (Shermer and Johnson, 2010; Ulmer, 2005; Ulmer et al., 2010). It has been long recognized that changes in sentencing schemes affect the distribution of discretion among court actors (Engen, 2009; Reitz, 1998), and it is likely that federal prosecutors' decisions and behavior in the charging and plea agreement process have changed significantly in the wake of *Booker* and have changed in nonuniform ways. Some evidence for this is found in our results for substantial-assistance, government-sponsored, and fast-track departures. Our lack of presentence stage data means that we may be understating overall, process-wide disparity stemming from prosecutors' charging decisions and plea agreement behavior. If prosecutors have exhibited a greater tendency to consider extralegal factors in their charging decisions and in their plea agreement concessions in the post-*Booker* periods, our analyses would be unable to detect it.

The USSC 2010 report points to greater sentence-length disparity affecting Black males in the post-*Booker/Gall* periods, although it does not claim that *Booker* and *Gall* caused this increase. We have no wish to impugn the USSC or its commendable attention to the issue of unwarranted disparity. However, based on our differing results using alternative procedures that are reasonable in light of prior federal sentencing literature, as well as our analysis of Guidelines departures, we question the notion that *Booker* and *Gall* have caused increases in race/ethnic and gender sentence-length disparity compared with the full range of years when the Guidelines were mandatory.

We do find an unexplained increase in Black males' odds of imprisonment post-*Gall*, an empirical possibility that the *Booker* report cannot discern. There also seems to be notable disparity affecting Black males in immigration cases. These specific situations warrant additional scrutiny and perhaps discussions of policy changes targeted specifically to those two circumstances. Consideration of where disparities occur is fundamentally important to policy makers because, depending on where disparities are most prevalent, policy solutions differ. For example, based on the 2010 *Booker* report, a policy observer may favor restricting the sentencing ranges in the Guidelines table to reduce the amount of sentence-length disparity. However, if the bulk of disparity is located in the incarceration decision, such a "solution" would be misguided and would do little to help reduce this form of inequality. The same can be said for suggestions to "mandatorize" the Guidelines (see the Consumer Privacy Protection Act of 2005) to reduce sentencing disparities. In addition, if immigration cases are a particularly glaring locus of sentence-length disparity, but other kinds of offenses are not, then attention might be paid to the causes of such disparity and solutions drafted to target immigration cases. We argue that there is insufficient empirical support for broad-based policies, such as the Consumer Privacy Protection Act, that would globally constrain federal judges' sentencing discretion as a remedy for disparity. Such a policy would not only be a blanket, blunt instrument solution to fairly specific loci of disparity but also would do nothing about prosecutorial decisions that affect sentencing outcomes (i.e., substantial-assistance, government-sponsored, and fast-track departures), which we have shown to be as great or greater a locus of disparity as judicial discretion.

References

American Bar Association. 2005. ABA Criminal Justice Section, report and recommendation on *Booker*. *Federal Sentencing Reporter*, 17: 335–340.

Albonetti, Celesta A. 1997. Sentencing under the federal sentencing Guidelines: Effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991–1992. *Law & Society Review*, 31: 789–822.

Albonetti, Celesta A. 1998. The role of gender and departures in the sentencing of defendants convicted of a white collar offense under the federal sentencing Guidelines. In (Jeffery T. Ulmer, ed.), *Sociology of Crime, Law, and Deviance*, Vol. 1. Greenwich, CT: JAI Press.

Anderson, Amy and Cassia Spohn. 2010. Lawlessness in the federal sentencing process: A test for uniformity and consistency in sentencing practices. *Justice Quarterly*, 27: 362–393.

Berman, Douglas A. 2005. Assessing federal sentencing after *Booker*. *Federal Sentencing Reporter*, 17: 291–294.

Bowman, Frank O. III. 2003. *Only Suckers Pay the Sticker Price: The Effect of "Fast Track" Programs on the Future of the Sentencing Guidelines as a Principled Sentencing System*. Testimony before the U.S. Sentencing Commission, Sept. 23, 2003, Washington, DC.

Bowman, Frank O. III. 2005. Letter from Frank Bowman concerning H.R. 1528. *Federal Sentencing Reporter*, 17: 311–314.

Bushway, Shawn D., Brian D. Johnson, and Lee Ann Slocum. 2007. Is the magic still there? The relevance of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*, 23: 151–178.

Bushway, Shawn D. and Anne Morrison Piehl. 2001. Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law & Society Review*, 35: 733–764.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100: 1261–1293.

Doerner, Jill K. and Stephen Demuth. 2009. The independent and joint effects of race/ethnicity, gender, and age on sentencing outcomes in U.S. federal courts. *Justice Quarterly*, 27: 1–27.

Engen, Rodney. 2009. Assessing determinate and presumptive sentencing: Making research relevant. *Criminology and Public Policy*, 8: 323–335.

Engen, Rodney and Randy R. Gainey. 2000. Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: the rules have changed. *Criminology*, 38: 1207–1230.

Feldmeyer, Ben and Jeffery T. Ulmer. 2011. Racial/ethnic threat and federal sentencing. *Journal of Research in Crime and Delinquency*, 48: 238–270.

Frase, Richard. 2007. The Apprendi-Blakely cases: Sentencing reform counter-revolution? *Criminology & Public Policy*, 6: 403–432.

Gonzales, Alberto. 2005. Federal sentencing guidelines speech by Attorney General Alberto Gonzales. *Federal Sentencing Reporter*, 17: 324–326.

Hartley, Richard, Sean Maddan, and Cassia Spohn. 2007. Prosecutorial discretion: An examination of substantial assistance departures in federal crack-cocaine and powder cocaine cases. *Justice Quarterly*, 24: 382–407.

Hofer, Paul J. 2007. *United States v. Booker* as a natural experiment: Using empirical research to inform the federal sentencing policy debate. *Crime & Public Policy*, 6: 433–460.

Holder, Eric. 2009. Remarks for the Charles Hamilton Houston Institute for Race and Justice and Congressional Black Caucus Symposium "Rethinking Federal Sentencing Policy 25th Anniversary of the Sentencing Reform Act." June 24, 2009. Washington, DC.

Johnson, Brian D. 2006. The multilevel context of criminal sentencing: Integrating judge and county level influences in the study of courtroom decision making. *Criminology*, 44: 259–298.

Johnson, Brian D. and Sara Betsinger. 2009. Punishing the 'model minority': Asian-American criminal sentencing outcomes in federal district courts. *Criminology*, 47: 1045–1090.

Johnson, Brian D., Jeffery T. Ulmer, and John H. Kramer. 2008. The social context of Guideline circumvention: the case of federal district courts. *Criminology*, 46: 711–783.

Kautt, Paula M. 2002. Location, location, location: Interdistrict and intercircuit variation in sentencing outcomes for federal drug-trafficking offenses. *Justice Quarterly*, 19: 633–671.

Kempf-Leonard, Kimberly and Lisa Sample. 2001. Have federal sentencing guidelines reduced severity? An examination of one circuit. *Journal of Quantitative Criminology*, 17: 111–144.

Kramer, John H. 2009. Mandatory sentencing guidelines: The framing of justice. *Criminology & Public Policy*, 8: 313–321.

Kramer, John H. and Jeffery T. Ulmer. 2009. *Sentencing Guidelines: Lessons from Pennsylvania*. Boulder, CO: Lynne Rienner.

Kurlychek, Megan and Brian D. Johnson. 2004. The juvenile penalty: A comparison of juvenile and young adult sentencing outcomes in adult criminal court. *Criminology*, 42: 485–517.

Maxfield, Linda D. and Keri Burchfield. 2002. Immigration offenses involving unlawful entry: Is federal practice comparable across districts? *Federal Sentencing Reporter*, 14: 260–266.

Maxfield, Linda D. and John H. Kramer. 1998. Substantial assistance: An empirical yardstick for gauging equity in current federal policy and practice. *Federal Sentencing Reporter*, 11: 6–12.

Mitchell, Ojmarrh and Doris L. MacKenzie. 2004. *The Relationship Between Race, Ethnicity, and Sentencing Outcomes: A Meta-Analysis of Sentencing Research: Final Report*. Submitted to the National Institute of Justice, Washington, DC.

Mustard, David. 2001. Racial, ethnic, and gender disparities in sentencing: Evidence from the U.S. federal courts. *Journal of Law and Economics*, 44: 285–314.

Paternoster, Raymond, Robert Brame, Paul Mazzerolle, and Alex R. Piquero. 1998. Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36: 859–866.

Peterson, Ruth D. and John Hagan. 1984. Changing conceptions of race: Towards an account of anomalous findings in sentencing research. *American Sociological Review*, 49: 56–70.

Reitz, Kevin R. 1998. Modeling discretion in American sentencing systems. *Law and Policy*, 20: 389–428.

Reitz, Kevin R. 2005. The enforceability of sentencing guidelines. *Stanford Law Review*, 58: 155–174.

Richter, John C. 2008. Deja vu all over again: How post-*Booker* sentencing threatens equal justice under the law. *Federal Sentencing Reporter*, 20: 340–342.

Shermer, Lauren O'Neill and Brian D. Johnson. 2010. Criminal prosecutions: Examining prosecutorial discretion and charge reductions in U.S. federal district courts. *Justice Quarterly*, 27: 394–430.

Spohn, Cassia. 2000. Thirty years of sentencing reform: The quest for a racially neutral sentencing process. In *Policies, Processes and Decisions of the Criminal Justice System*. Vol. 3, *Criminal Justice 2000*. Washington, DC: U.S. Department of Justice.

Spohn, Cassia and David Holleran. 2000. The imprisonment penalty paid by young unemployed Black and Hispanic male offenders. *Criminology*, 38: 281–306.

Steffensmeier, Darrell and Stephen Demuth. 2000. Ethnicity and sentencing outcomes in U.S. federal courts: Who is punished more harshly? *American Sociological Review*, 65: 705–729.

Steffensmeier, Darrell, John H. Kramer, and Cathy Streifel. 1993. Gender and imprisonment decisions. *Criminology*, 31: 411–446.

Steffensmeier, Darrell, Jeffery T. Ulmer, and John H. Kramer. 1998. The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, Black, and male. *Criminology*, 36: 763–798.

Stith, Kate. 2008. The arc of the pendulum: Judges, prosecutors, and the exercise of discretion. *The Yale Law Journal*, 117: 1420–1497.

Stith, Kate and Jose Cabranes. 1998. *Fear of Judging*. Chicago, IL: University of Chicago Press.

Stolzenberg, Ross M. and Daniel A. Relles. 1997. Tools for intuition about sample selection bias and its correction. *American Sociological Review*, 62: 494–507.

Thompson, Sandra G. 2006. The *Booker* project: The future of federal sentencing. *Houston Law Review*, 43: 269–278.

Ulmer, Jeffery T. 1997. *Social Worlds of Sentencing: Court Communities Under Sentencing Guidelines.* Albany: State University of New York Press.

Ulmer, Jeffery T. 2005. The localized uses of federal sentencing guidelines in four U.S. district courts: Evidence of processual order. *Symbolic Interaction*, 28: 255–279.

Ulmer, Jeffery T., James Eisenstein, and Brian D. Johnson. 2010. Trial penalties in federal sentencing: Extra-guidelines factors and district variation. *Justice Quarterly*, 27: 560–592.

Ulmer, Jeffery T. and Brian D. Johnson. 2004. Sentencing in context: A multilevel analysis. *Criminology*, 42: 137–177.

U.S. Sentencing Commission. 2004. *Fifteen Years of Guidelines Sentencing: An Assessment of How Well the Federal Criminal Justice System Is Achieving the Goals of Sentencing Reform*. Washington, DC: Author.

U.S. Sentencing Commission. 2006. *Final Report on the Impact of United States v. Booker on Federal Sentencing*. Washington, DC: Author.

U.S. Sentencing Commission. 2010. *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis*. Washington, DC: Author.

Zatz, Marjorie. 2000. The convergence of race, ethnicity, gender, and class on court decision making: Looking toward the 21st century. In *Policies, Processes and Decisions of the Criminal Justice System*, Vol. 3, *Criminal Justice 2000*. Washington, DC: U.S. Department of Justice.

Statutes Cited

Consumer Privacy Protection Act of 2005, H.R. 1528.

PROTECT Act of 2003, Pub.L. 108–21, 117 Stat. 650, S. 151, enacted April 30, 2003; amendment Stat. 667.

Court Cases Cited

Blakely v. Washington, (02–1632) 542 U.S. 296 (2004) 111 Wash. App. 851, 47 P.3d 149, reversed and remanded.

Gall v. United States, 552 U.S. 38 (2007).

Kimbrough v. United States, 552 U.S. 85 (2007).

Koon v. United States, 518 U.S. 81 (1996).

Rita v. United States, 551 U.S. 338 (2007).

United States v. Booker, 543 U.S. 220 (2005).

Jeffery T. Ulmer is currently Associate Professor of Sociology and Crime, Law, and Justice. His articles in journals such as *American Sociological Review*, *Social Problems*, *Criminology*,

Criminology and Public Policy, *Justice Quarterly*, *The Sociological Quarterly* and *Journal of Research in Crime and Delinquency* on courts and sentencing, criminological theory, race, ethnicity, and violence, religion and crime, symbolic interactionism, criminal enterprise, and the integration of qualitative and quantitative methods. He is the author of *Social Worlds of Sentencing: Court Communities Under Sentencing Guidelines* (1997, State University of New York Press), and coauthor (with Darrell Steffensmeier) of *Confessions of a Dying Thief: Understanding Criminal Careers and Illegal Enterprise* (2005, Aldine-Transaction) which won the 2006 Hindelang Award from the American Society of Criminology. His newest book (with John Kramer), *Sentencing Guidelines: Lessons from Pennsylania* was published in 2009 by Lynne Rienner Publishers.

Michael T. Light is a PhD candidate in the Department of Sociology and Crime, Law and Justice at The Pennsylvania State University. His research interests include punishment, immigration, race, ethnicity and criminal offending, and spatial analysis of crime. He (with Casey Harris) is the recipient of American Sociological Associations Crime, Law, and Deviance best student paper award and his recent work has been published in *Justice Quarterly*, *The Journal of Gender, Race, & Justice*, and *Federal Sentencing Reporter*.

John H. Kramer is Professor of Sociology and Crime, Law, and Justice. He served as Executive Director of the Pennsylvania Commission on Sentencing from 1978 to 1998, Advisor to the U.S. Sentencing Commission from 1995 to 1996, and Staff Director of the U.S. Sentencing Commission from 1996 to 1998. His research and policy development focus have been in the areas of sentencing disparity, the impact of sentencing reform, and evaluating the effectiveness of intermediate punishment and drug courts. His research has been published in numerous journals including *Criminology, Justice Quarterly, Social Problems*, *Journal of Research in Crime and Delinquency, and Criminology and Public Policy.* He is coauthor (with Jeffery Ulmer) of *Sentencing Guidelines: Lessons from Pennsylvania* (2009, Lynne Rienner Publishers)

Appen0dix A.

U.S. Sentencing Commission "Refined" OLS Models of Sentence Length Based on 2010 *Booker* Report

Independent Variables	(1) pre-PROTECT Ln Length b	(1) SE	(2) post-PROTECT Ln Length b	(2) SE	(3) post-*Booker* Ln Length b	(3) SE	(4) post-*Gall* Ln Length b	(4) SE	(2) vs. (1) z[a]	(3) vs. (1) z	(3) vs. (2) z	(4) vs. (1) z	(4) vs. (2) z	(4) vs. (3) z
White male (reference)														
White female	−0.244	0.023***	−0.275	0.024***	−0.216	0.015***	−0.340	0.021***	−0.930	1.003	2.050*	−3.100***	−2.043*	−4.772***
Black male	0.130	0.014***	0.089	0.015***	0.164	0.010***	0.217	0.013***	−1.942†	1.989*	4.182***	4.507***	6.398***	3.287***
Black female	0.258	0.026***	−0.422	0.028***	−0.300	0.018***	−0.349	0.027***	−4.359***	−1.343	3.687***	−2.480	1.901†	−1.529
Hispanic male	−0.058	0.016***	−0.023	0.017	−0.019	0.011†	0.049	0.014***	1.527	2.041*	0.192	5.106***	3.346***	3.938***
Hispanic female	−0.391	0.027***	−0.286	0.028***	−0.280	0.017***	−0.177	0.024***	2.676**	3.443***	0.184	5.920***	2.944**	3.496
Other male	−0.008	0.030	−0.097	0.030***	0.020	0.018	−0.014	0.025	−2.095*	0.795	3.336***	−0.155	2.115*	−1.094***
Other female	−0.432	0.056***	−0.592	0.056***	−0.247	0.034***	−0.499	0.049***	−2.008*	2.809**	5.254***	−0.889	1.254	−4.218***
>25 years old	0.033	0.012**	0.027	0.012*	0.033	0.008***	0.025	0.011*	−0.351	0.011	0.417	−0.539	−0.161	−0.651
U.S. citizen	−0.086	0.015***	−0.021	0.016	−0.052	0.010***	−0.074	0.013***	3.052**	1.956*	−1.682†	0.596	−2.608**	−1.369
Attending college	−0.103	0.012***	−0.073	0.013***	−0.077	0.008***	−0.075	0.011***	1.722†	1.807†	−0.268	1.687†	−0.154	0.105
Sexual abuse	−0.100	0.065	0.057	0.065	−0.012	0.040	−0.026	0.053	1.705†	1.153	−0.910	0.875	−0.994	−0.216
Pornography	−0.015	0.054	0.050	0.053	0.245	0.028***	0.329	0.034***	0.862	4.301***	3.230***	5.412***	4.398***	1.908†
Drug trafficking	−0.128	0.024***	−0.063	0.026*	−0.047	0.016**	−0.179	0.022***	1.870†	2.829**	0.522	−1.573	−3.427***	−4.809***
White collar	−0.492	0.026***	−0.519	0.028***	−0.480	0.018***	−0.558	0.024***	−0.710	0.369	1.161	−1.866†	−1.058	−2.565*
Immigration	0.022	0.027	0.025	0.029	0.054	0.018**	−0.003	0.024	0.071	0.978	0.854	−0.683	−0.735	−1.870†
Other offense	−0.337	0.025***	−0.303	0.026***	−0.297	0.017***	−0.414	0.023***	0.935	1.345	0.204	−2.315*	−3.206***	−4.179***
Upward departure	0.928	0.054***	1.050	0.054***	0.979	0.025***	0.884	0.031***	1.583	0.851	−1.186	−0.698	−2.637**	−2.382*
Substantial assistance dep.	−1.217	0.014***	−1.182	0.015***	−1.119	0.010***	−1.092	0.013***	1.772†	5.864***	3.580***	6.531***	4.528***	1.661†
Other downward departure	−1.076	0.014***	−0.885	0.016***	−0.913	0.008***	−0.979	0.010***	9.058***	10.261***	−1.575	5.645***	−4.998***	−5.182***
Mandatory minimum	0.333	0.014***	0.295	0.015***	0.354	0.009***	0.524	0.013***	−1.815†	1.197	3.265***	9.990***	11.542***	10.846***
Trial	0.224	0.026***	0.240	0.024***	0.284	0.015***	0.375	0.022***	0.442	2.023*	1.590	4.436***	4.156***	3.401***
Presentence detention	0.715	0.012***	0.744	0.013***	0.759	0.008***	0.867	0.012***	1.637	3.018**	0.987	9.133***	7.102***	7.601***
(ln) presumptive sentence	0.715	0.002***	0.686	0.002***	0.698	0.002***	0.642	0.002***	−8.741***	−6.046***	4.119***	−22.885***	−13.571***	−20.555***
Constant	0.738	0.031***	0.738	0.034***	0.621	0.022***	0.722	0.029***	−0.002	−3.077**	−2.927**	−0.363	−0.348	2.780**
N	88,856		70,505		187,793		124,118							
Adjusted R^2	0.685		0.715		0.678		0.620							

[a]Based on a two-tailed test. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Appendix B.

Alternative OLS Models of Sentence Length for Those Who Received Incarceration Sentences

	(1) pre-PROTECT Ln Length		(2) post-PROTECT Ln Length		(3) post-*Booker* Ln Length		(4) post-*Gall* Ln Length		(2) vs. (1)	(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
Independent Variables	***b***	***SE***	***b***	***SE***	***b***	***SE***	***b***	***SE***	z^a	*z*	*z*	*z*	*z*	*z*
White male (reference)														
White female	−0.125	0.010***	−0.128	0.010***	−0.099	0.006***	−0.109	0.008***	−0.265	2.202*	2.412*	1.226	1.466	−0.975
Black male	0.066	0.005***	0.045	0.006***	0.053	0.003***	0.077	0.004***	−2.748**	−2.164*	1.172	1.547	4.424***	4.326***
Black female	−0.112	0.011***	−0.115	0.013***	−0.084	0.008***	−0.075	0.011***	−0.205	2.057*	2.105*	2.413*	2.449*	0.705
Hispanic male	−0.034	0.006***	−0.039	0.006***	−0.025	0.004***	0.011	0.005*	−0.610	1.292	1.941†	5.986***	6.388***	5.974***
Hispanic female	−0.162	0.011***	−0.156	0.011***	−0.132	0.007***	−0.085	0.009***	0.371	2.367*	1.833†	5.631***	5.041***	4.421***
Other male	−0.004	0.011	−0.036	0.011**	−0.005	0.007	−0.018	0.009*	−1.953†	−0.079	2.303*	−0.983	1.202	−1.190
Other female	−0.123	0.026***	−0.134	0.026***	−0.086	0.014***	−0.087	0.019***	−0.318	1.226	1.622	1.115	1.479	−0.013
Age (in years)	0.001	0.000***	0.001	0.000***	0.001	0.000	0.001	0.000***	0.335	2.743**	2.212*	2.189*	1.726†	−0.443
U.S. citizen	0.003	0.005	−0.007	0.006	−0.007	0.004*	0.026	0.005***	−1.273	−1.645	−0.060	3.229***	4.432***	5.810***
Less than high school	−0.008	0.009	−0.003	0.010	0.012	0.006*	−0.023	0.008***	0.322	1.851†	1.335	−1.297	−1.558	−3.668***
High school graduate	0.000	0.009	−0.002	0.010	0.018	0.006**	−0.007	0.008	−0.106	1.682†	1.676†	−0.555	−0.405	−2.576**
Some college	−0.015	0.009†	−0.005	0.010	0.008	0.006	−0.016	0.008*	0.731	2.105*	1.095	−0.015	−0.799	−2.388*
Drug	0.281	0.012***	0.319	0.014***	0.231	0.010***	0.215	0.014***	2.044*	−3.126***	−5.109***	−3.536***	−5.272***	−0.960
Immigration	0.120	0.013***	0.098	0.015***	0.050	0.010***	0.044	0.014***	−1.086	−4.220***	−2.697**	−3.937***	−2.655**	−0.346
Violent	0.284	0.015***	0.301	0.017***	0.183	0.011***	0.167	0.015***	0.762	−5.556***	−5.919***	−5.533***	−5.909***	−0.824
Fraud	0.005	0.013	0.010	0.015	0.013	0.010	0.023	0.014†	0.277	0.510	0.162	0.944	0.617	0.562

(Continued)

Appendix B.

(Continued)

	(1) pre-PROTECT Ln Length		(2) post-PROTECT Ln Length		(3) post-*Booker* Ln Length		(4) post-*Gall* Ln Length		(2) vs. (1)	(3) vs. (1)	(3) vs. (2)	(4) vs. (1)	(4) vs. (2)	(4) vs. (3)
Independent Variables	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	z^a	*z*	*z*	*z*	*z*	*z*
Firearms	0.128	0.013***	0.148	0.015***	0.106	0.010***	0.133	0.014***	1.007	−1.320	−2.353*	0.258	−0.735	1.560
Other offense	0.068	0.015***	0.100	0.016***	0.127	0.011***	0.156	0.015***	1.464	3.278***	1.392	4.272***	2.563*	1.608
Upward departure	0.721	0.018***	0.783	0.019***	0.781	0.008***	0.751	0.010***	2.390*	3.046**	−0.098	1.477	−1.526	−2.419*
Substantial assistance dep.	−0.513	0.005***	−0.486	0.006***	−0.533	0.003***	−0.515	0.005***	3.469***	−3.342***	−7.132***	−0.382	−3.993***	3.132***
Other downward departure	−0.537	0.005***	−0.474	0.006***	−0.459	0.003***	−0.463	0.003***	8.104***	13.615***	2.279*	12.116***	1.500	−1.079
Trial	0.268	0.009***	0.261	0.008***	0.232	0.005***	0.257	0.007***	−0.623	−3.556***	−2.899**	−0.988	−0.318	2.917**
Presentence detention	0.182	0.005***	0.190	0.005***	0.180	0.003***	0.206	0.004***	1.131	−0.310	−1.581	3.811***	2.396*	4.862***
Criminal history	0.051	0.001***	0.049	0.001***	0.038	0.001***	0.030	0.001***	−1.355	−9.201***	−7.172***	−13.522***	−11.505***	−6.471***
(ln) presumptive sentence	0.669	0.002***	0.653	0.002***	0.712	0.001***	0.689	0.001***	−6.399***	21.610***	28.562***	9.476***	16.298***	−13.217***
Constant	0.820	0.017***	0.872	0.020***	0.721	0.013***	0.802	0.017***	1.974*	−4.600***	−6.472***	−0.752	−2.694**	3.782***
N	73,401		58,734		157,685		102,643							
Adjusted R^2	0.813		0.830		0.836		0.838							

[a] Based on a two-tailed test. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Appendix C.

Logistic Regression Models of "Other" Downward Departures

Independent Variables	(1) pre-PROTECT Downward Dep b	SE	(2) post-PROTECT Downward Dep b	SE	(3) post-*Booker* Downward Dep b	SE	(4) post-*Gall* Downward Dep b	SE	(2) vs. (1) z^a	(3) vs. (1) z	(3) vs. (2) z	(4) vs. (1) z	(4) vs. (2) z	(4) vs. (3) z
White male (reference)														
White female	0.173	0.049***	0.189	0.064**	0.122	0.029***	0.077	0.034*	0.200	−0.881	−0.951	−1.607	−1.556	−1.017
Black male	−0.678	0.036***	−0.538	0.046***	−0.397	0.020***	−0.661	0.023***	2.401*	6.900***	2.793**	0.398	−2.387*	−8.715***
Black female	−0.167	0.060**	−0.172	0.082*	−0.185	0.037***	−0.266	0.044***	−0.040	−0.248	−0.149	−1.323	−1.015	−1.413
Hispanic male	0.130	0.032***	0.167	0.042***	−0.019	0.020	−0.432	0.022***	0.694	−3.925***	−4.002***	−14.340***	−12.627***	−13.817***
Hispanic female	0.551	0.049***	0.486	0.062***	0.129	0.031***	−0.290	0.038***	−0.810	−7.218***	−5.128***	−13.554***	−10.673***	−8.581***
Other male	−0.147	0.065*	−0.017	0.078	−0.073	0.035*	−0.111	0.041**	1.274	0.998	−0.650	0.470	−1.060	−0.700
Other female	0.168	0.116	0.145	0.143	0.177	0.063**	0.089	0.078	−0.125	0.064	0.202	−0.568	−0.347	−0.880
Age (in years)	0.005	0.001***	0.007	0.001***	0.003	0.001***	0.003	0.001***	1.233	−1.084	−2.337*	−1.586	−2.734**	−0.730
U.S. citizen	−0.174	0.030***	−0.229	0.038***	−0.289	0.018***	−0.245	0.022***	−1.148	−3.298***	−1.423	−1.927†	−0.372	1.503
Less than high school	−0.119	0.047*	−0.268	0.060***	−0.228	0.028***	−0.240	0.033***	−1.949†	−1.990*	0.599	−2.099*	0.405	−0.274
High school graduate	−0.283	0.048***	−0.389	0.061***	−0.276	0.028***	−0.262	0.033***	−1.378	0.120	1.691†	0.363	1.843†	0.333
Some college	−0.162	0.050***	−0.305	0.063***	−0.158	0.029***	−0.120	0.034***	−1.777†	0.062	2.105*	0.689	2.563*	0.848
Drug	1.090	0.072***	0.740	0.092***	0.641	0.044***	0.754	0.054***	−2.995**	−5.324***	−0.959	−3.753***	0.131	1.609
Immigration	1.548	0.074***	1.482	0.095***	1.228	0.046***	1.209	0.056***	−0.549	−3.663***	−2.414*	−3.648***	−2.483*	−0.260
Violent	1.048	0.085***	0.697	0.114***	0.550	0.054***	0.453	0.065***	−2.464*	−4.929***	−1.171	−5.539***	−1.862†	−1.142
Fraud	0.509	0.073***	0.398	0.093***	0.400	0.045***	0.524	0.054***	−0.944	−1.287	0.020	0.157	1.167	1.771†

(Continued)

Appendix C.

(Continued)

Independent Variables	(1) pre-PROTECT Downward Dep *b*	*SE*	(2) post-PROTECT Downward Dep *b*	*SE*	(3) post-*Booker* Downward Dep *b*	*SE*	(4) post-*Gall* Downward Dep *b*	*SE*	(2) vs. (1) *z*[a]	(3) vs. (1) *z*	(3) vs. (2) *z*	(4) vs. (1) *z*	(4) vs. (2) *z*	(4) vs. (3) *z*
Firearms	0.872	0.078***	0.475	0.099***	0.407	0.047***	0.503	0.057***	−3.152***	−5.117***	−0.617	−3.837***	0.242	1.297
Other offense	0.748	0.081***	0.353	0.106***	0.476	0.048***	0.628	0.057***	−2.960**	−2.893**	1.052	−1.216	2.278*	2.031*
Trial	−0.602	0.063***	−0.577	0.072***	−0.301	0.028***	−0.344	0.036***	0.254	4.333***	3.565***	3.540***	2.897**	−0.934
Presentence detention	−0.248	0.026***	−0.312	0.035***	−0.549	0.017***	−0.583	0.019***	−1.447	−9.633***	−6.123***	−10.185***	−6.762***	−1.311
Criminal history	0.068	0.007***	0.057	0.009***	0.019	0.004***	0.028	0.005***	−1.007	−6.047***	−3.963***	−4.776***	−2.960**	1.411
Presumptive sentence	0.000	0.000	0.001	0.000***	0.003	0.000***	0.003	0.000***	3.289***	12.440***	6.511***	14.352***	8.332***	3.491***
Constant	−2.290	0.096***	−2.374	0.123***	−1.091	0.059***	−0.691	0.071***	−0.534	10.649***	9.401***	13.419***	11.860***	4.370***
N	72,807		58,662		156,406		104,390							
−2 log likelihood	68,128.8		46,263.2		180,084.9		129,834.2							

[a]Based on a two-tailed test. †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

POLICY ESSAY

RACIAL DISPARITY IN WAKE OF THE *BOOKER/FANFAN* DECISION

Unwarranted disparity in the wake of the *Booker/Fanfan* decision

Implications for research and policy

Cassia Spohn
Arizona State University

In 1972, Marvin Frankel, U.S. District Judge for the Southern District of New York, issued an influential call for reform of the federal sentencing process (Frankel, 1972a, 1972b). Judge Frankel characterized the indeterminate sentencing system that existed at that time as "a bizarre 'nonsystem' of extravagant powers confided to variable and essentially unregulated judges, keepers, and parole officials" (Frankel, 1972b: 1; see also Gertner, 2007). Frankel was particularly concerned about the degree of discretion given to judges, which he maintained led to "lawlessness" in sentencing. Claiming that judges "were not trained at all" for "the solemn work of sentencing" (Frankel, 1972b: 6), Judge Frankel called for legislative reforms designed to regulate "the unchecked powers of the untutored judge" (Frankel, 1972b: 41). More to the point, he called for the creation of an administrative agency called a sentencing commission that would create rules for sentencing that judges would be required to follow.

Congress responded to Judge Frankel's call for reform of federal sentencing by enacting the Sentencing Reform Act of 1984 (SRA). The SRA created the United States Sentencing Commission (USSC), which was authorized to develop and implement presumptive sentencing guidelines designed to achieve "honesty," "uniformity," and "proportionality" in sentencing (USSC, 2001). The Act also abolished discretionary release on parole, stated that departures from the guidelines would be permitted only with written justification, and provided for appellate review of sentences to determine whether the guidelines were

Direct correspondence to Cassia Spohn, School of Criminology and Criminal Justice, Arizona State University, 411 N. Central Avenue, Phoenix, AZ 85004 (e-mail: Cassia.Spohn@asu.edu).

DOI:10.1111/j.1745-9133.2011.00768.x

applied correctly or whether a departure was reasonable. The federal sentencing guidelines promulgated by the USSC went into effect in 1987.

The goals of those who championed the enactment of the federal sentencing guidelines varied. Liberals argued that structured sentencing practices would enhance fairness and equity in sentencing and would hold judges accountable for their decisions, whereas conservatives asserted that the reforms would lead to harsher penalties that eventually would deter criminal behavior. Reformers on both sides of the political spectrum, however, agreed that the changes were designed to curb discretion and reduce unwarranted disparity. Both conservatives and liberals urged sentencing reform as a means of reducing "lawlessness" (Frankel, 1972b) in sentencing. Reflecting this mindset, the *Federal Sentencing Guidelines Manual* states that one of the three objectives Congress sought to achieve in enacting the SRA was "reasonable uniformity in sentencing by narrowing the wide disparity in sentences imposed for similar criminal offenses committed by similar offenders" (USSC, 2001: Ch.1, Pt. A. 3). It is also clear that the guidelines were designed to eliminate discrimination based on legally irrelevant characteristics of the offender; in determining the appropriate sentence, judges are prohibited explicitly from considering the offender's race, sex, national origin, creed, religion, and socioeconomic status (USSC, 2001, Ch. 5, Pt. H1.10). As the Department of Justice stated in 1987, "unwarranted disparity caused by broad judicial discretion is the ill that the Sentencing Reform Act seeks to cure" (Trott, 1995: 197).

Although evidence on whether the federal sentencing guidelines were able to achieve their goal of eliminating unwarranted disparity in sentencing is mixed (for reviews, see Mitchell, 2005; Spohn, 2000; USSC, 2004), both social scientists (Miethe and Moore, 1985; Wooldredge, 2009) and legal scholars (Hofer, 2006) argued that the potential for racial and ethnic discrimination was limited by the fact that the guidelines were mandatory and that judge-initiated departures were regulated closely. These aspects of the guidelines, in other words, constrained judicial discretion and made it less likely that judges would resort to stereotypes linked to race/ethnicity and other legally irrelevant factors in attempting to fashion sentences that fit offenders and their crimes (for an alternative view, see Stith and Cabranes, 1998: 126–128).

A series of decisions by the U.S. Supreme Court that reshaped the state and federal sentencing process made the federal sentencing guidelines "effectively advisory" rather than mandatory (see *U.S. v. Booker,* 2005) and held that federal judges must make an individualized determination of the appropriate sentence based on the facts presented (see *Gall v. United States*, 2007) led several commentators (Frase, 2007; Hofer, 2007; Klein, 2005) to suggest that the result might be an increase in unwarranted disparity in federal sentencing. For example, Klein (2005: 720–721) noted that the *Booker* decision provided judges with the authority to fashion sentences that reflect the purposes of sentencing set forth in the SRA and argued that:

> [T]his authority, coupled with the admonition in 18 U.S.C. § 3661 that 'no limitations be placed on the information concerning the background, character, and conduct of a person convicted of an offense which a court of the United States may receive and consider for purposes of imposing an appropriate sentence,' allow trial judges *free reign* in gathering information and making discretionary sentencing decisions" (emphasis added).

Other scholars disagree, arguing that the Court's decision in *Booker* did not grant judges unfettered discretion at sentencing and that in the post-*Booker* era, the federal sentencing guidelines "remain as restrictive of judicial sentencing discretion as any system in the United States" (Reitz, 2005: 171; see also Berman, 2006; Bowman, 2005).

Advisory Guidelines and Unwarranted Disparity

Determining whether the Court's decisions have led to increased disparity in the federal sentencing process—and, more specifically, whether the decisions have increased disparity based on legally irrelevant offender characteristics such as race, ethnicity, and sex—clearly is important. In fact, Hofer (2007: 451) contended that this is "perhaps the most important empirical question facing policy makers." If, as critics of *Booker* and subsequent decisions assert, unwarranted disparity has increased as a result of the loosening of constraints on judicial discretion to determine the appropriate sentence and to depart from the guideline range, then some type of policy "fix" designed to refashion the guidelines as mandatory may be necessary. It also is a significant issue for sentencing scholars, whose work will inform the debate and influence subsequent policy changes. As Hofer (2007: 456) noted, our understanding of the impact of changes in sentencing policy must be "grounded in hard evidence and not in mere anecdotes or speculation."

Given the centrality of this issue, it is interesting that two major studies have reached different conclusions. Research by the USSC (2010) found that the offender's race, sex, citizenship status, education, and age affected the length of the prison sentence during at least some of the time periods under study. More to the point, that study's refined analysis found that disparity based on the *combination* of the offender's race and sex increased both post-*Booker* and post-*Gall*. To illustrate, compared with White male offenders, Black male offenders received 5.5% longer sentences in the pre-*Booker* period, but they received 15.2% longer sentences in the post-*Booker* period and 23.3% longer sentences in the post-*Gall* period. Hispanic male offenders received sentences that were 4.4% shorter than those imposed on White males in the pre-*Booke*r period, but they received 6.8% longer sentences than those imposed on White males in the post-*Gall* era. In contrast, the study by Ulmer, Light, and Kramer (2011, this issue), which (a) analyzed data over a longer period of time, (b) examined the likelihood of incarceration and the likelihood of departure as well as sentence length, (c) included somewhat different control variables, and (d) used a different modeling strategy, found no increase in racial disparity in sentence length post-*Booker* and post-*Gall*. In fact, *less* disparity was found in sentence lengths for Black males (compared with White males) in the post-*Booker* and post-*Gall* period than found in 1994 and 1995, when the guidelines were mandatory and the standard for appellate review of sentences outside the guideline range was more stringent. Ulmer et al. also found, however, that though disparity in the odds of incarceration for Black male offenders relative to White male offenders was stable through the post-*Booker* period, it increased significantly post-*Gall.* Other important findings were that the increase in sentence length disparity for Black male offenders found by the USSC study was reduced substantially when a variable measuring the offender's criminal history points was included in the model, that a nontrivial portion of the sentence length disparity for Black males across all time periods was attributable to immigration offenses, and that greater disparity occurred in government-sponsored departures than in judge-initiated departures. Ulmer et al. concluded that their results suggest that Black male offenders' odds of imprisonment have increased post-*Gall* but call into question the "notion that *Booker* and *Gall* have caused increases in race/ethnic and gender sentence length disparity compared with the full range of years when the Guidelines were mandatory."

The contrasting findings of these two studies suggest that definitive answers to questions regarding the impact of the Supreme Court's *Booker* and *Gall* decisions remain elusive. They also suggest, consistent with previous work on the appropriate way to control for offense seriousness and criminal history in jurisdictions with presumptive guidelines (Engen and Gainey, 2000) and the use of a "total incarceration variable" combining prison and jail sentences (Holleran and Spohn, 2004), that the methodological choices made by researchers have important consequences for conclusions regarding the locus and extent of unwarranted disparity in sentencing. This finding, in turn, implies that if their work is to be relevant to policy and practice, then researchers whose focus is the federal sentencing process must agree on a general framework for analysis. It is not helpful to policy makers to be presented with the results of studies whose conflicting findings can be attributed simply to the fact that the authors used different methodologies and analytical strategies.

The results of the study conducted by Ulmer et al. (2011) have other implications for both research and policy. In terms of research, their study illustrates the importance of estimating models of the likelihood of incarceration and of the likelihood of a downward departure, as well as of the length of the prison sentence, as doing so can identify the locus of unwarranted disparity and point to appropriate policy remedies. If, as Ulmer et al. reveal, judges in the post-*Booker* era take the offender's race/ethnicity and sex into account primarily when deciding whether the offender should be sentenced to prison, it would make little sense to attempt to structure discretion by decreasing the range of sentences for each cell in the guideline matrix. Similarly, if unwarranted disparity is tied to government-initiated downward departures, especially those for providing substantial assistance, then an appropriate policy fix would be one targeting discretionary charging and plea bargaining decisions by prosecutors. The sentence imposed on an offender who has been found guilty of a crime is the result of a collaborative exercise that involves criminal justice officials other than the judge. It results from a sentencing *process,* not from a single sentencing *decision*, and researchers must attempt to identify where in this process unwarranted disparity is found.

The study by Ulmer et al. (2011) also highlights the value of estimating models that include interaction effects and that test for indirect, as well as direct, effects. As numerous commentators have pointed out, simply attempting to determine whether race and ethnicity have direct effects on sentence outcomes is a theoretically unsophisticated and incomplete approach to a complex phenomenon. The more interesting question is, "When does the particular social characteristic matter—under what circumstances, for whom, and in interaction with what other factors?" (Wonders, 1996: 617). For example, a growing body of research suggests that findings of leniency for female offenders may be conditioned by race/ethnicity and that findings of preferential treatment for White offenders (or more punitive treatment for minorities) may be conditioned by sex (for a review of this research, see Brennan, 2009). Thus, a failure to consider the intersection of sex and race/ethnicity may result in inaccurate conclusions about the effects of these variables on sentencing outcomes. Both the USSC and Ulmer et al. test for the joint effects of the offender's race/ethnicity and sex, finding that Black and Hispanic male offenders are singled out for harsher treatment at various decision points. The same pattern—that is, harsher treatment of racial minorities—was not found for female offenders; in fact, Black female offenders faced lower odds of incarceration than White female offenders, and the lengths of the prison sentences imposed on Black, Hispanic, and White females were similar.

These results—coupled with Ulmer et al.'s (2011) findings that the offender's criminal history had a significant effect on sentence length, above and beyond its effect through the presumptive sentence, and that the effect was particularly pronounced (in all time periods) for Black male offenders—suggest that judges believe, first, that offenders with lengthier and more serious criminal histories deserve more punitive treatment; second, that the incorporation of the offender's criminal history score in the presumptive sentence does not account for this adequately; and, third, that Black male repeat offenders are particularly deserving of more punitive sentences. So, what are the policy implications of these findings? If the guidelines do not capture judges' assessments of the punitive impact of the offender's criminal history adequately, then perhaps the guideline matrix should be revised so that increases in the criminal history score result in larger increases in the range of presumptive sentences for each cell. The problem with this solution is that it would result in more punitive presumptive sentences for all offenders with more serious criminal histories, and we have no guarantee that judges would not continue to impose more severe sentences than the guidelines called for on repeat Black male offenders. If judges' assessments of offenders' dangerousness, threat, and likelihood of recidivism are tied to their evaluations of offenders' criminal histories, and if this connection is more salient for Black male offenders than for other offenders, then tinkering with the guideline matrix in this way will not reduce the racial disparity in sentence outcomes.

Ulmer et al.'s (2011) study also illustrates the importance of disaggregating the data by offense type. They point out correctly that immigration cases, which now represent more than 25% of all cases adjudicated in federal courts, present unique challenges, are concentrated in a handful of districts, and are handled in distinct ways from most other offenses. Their results revealed that inclusion of these offenses resulted in larger estimates of racial/ethnic and gender disparity in sentence length than would be the case if these offenses were excluded, particularly for Black and Hispanic male offenders. This finding is important in that it provides policy makers with information regarding the circumstances under which, and the contexts in which, legally irrelevant offender characteristics come into play. It also suggests that researchers interested in pinpointing the location of unwarranted disparity in the federal sentencing process should analyze immigration cases, drug trafficking cases, fraud cases, and firearms cases separately. As Hawkins (1986–1987: 724) pointed out more than two decades ago, it is overly simplistic to assert that Blacks and Hispanics will receive "more severe punishment than whites for all crimes, under all conditions, and at similar levels of disproportion over time."

The Future of Federal Sentencing

Shortly after the Supreme Court ruled in the *Booker* case that the federal sentencing guidelines would be advisory rather than mandatory, calls came for Congress to act to "mandatorize" the guidelines (see House Bill H.R. 1528, which would have made the federal sentencing guidelines a series of mandatory minimum penalties). Other commentators, including the American Bar Association (2005), argued that any type of "*Booker*-fix" was premature and urged Congress to allow the advisory guidelines to operate for at least 1 year before enacting legislation designed to constrain judicial discretion at sentencing.

The advisory guideline system put into place by *Booker* (and reiterated by *Gall*) has now been in operation for more than 6 years. The result has not been the instability and lack of uniformity that critics of *Booker* predicted. In fact, a report by the USSC (2006) on the first post-*Booker* year of sentencing revealed that judges continued to consult and apply the guidelines, with the result that a within-guideline sentence was imposed in nearly two out of every three cases. Moreover, the severity of sentences imposed did not change substantially. According to Berman (2006: 157), the reality is that *Booker* "has not radically altered many central features of the federal sentencing system."

In 2006, a series of articles designed to "take stock concerning the state and direction of federal sentencing" (Berman, 2006: 157) appeared in the *Federal Sentencing Reporter*. The authors of one of these articles, both of whom were district court judges, commented on the remarkable stability in sentencing in the post-*Booker* period and noted that it was "impossible to determine" whether the variations from the guidelines that did exist represented unwarranted disparity (Adelman and Deitrich, 2006: 160). The authors concluded that "at the present time, the argument that mandates are needed to eliminate unwarranted disparity is utterly unsupported by evidence" (2006: 160).

The question, of course, is whether this conclusion is still valid 5 years later. As the study conducted by Ulmer et al. (2011) makes clear, the answer to this question depends to some extent on the methodological and analytical choices made by researchers attempting to tease out the effects of offender race, ethnicity, and sex on sentence outcomes. One study (USSC, 2010) concluded that the racial/ethnic disparity in sentence length increased substantially in the post-*Booker* (for Black males) and the post-*Gall* periods (for both Black males and Hispanic males), but another study by researchers who made different methodological choices and whose findings challenge the conclusion of increasing unwarranted disparity in the federal sentencing process (Ulmer et al.) provides a different conclusion. However, it is important to point out that the authors of this latter study *do not* conclude that no unwarranted disparity exists in the federal sentencing process or that disparities have not surfaced for some types of offenders (i.e., Black male offenders, especially those with more serious criminal histories), some types of decisions (i.e., the decision to incarcerate or not), and some types of offenses (i.e., immigration offenses). Although it may be true that *Booker* and *Gall* have not "caused increases in race/ethnic and gender sentence length disparity compared with the full range of years when the Guidelines were mandatory" (Ulmer et al.), it does not necessarily follow from this that the now-advisory guidelines are being implemented without any consideration of legally irrelevant offender characteristics. As Ulmer et al. point out, the situations in which racial disparity surfaced in their study "warrant additional scrutiny and perhaps discussions of policy changes targeted specifically to those two circumstances."

References

American Bar Association, Criminal Justice Section. 2005. Report and recommendation on *Booker*. Reprinted in *Federal Sentencing Reporter*, 17: 334–337.

Adelman, Lynn and Jon Deitrich. 2006. Disparity: Not a reason to "fix" *Booker*. *Federal Sentencing Reporter*, 18: 160–163.

Berman, Douglas. 2005. Forward: Beyond *Blakely* and *Booker*: Pondering modern sentencing process. *Journal of Criminal Law and Criminology*, 95: 653–688.

Berman, Douglas. 2006. Now what? The post-*Booker* challenge for Congress and the courts. *Federal Sentencing Reporter*, 18: 157–159.

Bowman, Frank O. 2005. Beyond band aids: A proposal for reconfiguring federal sentencing after *Booker*. *University of Chicago Legal Forum*, 2005: 149–216.

Brennan, Pauline K. 2009. The joint effects of offender race/ethnicity and sex on sentencing outcomes. In (Marvin D. Krohn, Alan J. Lizotte, and Gina Penly Hall, eds.), *Handbook of Crime and Deviance*. New York: Springer Science and Business Media.

Engen, Rodney L. and Randy R. Gainey. 2000. Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: The rules have changed. *Criminology*, 38: 1207–1229.

Frankel, Marvin E. 1972a. *Criminal sentences: Law without order*. New York: Hall and Wang.

Frankel, Marvin E. 1972b. Lawlessness in sentencing. *University of Cincinnati Law Review*, 41: 1–54.

Frase, Richard S. 2007. The *Apprendi-Blakely* cases: Sentencing reform counter-revolution? *Criminology & Public Policy*, 6: 403–432.

Gertner, Nancy. 2007. From omnipotence to impotence: American judges and sentencing. *Ohio State Journal of Criminal Law*, 4: 523–529.

Hawkins, Darnell F.. 1986–1987. Beyond anomalies: Rethinking the conflict perspective on race and criminal punishment. *Social Forces*, 65: 719–745.

Hofer, Paul J. 2006. Immediate and long-term effects of *United States v. Booker*: More discretion, more disparity, or better reasoned sentences? *Arizona State Law Journal*, 38: 425–468.

Holleran, David and Cassia Spohn. 2004. On the use of the total incarceration variable in sentencing research. *Criminology* 42: 211–240.

Hofer, Paul J. 2007. *United States v. Booker* as a natural experiment: Using the empirical research to inform the federal sentencing policy debate. *Criminology & Public Policy*, 6: 433–460.

Klein, Susan R. 2005. The return of federal judicial discretion in criminal sentencing. *Valparaiso University Law Review*, 39: 693–740.

Miethe, Terance D. and Charles Moore. 1985. Racial differences in criminal processing: The consequences of model selection on conclusions about differential treatment. *Journal of Criminal Law and Criminology*, 78: 155–176.

Mitchell, Ojmarrh. 2005. A meta-analysis of race and sentencing research: Explaining the inconsistencies. *Journal of Quantitative Criminology*, 21: 439–466.

Reitz, Kevin R. 2005. The enforceability of sentencing guidelines. *Stanford Law Review* 58: 155–174.

Spohn, Cassia. 2000. Thirty years of sentencing reform: The quest for a racially neutral sentencing process. In *Policies, Processes and Decisions of the Criminal Justice System*, Vol. 3, *Criminal Justice 2000*. Washington, DC: U.S. Department of Justice.

Stith, Kate and Jose A. Cabranes. 1998. *Fear of Judging: Sentencing Guidelines in the Federal Courts*. Chicago: University of Chicago Press.

Trott, Stephen S. 1995. Letter to the Honorable William W. Wilkins, Jr. Reprinted in the *Federal Sentencing Reporter*, 8: 196–198.

Ulmer, Jeffery T, Michael T. Light, and John H. Kramer. 2011. Racial disparity in the wake of the *Booker/Fanfan* decision: An alternative analysis to the USSC's 2010 report. *Criminology & Public Policy*. This issue.

United States Sentencing Commission. 2001. *Federal Sentencing Guidelines Manual*. St. Paul, MN: West.

United States Sentencing Commission. 2004. *Fifteen Years of Guidelines Sentencing: An Assessment of How Well the Federal Criminal Justice System Is Achieving the Goals of Sentencing Reform*. Washington, DC: Author.

United States Sentencing Commission. 2006. *Report on the Impact of United States v. Booker on Federal Sentencing*. Washington, DC: Author.

United States Sentencing Commission. 2010. *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis*. Washington, DC: Author.

Wonders, Nancy A. 1996. Determinate sentencing: A feminist and postmodern story. *Justice Quarterly*, 13: 611–648.

Wooldredge, John. 2009. Short- versus long-term effects of Ohio's switch to more structured sentencing on extralegal disparities in prison sentences in an urban court. *Criminology & Public Policy*, 8: 285–312.

Court Cases Cited

U.S. v. Booker, 543 U.S. 220 (2005).

Gall v. United States, 128 S. Ct. 586 (2007).

Statutes Cited

Consumer Privacy Protection Act of 2005, H.R. 1528.

Sentencing Reform Act, P.L. No. 98–473, 98 Stat. 1987 (1984).

Cassia Spohn is a Professor in the School of Criminology and Criminal Justice at Arizona State University, where she also serves as the director of the doctoral program. She is the coauthor of five books, including *The Color of Justice: Race, Ethnicity, and Crime in America* and *How Do Judges Decide? The Search for Fairness and Justice in Punishment.* Dr. Spohn has published extensively on prosecutors' charging decisions in sexual assault cases, the effect of race/ethnicity and gender on sentencing decisions, sentencing of drug offenders, and the deterrent effect of imprisonment. She is currently working on an NIJ-funded study of police decision making in sexual assault cases in Los Angeles.

POLICY ESSAY

RACIAL DISPARITY IN WAKE OF THE *BOOKER/FANFAN* DECISION

Race disparity under advisory guidelines

Dueling assessments and potential responses

Ryan W. Scott
Indiana University

Dueling studies of race disparity, one by the U.S. Sentencing Commission (USSC, 2010) and an alternative analysis published in this issue by Ulmer, Light, and Kramer (2011), diverge sharply in their methodological choices and in their characterization of trends in federal sentencing. The Commission's study suggests a marked increase in race disparity, differences in sentencing outcomes between racial groups that cannot be explained by controlling for relevant nonrace factors, after the Supreme Court's decisions in *United States v. Booker* (2005) and *Gall v. United States* (2007). Those decisions rendered the federal Sentencing Guidelines advisory and set a highly deferential standard of appellate review. The alternative analysis finds more modest changes, which are largely confined to immigration offenses and to the decision whether to impose a sentence of prison or probation.

Yet, in several of their key findings, the Commission's research and the new analysis by Ulmer et al. (2011) reach similar conclusions. Both agree that for Black male offenders compared with White male offenders, the "in/out" decision—whether to impose a sentence of imprisonment or probation—is a source of persistent and increasing disparity. Both suggest that evidence of race disparity under the mandatory Guidelines, before 2003, was unstable and inconclusive. And surprisingly, both also indicate that race disparity affecting Black male offenders reached its lowest levels ever under the PROTECT Act in 2003 and 2004, when the Guidelines were at their most mandatory and inflexible and departures were closely policed through *de novo* appellate review.

Although narrow, those areas of agreement have potentially important implications for sentencing law. This policy essay evaluates the support that the new research lends

Direct correspondence to Ryan W. Scott, Maurer School of Law, Indiana University, Law Building 247, 211 South Indiana Avenue, Bloomington, IN 47405 (e-mail: ryanscot@indiana.edu).

DOI:10.1111/j.1745-9133.2011.00779.x

to several paths forward for federal sentencing. It focuses on three possibilities: a system of "dispositional departures" to regulate the prison/probation decision; a rollback of the *Booker* remedial opinion that would restore the PROTECT Act regime, augmented by jury fact finding; and a new proposal to simplify the Guidelines championed by Judge William Sessions, the former Chair of the Sentencing Commission (the "Sessions proposal"). It concludes that the best approach, based on the current body of research, may be "none of the above." As a postscript, however, it urges that the new studies of race disparity be evaluated in the context of related research on interjudge sentencing disparity.

Dueling Studies, Common Ground

The Commission's 2010 report and the alternative analysis by Ulmer et al. (2011) diverge sharply in their characterization of recent trends in race disparity at the federal level. Reversing its previous conclusions about post-*Booker* race disparity (USSC, 2006), the Commission found that after controlling for legally relevant factors for which data are available, Black male offenders consistently have received longer sentences than White male offenders, and the degree of disparity has "increased steadily since *Booker*" (USSC, 2010).[1] That kind of trend, if borne out by the data, is deeply worrisome in light of already staggering incarceration rates among Black men in America. Disparate levels of imprisonment for Black male offenders have resulted principally from facially race-neutral sentencing rules, such as mandatory minimum sentences for drug offenses and the now-repealed 100-to-1 crack/powder cocaine ratio (Tonry 2010). But a growing gap between Black and White offenders in sentencing decisions would exacerbate the problem and would deserve attention because the Sentencing Reform Act of 1984 was designed in part to counteract race disparity (Breyer, 1988).

The study by Ulmer et al. (2011), in contrast, finds more modest effects, which are largely confined to the prison/probation decision and to immigration offenses. Sentence-length disparity between Black male and White male offenders has indeed increased since *Booker*, the study concluded, but the effect was approximately 40% smaller than estimated by the Commission, and there is no significant difference between levels of disparity in the pre-PROTECT Act period (2002–2003) and the most recent post-*Gall* period. In other words, for sentence length, "Black male disparity returned to the pre-PROTECT state in the wake of *Gall*" (Ulmer et al.). Sentence-length effects for Black male offenders also shrink considerably in models that exclude immigration offenses. Yet the alternative analysis finds that some forms of disparity in imprisonment decisions have increased since *Booker* and *Gall*. For Black male offenders, regardless of whether immigration offenses are excluded,

1. The Commission's "refined" models found that during the PROTECT Act period, Black male offenders received sentences 5.5% longer than those for White male offenders. The gap grew to 15.2% in the post-*Booker* period and to 23.3% in the most recent post-*Gall* period.

unexplained disparity in the prison/probation decision is significantly higher under the advisory Guidelines post-*Gall* than under the PROTECT Act.[2]

The authors of the alternative analysis attribute the studies' divergent results to different methodological choices. Their analysis models the prison/probation decision separately, disentangling race disparity in the "in/out" decision from disparity in sentence length among offenders who receive a prison sentence. It also controls for criminal history using both the criminal history score and the guideline minimum sentence, which is determined in part by reference to criminal history. In some models, the alternative analysis by Ulmer et al. also excludes immigration offenses because of concerns raised by fast-track programs and noncitizen offenders.

Points of Agreement

Rather than attempt to referee the methodological sparring, this essay focuses on the substantial areas of overlap between the two studies. Despite their differences, the Commission's research and the analysis by Ulmer et al. (2011) reach similar conclusions in several key respects.

First, both the Commission's research and the analysis by Ulmer et al. (2011) seem to agree that the choice between prison and probation is a source of persistent and increasing race disparity. Ulmer et al.'s analysis found "an unexplained increase in Black males' odds of imprisonment post-*Gall*," which is statistically significant compared with the PROTECT Act period, both for federal offenses as a whole and for the subset of nonimmigration offenses. Although the Commission's most recent study did not analyze the "in/out" decision separately, its previous research found that, despite year-to-year fluctuations, for fiscal years 1998 to 2002 overall, the odds of imprisonment were 20% higher for Black males than for White males (USSC, 2004: Figure 4.4). Notably, neither study was able to control for the full range of considerations that a judge might consider in selecting a prison sentence, including the offender's employment record and the degree of violence in the offender's criminal history.[3] Yet within the limits of the available data, the studies seem to agree that for Black male offenders, significant and increasing unexplained race disparity exists in federal imprisonment decisions.

Second, both the Commission study and the analysis by Ulmer et al. (2011) seem to agree that the evidence of race disparity under the mandatory Guidelines, prior to the PROTECT Act in 2003, was volatile and inconclusive. For example, the Commission's earlier "unrefined" models found that Black offenders received longer prison terms than

2. Logistic regression models calculate a coefficient for Black males of 0.209 in the post-*Gall* period, which is significantly greater than the PROTECT Act period ($b = 0.046$) and post-*Booker* periods ($b = 0.084$), but not significantly greater than the pre-*Booker* period ($b = 0.101$).

3. Both studies control for the guideline minimum sentence, which reflects hundreds of non-race factors related to the nature of the offense and the history and characteristics of the offender, as detailed in chapters two, three, and four of the Guidelines Manual.

White offenders by as much as 14.2% in fiscal year 1999, but by a more modest 8.2% in fiscal year 2001, and by a statistically insignificant amount in fiscal year 2002 (USSC, 2006: Figure 13). Those results echoed the Commission's previous research, which reported that race disparity for Black men fluctuated considerably between 1998 and 2002, with no statistically significant race disparity for some sentencing outcomes in some years (USSC, 2004: Figures 4.7 and 4.8). That instability makes it difficult to draw reliable comparisons between the current advisory guidelines and the pre-PROTECT Act mandatory Guidelines.

Third, and most surprisingly, both the Commission's models and the analysis by Ulmer et al. (2011) seem to agree that race disparity for Black men reached its lowest levels under the PROTECT Act in 2003 and 2004. The Commission's "refined" models for Black male offenders, as replicated by Ulmer et al., found significantly lower levels of race disparity in sentence length under the PROTECT Act than under the advisory Guidelines after *Booker* and *Gall*.[4] The Ulmer et al. alternative analysis reached the same conclusion. Its models of sentence length for Black male offenders receiving a prison sentence found significantly lower levels of unexplained race disparity under the PROTECT Act than in the pre-PROTECT Act and post-*Gall* periods (Ulmer et al., 2011: Table 2).[5] Likewise, models of the incarceration decision for Black males found no statistically significant Black-male effect during the PROTECT Act period and found a race effect significantly smaller than under the advisory Guidelines post-*Gall* (Ulmer et al., 2011: Figure 4).[6] According to Ulmer et al., "the post-*Booker* era has brought greater sentence length racial disparity disadvantaging Black males," but "*only* when one's basis of comparison is the PROTECT era." That finding is important and entirely consistent with the Commission report.

Possible Paths Forward

Although narrow, those apparent points of agreement in the race disparity research have potentially important implications for sentencing law and policy. Consider several possible paths forward: (a) Create a system of "dispositional departures" by tweaking the Guidelines to regulate explicitly the choice between prison and probation; (b) restore the PROTECT Act, with its strict controls over judicial discretion, augmented by jury fact finding; (c) Adopt Judge Sessions's proposal, which would simplify the Guidelines while making

4. The linear regression models of sentence length for Black males calculated a coefficient of 0.089 during the PROTECT Act period, which is significantly lower than the post-*Booker* period ($b = 0.164$), and the post-*Gall* period ($b = 0.217$), but not significantly different from the pre-PROTECT Act period ($b = 0.130$).

5. The linear models of sentence length for Black men receiving a prison sentence produced a coefficient of 0.045 in the PROTECT Act period, which is significantly lower than the pre-PROTECT Act period ($b = 0.066$) and the post-*Gall* period ($b = 0.077$) but not significantly different from the post-*Booker* period ($b = 0.053$).

6. The logistic models of the incarceration decision for Black males calculated a nonsignificant coefficient of 0.046 for the PROTECT Act period, which is significantly lower than the post-*Gall* period ($b = 0.209$) but not significantly different from the pre-PROTECT Act period ($b = 0.101$) or the post-*Booker* period ($b = 0.084$).

them “presumptive” and widening guideline imprisonment ranges; or (d) None of the above.

Dispositional Departures

Ulmer et al. (2011) suggest that because their analysis provides clear evidence of increasing race disparity in the prison/probation decision, policy makers should consider changes “specifically targeted” at that stage of sentencing. Currently, federal sentencing rules leave the initial choice between prison and probation almost entirely unregulated. For cases in zone A of the sentencing grid (0–6 months of imprisonment), a sentence of probation is clearly permitted. Likewise, under U.S.S.G. § 5B1.1(a), for cases in zone B of the sentencing grid, a sentence of probation is “authorized” in combination with other conditions, subject to a few restrictions. But in neither case is a sentence of probation required, or even presumed; the decision rests entirely with the sentencing judge. Stronger guidance from the Commission might help counteract the emergence of unexplained race disparity in choosing from among available sanctions.

One option for addressing race disparity in the prison/probation decision is to create a presumption of prison or probation for different cells of the sentencing grid, coupled with a system of appellate review for “dispositional departures” that deviate from that presumption. Several state sentencing guidelines systems operate in that manner, including Minnesota and Kansas (Frase, 2006). And there is reason to believe the dispositional-departure model can succeed. Frase (2009) uncovered little evidence of race disparity in Minnesota sentencing decisions, although unexplained disparity was found in other parts of the state’s criminal justice system. Of particular relevance, after other legal and extralegal factors were taken into account, race has not been a significant predictor of an executed (rather than a stayed) prison sentence in Minnesota for the last 15 years (Frase, 2009).

Given the complexity of the federal Guidelines, there cannot be much appetite among judges and lawyers for yet another layer of presumptions, departures, and appellate review standards. In addition, to avoid constitutional problems, a system of dispositional departures would be required either to afford offenders a right to a jury trial with respect to facts that rebut a presumption of probation, or to specify that the presumption is merely advisory. Nonetheless, because the latest research on race disparity suggests that prison/probation decisions are an area of special concern, explicit regulation of the choice among sanctions deserves a look.

Restore the PROTECT Act, with Jury Fact Finding

A much more drastic option is to roll back the *Booker* remedial opinion and restore federal law as it stood under the PROTECT Act in 2003 and 2004, augmented by the right of the accused to have a jury find any facts required by the Sixth Amendment. That was precisely the remedy proposed by Justice Stevens in dissent in *Booker*, and some members of Congress introduced legislation to the same effect in the immediate aftermath of the decision (Bowman, 2005). The theory is that tight constraints on judicial discretion, of the kind that prevailed under the PROTECT Act, can minimize the chances that implicit race bias on the part of judges will taint sentencing outcomes.

Ulmer et al. (2011) find no reason, in light of their research, to "globally constrain federal judges' sentencing discretion as a remedy for disparity." They suggest that the PROTECT period was an "anomaly" because "racial and gender sentence length disparities are less today, under advisory Guidelines, than they were [in the pre-*Koon* period in 1994 and 1995; see *Koon v. United States*, 1996] when the Guidelines were arguably their most rigid and constraining." It is a mistake, however, to equate the pre-*Koon* and PROTECT Act periods. The PROTECT Act ushered in the "most rigid and constraining" period in federal sentencing history, and by a country mile. Whereas the pre-*Koon* standard of appellate review was unsettled, with most circuits adopting a three-tier approach with a mixture of "reasonableness" and *de novo* review (Lee, 1997), the PROTECT Act rendered the Guidelines more inflexible and unyielding than ever. The Act not only repudiated the abuse-of-discretion standard of review announced in *Koon* but also specified that an appellate review of sentences would be *de novo*, directed the Commission to reduce the incidence of downward departures, saddled judges with new reporting requirements for non-guideline sentences, and directed prosecutors to resist downward departures (Scott, 2010). The rate of judge-initiated, below-Guidelines sentences plunged to an estimated 5% after the PROTECT Act, lower than during the run-up to *Koon* (Stith, 2008: Figure 2; USSC, 2003: Figure 16; USSC, 2006).[7] The Act was widely described as the most fundamental power shift in federal sentencing since the inauguration of the Guidelines (Barkow, 2005; Stith, 2008; Tiede, 2009). To test the effectiveness of "rigid and constraining" sentencing guidelines, one cannot do better than the PROTECT Act, and at least with respect to unexplained race disparity, the results are surprisingly encouraging.

I harbor no affection for PROTECT Act, which I have criticized previously for its hasty enactment, its reliance on flawed departure data, and its myriad intrusions into the province of judges, the Commission, and even prosecutors (Scott, 2010). The primary objective was not to rectify racial injustice but simply to decrease downward departures. Moreover, like other proposals laser-focused on judicial discretion, restoring the PROTECT Act regime would do nothing to alleviate race disparity in the decisions of police and prosecutors, or in the content of the Guidelines and statutory sentencing ranges. Still, the favorable marks that the new studies give to the PROTECT Act, which Ulmer et al. (2011) describe as a "truly unusual period in the history of the Guidelines," provide at least modest support for the premise that mandatory guidelines with robust appellate review can reduce unexplained race disparity in sentencing decisions.

7. The Commission's reported downward departure rates throughout these years are misleading because of the misclassification of fast-track sentences in border districts throughout the 1990s. Stith (2008) estimated revised rates for 2001–2007 based on the Commission's research.

The Sessions Proposal

In a widely discussed new article, Judge William K. Sessions III (in press), a Clinton appointee and President Obama's choice to chair the Sentencing Commission, proposed a dramatic restructuring and simplification of the federal Sentencing Guidelines. Among other changes, Judge Sessions recommended streamlining individual guidelines, simplifying the sentencing table, widening punishment ranges to afford judges greater flexibility, and abrogating mandatory minimum sentences. Most important, for the current purposes, he proposed to replace the current advisory system with "presumptive" guidelines, subject to meaningful appellate review. That change is necessary, he argued, to accomplish Congress's goal of eliminating unwarranted sentencing disparity, including demographic disparity. Indeed, Judge Sessions cited the Commission's research as evidence of increasing race disparity in the wake of *Booker* and, thus, as support for his revisions (Sessions, in press).

The proposal by Sessions (in press) does not, however, call for a system of unyielding Guidelines with *de novo* appellate review, as under the PROTECT Act. To the contrary, Judge Sessions is critical of the Act and especially its directives to the Commission. Instead, he proposes to "resurrect" presumptive Guidelines and to install a new form of appellate review in which within-range sentences are essentially unreviewable, while review of departures would involve "relatively strict scrutiny." At the same time, the proposal by Sessions would widen sentencing ranges significantly to make the Guidelines simpler and more flexible. Thus, on a continuum between today's advisory Guidelines, on the one hand, and the strict and inflexible system of the PROTECT Act, on the other hand, the proposal by Sessions falls somewhere in between.

The Commission study and the alternative analysis by Ulmer et al. (2011) provide only mixed support for such a middle-ground proposal. The closest real-world analog to the system Judge Sessions (in press) envisions is the system of "presumptive" guidelines he seeks to partially resurrect, which prevailed from 1996 to the PROTECT Act in 2003. At best, however, that regime performed erratically. The Commission's research indicates that the levels of race disparity for Black and Hispanic men fluctuated considerably throughout that period, without apparent explanation (USSC, 2004, 2006). The best support for a Sessions-style proposal comes from the Commission's "refined" models, which report that levels of race disparity for Black male offenders were significantly lower in the immediate pre-PROTECT Act period than under the advisory Guidelines in the *Booker* and *Gall* periods (Ulmer et al., 2011: Figure 2). But Ulmer et al.'s analysis, both for incarceration decisions and sentence length, finds no significant difference between the pre-PROTECT and post-*Gall* periods for Black men. Indeed, setting aside immigration offenses, Ulmer et al. conclude that significantly *less* race disparity exists in the post-*Gall* period than in the pre-PROTECT period. Because much depends on the researchers' methodological choices, we have only mixed evidence that the proposal would reduce unexplained race disparity.

None of the Above

In the end, however, the best approach may be "none of the above." The studies by USSC (2010) and Ulmer et al. (2011) underscore some fundamental challenges in identifying trends in race disparity in federal sentencing. First, as McDonald and Carlson (1993: 106) observed, "[a]ny findings that are sensitive to minor changes in model specifications such as these must be interpreted with caution." Here, basic choices about how to model the sentencing decision, how to control for criminal history, and how to disentangle noncitizen effects from race effects seem to have serious consequences. Signs of increasing race disparity deserve continuing vigilance, but the competing analyses in these studies suggest that no robust trend has yet emerged under the advisory Guidelines.

Second, the Commission's previous research suggests that considerable "noise" exists in race disparity trends. Between 1998 and 2002, for example, levels of unexplained race disparity in sentencing outcomes for Black and Hispanic men swung wildly from year to year, with no obvious explanation. Nonrace factors could be to blame for those "unstable" results, including omitted variable problems and changes in multicollinearity between race and other independent variables (USSC, 2004). It is encouraging that, according to both studies, the short-lived experiment with the PROTECT Act produced historically low levels of unexplained race disparity at sentencing for Black male offenders. Yet the Supreme Court brought the PROTECT Act era to an abrupt end after just 15 months, and we have no way of knowing whether the results would have persisted.

I do not mean to suggest that research into unexplained race disparity at sentencing is hopeless. Both the results and the methodological discussion in these studies make valuable contributions to debates over the future of federal sentencing. To move policy makers, however, evidence of a trend in race disparity will have to be robust and sustained. So far, the race disparity research, standing alone, is insufficient to justify sweeping changes.

Race Disparity and Interjudge Disparity

Of course, research on race disparity does not stand alone. Another primary objective of the Sentencing Reform Act was to curtail *interjudge* disparity, driven not by legitimate differences between offenses and offenders but by the preferences, punishment philosophies, and idiosyncrasies of individual judges (Breyer, 1988). Studies in the late 1990s found that the Guidelines had in fact succeeded in decreasing that form of unwarranted disparity (Anderson, Kling, and Stith, 1999; Hofer, Blackwell, and Ruback, 1999). But preliminary empirical work focused on one district court suggests a sharp increase in interjudge disparity in the wake of *Booker* and *Gall* (Scott, 2010).

If subsequent research on interjudge disparity were to detect the same trend nationwide, policy makers might consider the same options implicated by research on race disparity: the Sessions proposal, a modified PROTECT Act system, or more targeted changes. Today, evidence of a surge in unexplained race disparity is too equivocal to justify sweeping changes in sentencing law. But as our understanding of post-*Booker* sentencing improves, it could form a crucial part of a broader case for changes to the advisory Guidelines system.

References

Anderson, James M., Jeffrey R. Kling, and Kate Stith. 1999. Measuring interjudge sentencing disparity: Before and after the federal sentencing guidelines. *Journal of Law & Economics*, 42: 271–308.

Barkow, Rachel E. 2005. Our federal system of sentencing. *Stanford Law Review*, 58: 119–136.

Breyer, Stephen. 1988. The federal sentencing guidelines and the key compromises upon which they rest. *Hofstra Law Review*, 17: 1–50.

Bowman, Frank O. III. 2005. Letter from Frank Bowman concerning H.R. 1528. *Federal Sentencing Reporter*, 17: 311–314.

Frase, Richard S. 2006. *Blakely* in Minnesota, two years out: Guideline sentencing is alive and well. *Ohio State Journal of Criminal Law*, 4: 73–94.

Frase, Richard S. 2009. What explains persistent racial disproportionality in Minnesota's prison and jail populations? *Crime and Justice*, 38: 201–280.

Hofer, Paul J., Kevin R. Blackwell, and R. Barry Ruback. 1999. The effect of the federal sentencing guidelines on inter-judge sentencing disparity. *Journal of Criminal Law and Criminology*, 90: 239–322.

Lee, Cynthia K. Y. 1997. A new "sliding scale of deference" approach to abuse of discretion: Appellate review of district court departures under the federal sentencing guidelines. *American Criminal Law Review*, 35: 1–56.

McDonald, Douglas C. and Kenneth E. Carlson. 1993. *Sentencing in the Federal Courts: Does Race Matter? The Transition to Sentencing Guidelines, 1986–90*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Scott, Ryan W. 2010. Inter-judge sentencing disparity after *Booker*: A first look. *Stanford Law Review*, 63: 1–66.

Sessions III, William K. In press. At the crossroads of the three branches: The U.S. Sentencing Commission's attempts to achieve sentencing reform in the midst of inter-branch power struggles. *Journal of Law & Politics.* Retrieved August 24, 2011 from ssrn.com/abstract=1773045.

Stith, Kate. 2008. The arc of the pendulum: Judges, prosecutors, and the exercise of discretion. *The Yale Law Journal*, 117: 1420–1497.

Tiede, Lydia Brashear. 2009. The swinging pendulum of sentencing reform: Political actors regulating district court discretion. *BYU Journal of Public Law*, 24: 1–48.

Tonry, Michael. 2010. The social, psychological, and political causes of racial disparities in the American criminal justice system. *Crime and Justice*, 39: 273–312.

Ulmer, Jeffery T., Michael T. Light, and John H. Kramer. 2011. Racial disparity in the wake of the *Booker/Fanfan* decision: An alternative analysis to the USSC's 2010 report. *Criminology & Public Policy.* This issue.

U.S. Sentencing Commission. 2003. *Downward Departures from the Federal Sentencing Guidelines*. Washington, DC: Author.

U.S. Sentencing Commission. 2004. *Fifteen Years of Guidelines Sentencing: An Assessment of How Well the Federal Criminal Justice System Is Achieving the Goals of Sentencing Reform*. Washington, DC: Author.

U.S. Sentencing Commission. 2006. *Final Report on the Impact of United States v. Booker on Federal Sentencing*. Washington, DC: Author.

U.S. Sentencing Commission. 2010. *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis*. Washington, DC: Author.

Court Cases Cited

Gall v. United States, 552 U.S. 38 (2007).
Kimbrough v. United States, 552 U.S. 85 (2007).
Koon v. United States, 518 U.S. 81 (1996).
Rita v. United States, 551 U.S. 338 (2007).
United States v. Booker, 543 U.S. 220 (2005).

Statutes Cited

PROTECT Act of 2003, Pub. L. 108–21, 117 Stat. 650, S. 151, enacted April 30, 2003; amendment Stat. 667.

Sentencing Reform Act of 1984, 18 U.S.C. §§ 3551–3626 and 28 U.S.C. §§ 991–998, as amended 1985–1988, 1990, 1992, 1994 and 1996.

Ryan W. Scott is an Associate Professor at Indiana University Maurer School of Law in Bloomington, Indiana.

Racial disparity in the wake of *Booker/Fanfan* Making sense of "messy" results and other challenges for sentencing research

Rodney Engen
University of Arkansas

In the three decades since the first comprehensive sentencing guidelines were adopted in MN, PA, and WA scholars have produced a substantial body of research examining if, when, and how race, ethnicity, or gender still "matter" in the sentencing process (in state and federal systems). Yet it is difficult to comment on the impact of sentencing guidelines on sentencing disparity because there simply is little empirically rigorous research examining the effects of actual policy changes (i.e., the introduction, modification, repeal, etc., of sentencing laws) for sentencing practices (Engen, 2009). In this context, the study by Ulmer, Light, and Kramer (2011, this issue), and the U.S. Sentencing Commission report (USSC, 2010) to which it responds, constitute important contributions to the sentencing literature.[1]

Like the USSC study, Ulmer et al. (2011) examine racial/ethnic and gender disparity in sentencing under the U.S. guidelines prior to, and subsequent to, several key pieces of legislation and U.S. Supreme Court decisions affecting judicial discretion, most notably the *Booker* and *Gall* decisions. However, Ulmer et al. also employ numerous methodological "choices" (a wonderfully diplomatic phrase) different from those of the Commission's study. Consequently, Ulmer et al.'s findings also differ in some important ways from those of the USSC study. They find that the increase in disparity between White and Black males is primarily related to imprisonment decisions, as opposed to sentence length, and is largely due to sentencing disparity in immigration cases. Ulmer et al. also find that controlling for criminal history explains a substantial proportion of the disparity found by the Commission

Direct correspondence to Rodney Engen, Department of Sociology & Criminal Justice, University of Arkansas, 211 Old Main, Fayetteville, AR 72701 (e-mail: rengen@uark.edu).

1. See also Wooldredge, Griffin, and Rauschenberg (2005) and Wooldredge (2009) on the impact of reforms in Ohio. While conceptually similar, the guidelines in Ohio have little in common with the federal guidelines.

DOI:10.1111/j.1745-9133.2011.00773.x

and that departure sentences continue to contribute to sentencing disparity. Perhaps most importantly, they challenge the Commission's (USSC, 2010) main conclusion that disparity in sentence length associated with race and gender increased significantly post-*Booker* and post-*Gall,* finding that disparity post-*Booker/Gall* is not necessarily greater than it was in some previous guideline periods. If anything, any recent changes may reflect a return to previous patterns of sentencing behavior under the guidelines.

Overall, Ulmer et al. (2011) conclude that little compelling evidence exists of the large increases in sentencing disparity that some feared might follow in the wake of these rulings. I concur. At best, the evidence is inconsistent regarding whether disparity worsened post-*Booker* and post-*Gall*, but there clearly is no evidence of an urgent need for legislation to counteract the supposedly deleterious effects of increased judicial discretion. However, just as Ulmer et al.'s reanalysis of the USSC data demonstrates the value of independent analyses employing different strategies and methodological "choices," previous policy essays in CPP confirm the value of having more than one "set of eyes" examine and contemplate empirical findings. In this spirit, I offer a few observations regarding some of Ulmer et al.'s own methodological choices and their findings, some of which differ from the authors' interpretations. I then turn to some larger issues beyond the scope of their study, which may prove challenging for research assessing the impact of changes in sentencing policy. In short, like most good research, the study prompts as many questions as it provides answers.

Some Additional Thoughts re Ulmer et al.

Among Ulmer et al.'s (2011) methodological "choices," perhaps the most consequential one, with respect to policy implications, is the decision to include cases sentenced prior to the PROTECT Act. Including the previous time periods revealed that recent changes in sentencing practices—assuming they are in any way consequences of policy changes—may have more to do with undoing the PROTECT Act than with *Booker/Gall*. Regardless of the interpretation, however, the findings remind one of how sensitive interrupted time-series designs can be to the duration of the time series selected and the specific comparisons made. If Ulmer et al. had examined sentencing practices only in the same time periods included in the USSC's study they might have concluded, like the USSC, that disparity increased post-*Gall*. Of course, strictly speaking, this might be true, but the meaning of any such increase changes when viewed in this broader historical context.

This observation also leads me to wonder what we might find if we could extend the timeline back even further, to the pre-guideline era. How would post-*Gall* sentencing compare with sentencing prior to the sentencing guidelines? Then again, I am not even sure this is a meaningful question. In other words, what is the appropriate comparison or benchmark for assessing the impact of recent court decisions, sentencing practices as they were twenty five years ago, ten years ago, or practices just prior to these court decisions? Neither seems an ideal reference. It may be more meaningful to ask how the exercise of discretion is related to the structure of sentencing laws in place at any given time.

Ulmer et al. (2011) also emphasize that a "substantial amount of racial–gender disparity can be attributed to immigration offenses" (p. 1098) and they suggest future research examining "the greater racial disparity among immigration cases" explicitly. I would qualify and expand on this observation in several ways. First, if sentence length disparity is greater in immigration cases, their results suggests it does *not* pertain uniquely to Black males, as implied in their conclusions. With the exception of "other male" and "other female," *all* the race-gender effects on sentence length seem to be smaller when immigration cases are removed (see Figure 3 and Table 1 of Ulmer et al.). Second, the effect of being a Black male on the likelihood of incarceration is nominally *larger* when they remove immigration cases. Among the full sample, "Black male" is not significant until post-*Gall*, whereas among nonimmigration cases they find significant Black male effects pre-PROTECT and post-*Booker* as well. It seems, then, that including immigration cases *suppresses* a significant Black male effect in most eras and that Black males may be relatively *less* likely than other groups to receive a prison sentence in immigration cases. If so, then is it really Hispanic/non-Hispanic imprisonment disparity that is concentrated in immigration cases? What about drug offenses? Should we expect the effects of increased discretion to vary by offense type, generally? Future research probably should consider whether the impact of *Booker* and *Gall* differs by offense.

Making Sense of Messy Results

The consequences of policy changes are seldom, if ever, predicted accurately either by scholars, by practitioners, or by the policy makers themselves. This result is well known, at least among social scientists, since Merton described "The unanticipated consequences of purposive social action" in 1936. Ulmer et al.'s findings are no exception (see also Wooldredge, Griffin and Rauschenberg, 2005). Although they find limited evidence of the "anticipated" increase in racial disparity (namely, between Black and White men) the study is also replete with "unanticipated" results. Would anyone have anticipated an increase in disparity *only* between Black and White males? Why did disparity between Hispanic and non-Hispanic White males not increase (it actually decreased, in the sense that Hispanic men received slightly *shorter* sentences pre-*Gall*)? And why did disparity *decrease* post-*Gall* for Black and Hispanic women versus White men? Equally interesting to me, no significant difference was found in the likelihood of incarceration between White and Black men in any of the earlier periods; yet in these same periods, judges were less likely to sentence women of all races to prison and were more likely to sentence Hispanic men to prison. Why are they discriminating against Black men only now, if indeed that is what these findings represent? In short, no consistent evidence exists of an increase in disparity and, taken at face value, the findings suggest that disparity overall *declined*. What is more, this decline suggests that three of four minority groups received *preferential* (or more preferential) treatment before *Gall*.

How are we to make sense of such "messy" and unanticipated results? In short, I am not sure we can make sense of them, or that we should even accept them at face value. Rather, in the discussion that follows I suggest that some of these confusing findings (as with many results of sentencing research) may be a consequence of other methodological choices made by Ulmer et al. (2011) and by most sentencing researchers, and of data limitations plaguing most sentencing research.

The Problem with Statistical Significance

Ulmer et al. test whether unwarranted disparity increased or decreased in the wake of *Booker* and *Gall* by estimating race-gender coefficients before and after these rulings and by testing whether they differ across time periods by statistically significant margins. Unfortunately, statistical significance has never been a good measure of the substantive importance or "size" of effects in multivariate analyses, and it can be especially misleading when analyzing large data sets like those commonly used in sentencing research. The problem is not, as I have heard it described, that very large "*N*s" somehow produce significant differences or effects where none exist but that large *N*s allow us to measure associations with great precision, thus increasing our confidence that the slope coefficients in our models were not obtained by chance (which is not a bad problem to have). The consequence, however, is that even small associations often achieve "significance," and we often will reject the null hypothesis of no difference in slope coefficients obtained in two or more time periods, or across groups, even when those differences are small.

For example, I refer readers to Ulmer et al.'s (2011) Appendix B, where their full models of sentence length are presented. Most coefficients in these models are statistically significant, and a majority of them differ significantly over time in at least one comparison. This includes the effect of the presumptive sentence on sentence length, which was *larger* post-*Booker* and post-*Gall*. Taken at face value, this suggests that judges sentenced more closely to the guidelines when they ostensibly were free to deviate (we might have predicted the opposite). I would find this unanticipated pattern interesting, but the differences are so small I am more inclined to view them as random fluctuations having little substantive importance. Whereas a 1% increase in the presumptive sentence length produced a 0.669 (or 67%) increase in sentence length pre-PROTECT, it produced a 0.689 (or 69%) increase in sentence length post-*Gall* ($z = -13.217$; $p < .001$). For an offender with a 5-year presumptive sentence, this difference amounts to a little more than 1 month. I cannot help wondering, also, how large really are the differences over time in the size of the race/ethnicity/gender coefficients? One challenge for policy analyses such as this, and for disparity research generally, is in determining when the observed disparities, or changes in disparity, are substantively meaningful (see Langan, 2001). Calculating predicted sentence lengths and probabilities of incarceration for each offender group, under different contingencies, in each time period, might give us a better sense of which of the observed changes are both "real" and substantively important.

The Problem of Departures

A final methodological choice made by Ulmer et al. (2011), and by many sentencing researchers, that may have substantive import is the decision to control for sentence departures in the models predicting sentence length and incarceration. Although this is common practice in the literature, controlling for what may be the main source of disparity under the guidelines—departures—means that, in essence, we are focusing our attention on disparity that *cannot* be attributed to departures. This is akin to pointing out that Black and Hispanic defendants are less likely to receive the benefit of downward departures but then limiting our analyses of disparity to cases that were denied the best breaks of all (departures). It should perhaps come as no surprise, then, that the findings are inconsistent. Although an interesting question, analyses that only examine disparity that is net of the disparity in departure do not reveal disparity in the sentencing process overall.

Making sense of any changes in sentencing or in disparity over time (and subsequent to major policy changes) also is difficult when we control for departures, especially in light of the evidence that disparity in the likelihood of receiving downward departure also has changed (see Appendix C in Ulmer et al., 2011). If the concern is that judges or prosecutors will be even more discriminatory in their use of departures post *Booker/Gall*, I believe we could learn more by first estimating changes in disparity in incarceration and sentence length *prior to* controlling for departures (i.e., changes in the "total" effects of race, ethnicity, and gender) and then testing whether controlling for departures explains the changes in disparity. Without first knowing the main or total effects of race-gender, we cannot say with much confidence whether disparity in the sentencing process has changed, remained constant, or simply moved.

Judging Judicial Discretion and Other Challenges for Sentencing Research

At this point, I would like to call attention to some larger challenges and limitations of sentencing research, generally, for assessing the effects of court rulings like *Blakely*, *Booker*, and *Gall*, or any other policy changes. I find it difficult to say much on the topic that is new, especially in light of the thoughtful essays in the 2007 special section of *Criminology & Public Policy* addressing this question (see Bushway and Piehl, 2007; Frase, 2007; Hofer, 2007; Wellford, 2007), but some ideas may bear repeating and further development. I encourage interested readers to consider those essays carefully. In my subsequent comments, I attempt to build on some of those arguments as well as two points raised by Ulmer et al. (2011), as they pertain to research testing the impact of policy changes on sentencing and disparity.

The Problem of Prosecutorial Discretion in the Sentencing Process

The general unavailability of data on charging and plea bargaining remains, in my opinion, the greatest challenge to the validity of sentencing research and perhaps especially research assessing the effects of policy changes on sentencing (Bushway and Piehl, 2007; Frase, 2007). It is well known that prosecutors and defense attorneys routinely negotiate over charges and other "facts" with an eye toward the sentence that is likely to result (Ulmer, 2005; Ulmer et al., this issue). Acknowledging this reality, sentencing researchers, including Ulmer et al., are careful to limit their inferences to "judicial" discretion and disparity at the formal sentencing stage, and to note that important sources of disparity might be overlooked. However, *the lack of data on charging and plea bargaining also threatens the validity of inferences about decision making at the sentencing stage.* Blumstein et al. (1983) described this as:

> [T]he problem of classifying "like cases".... Cases that appear alike initially may, on closer scrutiny, differ in subtle ways . . . or in not-so-subtle ways (e.g., two cases in which the conviction offenses are the same as a result of plea negotiations may differ substantially in the actual underlying offense behavior). (p. 267)

If judges take it into account these unmeasured but "real" differences, it might result in very different sentences for offenders who appear to be "similarly situated" based on their conviction offenses and criminal histories (Wilmot and Spohn, 2004). If these unmeasured differences are related to defendants' race, ethnicity, and sex, then estimates of sentencing disparity may be biased.

Controlling for prosecutorial decisions might be even more important in research assessing the impact of policy changes on sentencing. If we wish to make strong inferences about the consequences of policy changes we ideally must either include measures of offending behavior and/or plea bargaining that are independent of the conviction offense (Frase, 2007) or we must assume that prosecutorial practices remained relatively stable. If prosecutorial behavior did not remain stable—and Ulmer et al. report that, at least with regard to departures, it did not—then we have the same problem of being unable to compare "like with like" over time. Unfortunately, when sentencing policies change, court actors—including prosecutors—often adapt in unanticipated ways (Engen and Steen, 2000).[2] Referring to *Blakely* and related decisions, Richard Frase (2007: 404) remarked, "the cases [may] have their most important lasting effects not on sentencing practice but on charging decisions and on the design of sentencing laws and guidelines."

The Problem of Disparity in the Guidelines

A potentially more difficult issue raised by both Hofer (2007) and Bushway and Piehl (2007), and undoubtedly familiar to many readers, is that the sentencing guidelines themselves are

2. Hofer (2007), in his analysis of aggregate sentencing trends pre- and post-*Booker*, found both an increase in departures post-*Booker* and an increase in average presumptive sentences. Among the possible explanations he offered for this somewhat paradoxical pattern is that federal prosecutors may have increased the severity of charges and aggravating facts presented in response to the increase in judicial discretion. If this is correct, then cases sentenced for identical crimes pre- and post-*Booker* are almost certainly *not* alike.

not neutral. Racial disparity may be built into the guidelines, especially in the realm of drug offenses (Tonry, 1995). The crack versus powder cocaine differential is a textbook example (literally) of institutionalized bias. Similarly, because of the counting of drug offenses, Black offenders disproportionately "qualify" for the career criminal enhancement that, according to Hofer, contributes significantly to the disparity in average sentence lengths between Black and White offenders (see also Bushway and Piehl, 2007). Undoubtedly, we could find other examples where one group or another is inherently disadvantaged, and I second Bushway and Piehl's call for research examining the structure of the guidelines in this way.

This poses several difficult challenges for sentencing research: First, *how are we to evaluate the exercise of judicial discretion, or changes in sentencing disparity, relative to guidelines that many observers, including federal judges, believe are unjust?* Should we be relieved or concerned if several studies find that, for the most part, judges are continuing to sentence in accordance with the guidelines even post-*Booker/Gall*? If the laws are indeed unfair, then the uniform application of such laws would only perpetuate injustice. Although I expect many researchers would agree with this last statement, we implicitly assume guidelines neutrality when, in Bushway and Piehl's (2007: 479) words, "the sentencing grid defines the starting point of analysis." The problem, then, is if we cannot use the guidelines as the benchmark against which to judge judicial behavior, what do we use? Ultimately, sentencing research will need to develop some criteria for assessing punishment disparity, or measures capturing the seriousness of offending and other "legitimate" concerns that are independent of the guidelines.

Second, equal treatment (i.e., the absence of disparity) is not the only criterion relevant to judging the impact of these recent decisions or of sentencing guidelines generally. *Substantive "justice" also is important*. Although this concept is much harder to define than "disparity," evidence of widespread dissatisfaction with the federal sentencing guidelines (Frase, 2007) suggests the laws do not always reflect normative expectations for appropriate punishment. The important point for the current discussion is that the exercise of judicial discretion under the guidelines—and changes to the guidelines themselves—may increase one type of justice while diminishing the other. For instance, a number of studies suggest that the frequent use of departures in the sentencing of drug offenders may contribute to unwarranted disparity, undermining justice defined as equal treatment. At the same time, the use of departures in these cases may actually enhance substantive justice by mitigating the impact of sentencing laws that many view as excessively harsh. In principle, *Booker* and *Gall* could have similar effects. If judges use their increased discretion to counteract some of the more extreme and controversial aspects of the guidelines (e.g., mandatory minimums or the career offender enhancement), they might hand down sentences that most observers would agree are more "appropriate," on average, than if they had followed the guidelines closely. If so, it is most likely that all races, ethnicities, and genders will benefit from this discretion. At the same time, experience shows that these groups might not benefit equally. In this hypothetical, but highly plausible, scenario wherein all groups benefit, but some benefit more than others, would we conclude that the *Booker/Gall* rulings improved or undermined the quality of justice in U.S. courts? The answer might depend on how we define justice.

Third, we should bear in mind that *the consequences of any change to sentencing laws may depend on the level of analysis we choose to examine*. Sentencing research characteristically focuses on the effects of individual offender and offense characteristics. However, because Black, White, Hispanic, male, and female defendants differ in the aggregate on such factors as average offense seriousness, criminal history, and conviction offense, any changes in the importance placed on these criteria—whether discretionary in origin, legislated, or the result of court rulings—will impact groups differently in the aggregate (Bushway and Piehl, 2007; Hofer, 2007). For example, steps taken in recent years to reduce the "crack versus powder cocaine" differential have probably done more to reduce racial disproportionality in incarceration than any other policy changes adopted since the guidelines were introduced. Importantly, this effect is likely to happen *irrespective of disparity at the individual level*. Similarly, recent court rulings increasing judicial discretion—or future rulings and legislation restricting it once again—could impact racial disproportionality in the aggregate differently than they impact disparity at the individual level. For instance, to continue my earlier example, if judges use their newfound discretion to reduce sentences for drug offenders substantially, but still give preferential treatment to Whites, we could even observe an *increase* in disparity at the individual level accompanied by a *reduction* in racial disproportionality in the aggregate. Again, would we conclude that *Booker/Gall* have increased or decreased racial justice?

Conclusion: *"There's Nothing So Practical as a Good Theory"* (Kurt Lewin)

Perhaps the most difficult challenge for research hoping to make sense of major policy changes is that we have little pertinent theory to guide us. Why should we expect disparity to increase when guidelines are repealed or made advisory? This question assumes, first and foremost, that the guidelines worked to minimize disparity in the first place. Several previous essays appearing in this journal have questioned this assumption explicitly (Bushway and Piehl, 2007; Engen, 2009; Kramer, 2009). The prediction also rests on an assumption that, if left to their own devices, judges will discriminate on the basis of race and gender. And yet, this often seems not to be the case. John Kramer (2009) pointed out that local norms have always regulated sentencing, and this has continued even under the guidelines. Although they might disagree with some aspects of them, court communities long ago adapted to and incorporated the guideline model into their local legal culture (Kramer, 2009; Ulmer and Kramer, 1996). Were the guidelines to be repealed entirely it is likely that the philosophy, values, and even the content of those guidelines would be carried forward as a part of local legal culture.

Finally, we know that sentencing guidelines do more than simply provide a context in which social–psychological processes take place. Guidelines structure ways in which substantive sentencing rationales and goals are achieved (e.g., justice, rehabilitation, and community protection), they structure the plea-negotiation process by providing specific rewards and penalties, they facilitate court actors' abilities to manage caseloads, they provide "benchmarks" for determining appropriate punishment, they provide political "cover" for what might be unpopular decisions, they constrain the exercise of discretion, and in many instances, they probably determine the sentence *despite* judicial preferences (Bowen, 2009; Engen and Steen, 2000; Savelsberg, 1992; Ulmer, 1997, 2000; Ulmer and Kramer, 1996). If we are to predict what is likely to happen in the wake of court rulings like *Booker* and *Gall*, or what might happen if legislation once again limits their discretion, then we need theory that goes beyond social–psychological models of judicial decision making. We need theory that explains *how and toward what ends* judges and other court actors use the laws in day-to-day decision making and *how the structure of sentencing laws both facilitates and limits the ability of judges and prosecutors* to achieve their objectives. Reviewing some of the earlier qualitative work by Ulmer, Kramer, and their colleagues might be a good place to start.

References

Blumstein, Alfred, Jacqueline Cohen, Susan Martin, and Michael H. Tonry (eds). 1983. *Research on Sentencing: The Search for Reform*. Vol I. Washington, DC: National Academy Press.

Bowen, Dierdre M. 2009. Calling your bluff: How prosecutors and defense attorneys adapt plea bargaining to increased formalization. *Justice Quarterly*, 26: 2–29.

Bushway, Shawn D. and Anne Piehl. 2001. Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law & Society Review*, 35: 733–764.

Bushway, Shawn D. and Anne Morrison Piehl. 2007. Social science research and the legal threat to presumptive sentencing guidelines. *Criminology & Public Policy*, 6: 461–482.

Engen, Rodney. 2009. Assessing determinate and presumptive sentencing: Making research relevant. *Criminology & Public Policy*, 8: 323–335.

Engen, Rodney L. and Randy Gainey. 2000. Conceptualizing the role of legal and extra-legal factors under sentencing guidelines: Reply to Ulmer. *Criminology*, 38: 1245–1252.

Engen, Rodney L. and Sara Steen. 2000. The power to punish: Discretion and sentencing reform in the war on drugs. *American Journal of Sociology*, 105: 1357–1395.

Frase, Richard S. 2007. The *Apprendi–Blakely* cases: Sentencing reform counter-revolution? *Criminology & Public Policy*, 6: 403–432.

Griffin, Timothy and John Wooldredge. 2006. Sex-based disparities in felony dispositions before versus after sentencing reform in Ohio. *Criminology*, 44: 893–923.

Hofer, Paul J. 2007. *United States v. Booker* as a natural experiment: Using empirical research to inform the federal sentencing policy debate. *Criminology & Public Policy*, 6: 433–460.

Kramer, John H. 2009. Mandatory sentencing guidelines: The framing of justice. *Criminology & Public Policy*, 8: 313–321.

Langan, Patrick. 2001. Effect of choice of measure of the size of a racial disparity. *Journal of Quantitative Criminology*, 17: 273–290.

Merton, Robert K. 1936. The unanticipated consequences of purposive social action. *American Sociological Review*, 1: 894–904.

Savelsberg, Joachim J. 1992. Law that does not fit society: Sentencing guidelines as a neoclassical reaction to the dilemmas of substantivized law. *The American Journal of Sociology*, 97: 1346–1381.

Tonry, Michael. 1995. *Malign Neglect: Race, Crime, and Punishment in America*. New York: Oxford University Press.

Ulmer, Jeffery T. 1997. *Social Worlds of Sentencing: Court Communities Under Sentencing Guidelines.* Albany: State University of New York Press.

Ulmer, Jeffery T. 2000. The rules have changed—so proceed with caution: A comment on Engen and Gainey's method for modeling sentencing outcomes under guidelines. *Criminology*, 38: 1231–1243.

Ulmer, Jeffrey T. 2005. The localized uses of federal sentencing guidelines in four U.S. district courts: Evidence of processual order. *Symbolic Interaction*, 28: 255–279.

Ulmer, Jeffery T, Michael T. Light, and John H. Kramer. 2011. Racial disparity in the wake of the *Booker/Fanfan* decision: An alternative analysis to the USSC's 2010 report. *Criminology & Public Policy*. This issue.

Ulmer, Jeffrey T. and John Kramer. 1996. Court communities under sentencing guidelines: Dilemmas of formal rationality and sentencing disparity. *Criminology* 3: 306–332.

U.S. Sentencing Commission. 2010. *Demographic Differences in Federal Sentencing Practices: An Update of the Booker Report's Multivariate Regression Analysis*. Washington, DC: Author.

Wellford, Charles F. 2007. Sentencing research for sentencing reform. *Criminology & Public Policy*, 6: 399–402.

Wilmot, Keith A. and Cassia Spohn. 2004. Prosecutorial discretion and real-offense sentencing: An analysis of relevant conduct under the Federal Sentencing Guidelines. *Criminal Justice Policy Review*, 15: 324–343.

Wooldredge, John, Timothy Griffin, and Fritz Rauschenberg. 2005. (Un)anticipated effects of sentencing reform on the disparate treatment of defendants. *Law & Society Review*, 39(4): 835–874.

Court Cases Cited

Blakely v. Washington, (02-1632) 542 U.S. 296 (2004) 111 Wash. App. 851, 47 P.3d 149, reversed and remanded.

Gall v. United States, 552 U.S. 38 (2007).

United States v. Booker, 543 U.S. 220 (2005).

Statute Cited

PROTECT Act of 2003, Pub.L. 108–21, 117 Stat. 650, S. 151, enacted April 30, 2003; amendment Stat. 667.

Rodney Engen is Associate Professor in Sociology and Criminal Justice at the University of Arkansas. His research examines racial and gender disparities in sentencing, the effects of prosecutorial discretion under sentencing guidelines, and the effects of state sentencing policies on imprisonment. His research has appeared in journals including the *American Journal of Sociology*, *Criminology*, *Justice Quarterly*, and *Social Problems*.

POLICY ESSAY

RACIAL DISPARITY IN WAKE OF THE *BOOKER*/*FANFAN* DECISION

Judicial discretion in federal sentencing

An intersection of policy priorities and law

Celesta A. Albonetti
University of Iowa

Questions about who has authority to make which decisions with how much discretion has occupied the attention of organizational sociologists for years. As applied to the criminal legal system, the focus is on judicial and prosecutorial discretion. After decades of allowing judges extensive discretion within broadly defined statutory limits, coupled with unreviewable sentence outcomes in both federal and state courts, critics in the early 1970s and 1980s argued for limitations on judicial discretion (Frankel, 1972). Critics argued that judicial discretion produced uncertainty and disparity in sentence severity.

Some social scientists found that that federal sentence disparity prior to reforms of the 1980s was linked to extralegal variables such as the defendant's race/ethnicity, gender, and socioeconomic status (Albonetti, 1998, 1999; Hagan, Nagel, and Albonetti, 1982; Nagel and Hagan, 1982; Peterson and Hagan, 1984; Weisburd, Wheeler, Waring, and Bode, 1991; Wheeler, Weisburd, and Bode, 1982). Other researchers found no significant relationship between defendant's socioeconomic status and sentence severity (Benson and Walker, 1988).

Federal Reform Measures

After years of policy debate, in the mid-1980s, the U.S. Congress enacted laws that virtually transformed sentencing practices. Policy priorities were aimed at severely limiting judicial discretion in an attempt to eliminate unwarranted sentence disparity. How much discretion should judges' exercise at sentencing? Should limits be placed on that discretion? What reform mechanisms could be instituted that would limit judicial discretion effectively? Should judicial discretion be formally overseen? Who is to perform this oversight?

Direct correspondence to Celesta A. Albonetti, Department of Sociology, University of Iowa, W140 Seashore Hall West, Iowa City, IA 52242-1401 (e-mail: celesta-albonetti@uiowa.edu).

DOI:10.1111/j.1745-9133.2011.00772.x

Congress began answering some of these questions early on by enacting mandatory minimum drug laws and passing the Sentencing Reform Act of 1984. More specifically, The Anti-Drug Abuse Act of 1986 mandated minimum penalties for offenders who (a) sell drugs to persons under 21 years of age, (b) hire a person under 18 years of age in a drug offense, and (c) possess a firearm.[1] The Omnibus Anti-Drug Act of 1988 mandated 5-year minimum sentences for simple possession of more than 5 g of crack cocaine. The Act requires a minimum 20-year imprisonment for offenders convicted of involvement in a drug enterprise. It also applied the mandatory minimum penalties for substance distribution and importation/exportation conspiracies to commit these crimes. These laws expressed the "get tough with crime" polices of both houses of Congress and President Ronald Reagan. In 1984, Congress substantial amended the 1968 Gun Control Act[2] as part of Comprehensive Crime Control Act of 1984 by changing the previously wide range of sentences that could be imposed for use of a firearm in the commission of a felony to mandatory 5-year sentences for a first offense and 10-year for subsequent violations (18 U.S.C. § 924(c) (Supp. III), 1985). Two years later, Congress amended the original Act by providing mandatory longer sentences for the use of a machine gun or a firearm with a silencer or some form of muffler (18 U.S.C. § 924(c) (Supp. V), 1987). The change provided for an additional 10-year sentence for first offenders and an additional 20-year sentence for repeat offenders. These statutory changes from indeterminate sentencing for drug and firearm offenses to mandatory minimum penalties and sentence enhancements were aimed not only at deterring criminal behavior but also at reducing judicial sentencing discretion. With these changes, Congress translated policy into sentencing law and intruded into judicial autonomy by limiting judicial discretion at a point in the criminal justice system where judges have historically retained substantial control within an indeterminate sentencing scheme.

By far, the greatest intrusion into federal judicial discretion occurred with the enactment of the Sentencing Reform Act of 1984 and the subsequent implementation of the Federal Sentencing Guidelines in 1987 (hereafter Guidelines). The Guidelines codified congressional and administrative policies to reduce sentence disparity by all but eliminating judicial discretion. The Guidelines replaced indeterminate sentencing within statutorily defined limits with presumptive structured directives.

Current Study and Policy Implications

The current study contributes to our knowledge of federal sentencing by exploring the impact of the 2005 Supreme Court decision in the consolidated cases of *United States v. Booker* and *United States v. Fanfan* (hereafter *Booker*) and sentencing outcomes after the

1. This Act assigned a 5-year mandatory minimum for 100 g or more of a mixture or substance that contains heroin and 500 g or more of a mixture or substance containing cocaine. A 10-year mandatory minimum was assigned for second convictions of these offenses.

2. This Act is part of the Omnibus Crime Control and Safe Streets Act of 1968.

Gall v. United States decision in 2007. Based on the analyses by Ulmer, Light, and Kramer (2011, this issue), the authors "question the notion that *Booker* and *Gall* have caused increases in race/ethnic and gender sentence-length disparity compared with the full range of years when the Guidelines were mandatory" (2011, p. 1108).

Ulmer et al. (2011) conclude that:

> If post-*Booker/Gall* racial/gender length disparity levels were comparable to or lower than levels in previous periods when the Guidelines were also mandatory, this calls into question the notion that the post-*Booker/Gall* eras of advisory Guidelines have produced *uniquely high* levels of racial disparity in sentence lengths. In our view, this calls into question the need for blanket policy remedies that would attempt to curtail overall judicial sentencing discretion in the name of disparity in sentence lengths. [emphasis in original]

I agree with these policy recommendations. There is no need to institute statutory remedies for sentences that do not greatly differ from those imposed under pre-*Booker* mandatory guidelines structure.

One might ask why post-*Booker* racial/gender sentences lengths are not substantially different than those observed during the immediate years pre-*Booker*. Several reasons may explain Ulmer et al.'s (2011) findings. These reasons have policy relevance to what action, if any, should be taken to constrain judicial discretion. First, by the time *Booker* is decided in early 2005, most of the sitting lower court federal sentencing judges have known no other sentencing scheme but the Guidelines. For the relatively few judges that are holdovers from pre-Guidelines days, 18 years of following the step-by-step path to determine sentence outcomes probably has been routinized.

Second, because the Supreme Court in *Booker* replaced the then *de novo* standard of appellate review with a "reasonableness" standard that itself is tied to 18 U.S.C. § 3553(a), lower court judges' sentencing decisions are still made in reference to same policy related directives reflected in mandatory Guidelines. As provided by 18 U.S.C. § 3553(a) (2006), the court is to consider the following in determining the sentence to impose:

(1) the nature and circumstances of the offense and the history and characteristics of the defendant;
(2) the need for the sentence imposed—
  (A) to reflect the seriousness of the offense, to promote respect for the law, and to provide just punishment for the offense;
  (B) to afford adequate deterrence to criminal conduct;
  (C) to protect the public from further crimes of the defendant; and
  (D) to provide the defendant with needed educational or vocational training, medical care, or other correctional treatment in the most effective manner;
(3) the kinds of sentences available;

(4) the kinds of sentence and the sentencing range established for—
   (A) the applicable category of offense committed by the applicable category of the defendant set forth in the guidelines;
(5) any pertinent policy statement—
   (P) issued by the Sentencing Commission;
(6) the need to avoid unwarranted sentence disparities among defendants with similar records who have been found guilty of similar conduct; and
(7) the need to provide restitution to any victims of the offense (18 U.S.C. § 3553 (a)).

By tying post-*Booker* appellate review standard of "reasonableness" to the policy-related considerations found in 18 U.S.C. 3553(a), the Supreme Court constrained lower judges to the same policies that are the cornerstone of the mandatory federal sentencing guidelines enacted in 1987. In effect, the Supreme Court in *Booker* moderated the amount of sentence disparity and continued to limit the extent to which lower court judges could stray from policy related statutory constraints that were at the heart of the mandatory federal sentencing guidelines.

Third, the statutory mandatory minimum penalties of the 1980s and thereafter were let untouched by the *Booker or Gall* decisions. Lower court judges are still required to impose mandatory minimum sentences for drug and gun offenses. The statutory mandatory has always trumped the federal sentencing guidelines. The only exceptions to imposing these mandatory minimum penalties is that the defendant qualifies for the "safety valve" provision (18 U.S.C. §§ 3553 (1)-(5), (1994); U.S. Sentencing Guidelines Manual § 5C1.2, 1995) in drug cases or the government files a substantial assistance departure (18 U.S.C. § 3553(e) and 28 U.S.C. § 994(n); U.S. Sentencing Guidelines Manual §5K1.1, 1987). Absent either of these exceptions, lower court judges are required to impose the mandatory minimum sentences post-*Booker.* For the preceding three reasons, it is not surprising that post-*Booker* race/ethnicity and gender sentence length disparities are similar to those found during pre-*Booker* days. I agree with the authors that there is no need to implement policy remedies—the Supreme Court and Congress already put into place statutes that maintain virtually the same constrains on judicial discretion that existed pre-*Booker*.

References

Albonetti, Celesta A. 1998. Direct and indirect effects of case complexity, guilty pleadate, and offender characteristics on sentencing for offenders convicted of a white-collar offense prior to sentencing guidelines. *Journal of Quantitative Criminology*, 14: 353–378.

Albonetti, Celesta A. 1999. The avoidance of punishment: A legal-bureaucratic model of suspended sentences in federal white-collar cases prior to the federal sentencing guidelines. *Social Forces*, 78: 303–329.

Benson, Michael L. and Esteban Walker. 1988. Sentencing the white-collar offender. *American Sociological Review*, 53: 294–302.

Frankel, Marvin E. 1972. *Criminal Sentences: Law Without Order*. New York: Hill and Wang.

Hagan, John, Ilene H. Nagel (Bernstein), and Celesta A. Albonetti. 1980. The differential sentencing of white-collar offenders in ten federal district courts. *American Sociological Review*, 45: 802–820.

Nagel, Ilene H. and John L. Hagan. 1982. The sentencing of white-collar criminals in federal courts: An socio-legal exploration of disparity. *Michigan Law Review*, 80: 1427–1465.

Peterson, Ruth D. and John Hagan. 1984. Changing conceptions of race: Towards an account of anomalous findings of sentencing research. *American Sociological Review*, 49: 56–70.

Stith, Kate and Jose A. Cabranes. 1998. *Fear of Judging: Sentencing Guidelines in the Federal Courts*. Chicago, IL: The University of Chicago Press.

Ulmer, Jeffery T, Michael T. Light, and John H. Kramer. 2011. Racial disparity in the wake of the *Booker/Fanfan* decision: An alternative analysis to the USSC's 2010 report. *Criminology & Public Policy*. This issue.

Weisburd, David, Stanton Wheeler, Elin Waring, and Nancy Bode. 1991. *Crimes of the Middle Classes: White-Collar Offenders in the Federal Courts*. New Haven, CT: Yale University Press.

Wheeler, Stanton, David Weisburd, and Nancy Bode. 1982. Sentencing the white-collar offender: Rhetoric and reality. *American Sociological Review*, 47: 641–659.

Court Cases Cited

Gall v. United States, 552 U.S. 38 (2007).

United States v. Booker, 543 U.S. 220 (2005).

Statutes Cited

Anti-Drug Abuse Act of 1986, Pub.L. No.99–570, § 1105(a), 100 Stat.3207–11 (1986).

Comprehensive Crime Control Act of 1984, S. Rep. No. 98–225, at 1312 (1983); H.R. Rep. No. 98–1017, at 55–56 (1984).

Omnibus Anti-Drug Act of 1988, Pub.L. No. 100–690, § 6371, 102 Stat. 4181, 4370 (1988).

Celesta A. Albonetti's research focuses on judicial and prosecutorial discretion in the criminal legal sentencing. Her work develops the uncertainty avoidance/causal attribution theory of judicial and prosecutorial decision making and the legal bureaucratic model of criminal adjudication. Her research explores linkages between extra-legal defendant characteristics, sources of discretion, and legally relevant case-level variables. Recently, her research examines the federal circuit and district differences in federal sentencing linked to differences in interpretation of sentencing jurisprudence. Her research has appeared in *American Sociological Review*, *Social Forces*, *American Journal of Sociology*, *Law & Society Review*, *Criminology*, *Journal of Quantitative Criminology* and *Social Problems*.