

Florida State University Libraries

2007

Towards Rational and Evidence-Based Crime Policy

Daniel P. Mears



PRINT VERSION CITATION: Mears, Daniel P. 2007. "Towards Rational and Evidence-Based Crime Policy." *Journal of Criminal Justice* 35(6):667-682.

PRE-PRINT VERSION

Towards Rational and Evidence-Based Crime Policy*

Daniel P. Mears
Florida State University

RUNNING TITLE: Evidence-Based Crime Policy

* Direct correspondence to Daniel P. Mears, Ph.D., Associate Professor, Florida State University, College of Criminology and Criminal Justice, 634 West Call Street, Tallahassee, FL 32306, phone (850-644-7376), fax (850-644-9614), e-mail (dmears@fsu.edu).

Towards Rational and Evidence-Based Crime Policy

ABSTRACT

Rational crime policy constitutes a basic goal for society. If, however, evidence-based, cost-efficient crime prevention is the standard, there is little indication that current policies—including programs, laws, and court decisions—are rational. To support that assessment, this article uses an evaluation research perspective to highlight five prominent problems with extant crime policies: (1) a lack of empirical assessment of the need for them; (2) a range of design issues, including gaps between crime theory and policy, and, most notably, the pursuit of silver bullet solutions; (3) a range of implementation issues, including disjunctures between ideal and actual practice; (4) the lack of rigorous impact evaluations and the sometimes misplaced emphasis on them; and (5) a scarcity of cost-efficiency analyses for guiding investment decisions. It then discusses the implications of these problems and suggests steps that can be taken to place crime policy on a more evidence-based foundation.

Key words = crime policy evaluation prevention effectiveness

Introduction

Policymakers and the public typically do not possess an accurate understanding of the types or amounts of crime that occur, much less of the sanctions or crime prevention measures that society employs (Roberts & Hough, 2005; Roberts & Stalans, 1998). Regardless, crime occurs, the public generally believes it is worsening, and new policies proliferate. As one scholar has noted, “The fact that the public perceives an increased problem requires [legislatures] to act to eliminate this problem” (Dawson, 1988, p. 949).

This situation leads to perverse outcomes, as when the need for and amount of policy trend in opposite directions. To cite a recent example, despite a consistent decline in property and violent crime (Catalano, 2005, p. 1), there has been a dramatic increase in the population supervised by the correctional system. Nationally, individuals under any form of supervision, including probation and parole, increased 250 percent between 1980 and 2000, and by 2004 included 6,996,500 people (Glaze & Palla, 2005). The increase was greater for the most punitive and costly of sanctions, incarceration—state prison populations increased 309 percent between 1980 and 2000, and by the end of 2004 held 1,316,301 inmates, with federal prisons housing another 180,328 (Bureau of Justice Statistics, 2000; Harrison & Beck, 2005). The decline in crime trends may be explained only in part by the putative crime-reducing effects of incarceration (Blumstein & Wallman, 2000), and the question remains as to whether the amount of incarceration was needed or sufficiently effective to offset its costs.

The question of whether various crime policies are needed, effective, or efficient—whether, more generally, they rest on evidence—has assumed particular importance given the large-scale accumulation of such policies over the past three decades (Farabee, 2005; Lipsey, Adams, Gottfredson, Pepper, & Weisburd, 2005; Petersilia, 1991; Ruth & Reitz, 2003; Sherman, Farrington, Welsh, & MacKenzie, 2002; Stolz, 2002; Wilson & Petersilia, 2002). Incarceration perhaps represents the most prominent general policy—states enacted a plethora of sentencing initiatives, including three-strikes laws, sentencing guidelines, and statutes that create enhanced and diverse penalties for specific offenses or types of offenders. At the same time, specialized

community, drug, mental health, drug, and teen courts proliferated. Boot camps became a prominent staple of correctional systems. Numerous gun and drug laws were enacted to reduce supply and demand, while also toughening sanctions for gun- or drug-related crimes. Today, community policing can be found in almost every part of the country. At the same time, countless programs now exist that emphasize specific factors—family functioning, parenting, peer relationships, self-esteem, vocational and life skills, education, religiosity, and, more generally, criminogenic attitudes and behaviors—in the belief that doing so can reduce crime.

The amount of and variation in crime policy within and across states and jurisdictions is difficult to overstate. The intensity of effort alone—occurring during a time when government accountability, and a reliance on evidence-based practices, has increasingly been prioritized (Campbell, 2003; Eskridge, 2005; Lipsey et al., 2005; Riveland, 1999; Sherman, Gottfredson, MacKenzie, Eck, Reuter, & Bushway, 1997)—suggests a proactive, studied attempt to prevent and reduce crime. The question, however, remains: are current crime policies, including programs, laws, and court decisions, rational in the sense of being evidence-based and cost-efficient? That is, from the perspective of evaluation research (Rossi, Lipsey, & Freeman, 2004, p. 80): (1) Are the policies *needed*? (2) Are they premised on sound *theory and design*? (3) Are they fully and appropriately *implemented*? (4) Do the policies have an *impact* on relevant outcomes, especially crime? (5) Are they *cost-effective or cost-beneficial*?

This article addresses these questions and concludes that, in general, current crime policies are *not* in any obvious way rational from an evaluation research perspective. Specifically, it is not clear that they have been: needed, built around a solid theoretical foundation, well-implemented, had an impact, or produced an impact sufficient to justify the costs. Given this assessment, the article discusses the implications of the problem and explores several steps that might be taken to place crime policy on a more rational, evidence-based foundation.

Five evaluation research gaps prevalent in current crime policies

Current crime policies have been criticized along many dimensions. Scholars frequently point, for example, to the seemingly disproportionate influence of political expediency, ideology,

and special interests in the forging of policy (Beckett, 1997; Eskridge, 2005; Garland, 2001; Robinson, 2003; Stolz, 2002). They also emphasize barriers to policy evaluation research, such as the: lack of prestige that academic programs afford applied policy research; challenge of translating complicated research findings into language accessible to lay people, including policymakers; and limited funding for large-scale policy evaluations (Blumstein, 1997; Cullen, 2005; Petersilia, 1991; Rossi, 1980; Stolz, 2002). Still others point to ways in which the policymaking process itself, which frequently does not encourage or allow for careful deliberation and review of research, often undermines the creation of effective crime policy (Marion & Oliver, 2006; Welsh & Harris, 2004). Most recently, scholars have called for rigorous impact evaluations of crime policy effectiveness (Lipsey et al., 2005; Sherman, 2003).

These discussions do not, however, systematically critique crime policy through the perspective of each of the five major types of evaluation research identified by Rossi et al. (2004)—needs, theory, implementation (or process), impact, and cost-efficiency. Certainly, considerable attention has been given to impact evaluations (e.g., Lipsey et al., 2005; Sherman et al., 1997), due in no small part to increased calls for governmental accountability and evidence-based practice (Campbell, 2003; Riveland, 1999). Process evaluations also feature prominently in much crime policy research. The others typically receive far less attention, however, and rarely are all five mentioned in discussions of a particular policy. Each type of evaluation is nonetheless critical to ensuring that crime policies rest on a solid empirical foundation and are cost-efficient. To support that assessment and to highlight the importance of all five types of evaluations to rational crime policy, the following discussion examines each of the types. In so doing, it illustrates their relevance to placing crime policy on an evidence-based foundation and identifies the evaluation research gaps that plague federal and state crime policies.

1. Lack of assessment of need for specific policies

The question arises, even if policymaker and public perceptions about crime were accurate, what threshold of this problem merits a response beyond that which justice or retribution require (von Hirsch & Ashworth, 1992)? *That is, what is the need for a crime prevention response?*

There is no simple answer. As Rossi et al. (2004, p. 108) have emphasized, “It is much easier to establish that a problem exists than to develop valid estimates of its density and distribution.” Consider that in 2004, the United States experienced 21.4 violent crimes (excluding homicide) and 161 property crimes per every 1,000 households, and over 16,000 murders (Catalano, 2005, p. 2). Some foothold for assessing the relevance of these numbers comes from inspecting trends. The rates, for example, are markedly lower than in past years. Even with that context, it is unclear whether current crime levels warrant a response, much less a change in response. Indeed, unless one has a foundation for determining whether prior efforts were “sufficient,” there is no basis for knowing whether to increase or decrease them. Assume, for example, that a given state is spending more than it “needs” and then crime rates increase. Arguably, no additional funding should be invested since the state already is over-investing in crime-fighting efforts.

Many observers might argue that funding for crime policies is never adequate. That argument, however, assumes a definition of need. What, for example, is the proportion of crime that merits a response and how much of it can be addressed with available funding? Answers to either question require an assessment of the magnitude of the crime problem. Few jurisdictions or states, however, compile independent self-report offending or victimization data on crime, and so have little idea how much crime occurs. (Law enforcement data invariably capture only a part of all crime, and so are of questionable use in assessing the scope of or trends in a crime problem since the data may reflect law enforcement practices as much as they do crime—O’Brien, 1995.)

The question remains: what in fact is the “right” amount to spend on crime policies? Consider that between 1982 and 2001, the United States increased its justice system spending by 165 percent (in constant dollars), with spending, as of 2001, approximating \$167 billion (Bauer & Owens, 2004, p. 1). Should this money have been spent to prevent crime? Was it needed?

To understand the importance and challenge of defining “need,” it might help to consider prison forecasting, if only because prisons constitute one of the largest parts of any corrections budget. Forecasting the need for prison bed space holds obvious importance—incarceration is costly. Society thus has a vested interest in avoiding “needless” amounts of it. Although states

vary in their approach to forecasting prison bed space needs, from statistical times series analyses to microsimulation efforts (Butts & Adams, 2001; Sabol, 1999), most approaches attempt to incorporate information about the direction of crime in coming years. Such efforts, like any forecast attempts, entail error, especially when predictions are being made beyond a few years (Penner, 2001). Recall, for example, the dramatic predictions made as late as the mid-1990s of a juvenile violent crime epidemic that would emerge in ensuing years and the fact that juvenile offending instead rapidly declined (Fox, 2000; Howell, 2003). Predictions invariably involve assumptions, such as the relative importance of demographic versus social and economic trends, and when these assumptions prove to be incorrect, so, too, are the forecasts.

Almost all bed space forecasts make a highly problematic assumption—namely, that current bed space use is the “right” amount for the level of crime faced in a particular state or jurisdiction. Put differently, current bed space usage is assumed to be appropriate for achieving some amount of crime reduction. Empirically-based models then rely on current bed space use, along with estimates of future crime rates, to predict the need for future bed space. Observe that “need” is never actually defined. To illustrate, imagine a state where 1,000 assaults occurred. Assume that all could be prosecuted and would likely result in conviction, but that only 300 result in an arrest. Of course, from a purely legalistic perspective, virtually all of the assaults warrant punishment. Only 300 are arrested, however; so, from the outset one faces the problem that 700 of the assaulters (assuming no repeat offenders) will go unpunished. In virtually every state, crimes such as the 700 assaults occur that never come to the attention of law enforcement. The question, then, is, when is that amount—this “dark figure of unpunished crime”—“too much”? Jurisdictions do not typically answer the question, thus the “need” for legal punishment is unknown. Even for the 300 arrests, states rarely define need, save that the most “severe,” however operationalized, typically should be prioritized by prosecutors and the courts. Not all 300 will be given equal attention; indeed, some will be dismissed.

If one turns from purely legalistic to crime prevention definitions of need, questions arise about the amount of punishment needed to achieve some specific quantity of reduced crime in society at large. How many of the 300 arrested assaulters must be punished? What is, on the

one hand, the amount of crime that “needs” to be reduced, and, on the other hand, a reasonable amount of crime reduction to expect? Can a substantial reduction be had by punishing only a small percentage of the offenders, perhaps, for example, the most serious assaulters? Or is it necessary to sanction all of them? If so, by how much, and what about the 700 assaulters who went untouched? Not least, how much of a crime prevention effect will arise through specific and general deterrence mechanisms? That distinction is important because for some crimes, such as homicide, one can anticipate that certain sanctions, like the death penalty, will produce little by way of a general deterrent effect (Donohue & Wolfers, 2006). For other crimes, such as assault, it may be that a combination of specific and general deterrence collectively gives rise to an appreciable crime reduction. (The second issue examined in this article concerns policy theory and design; technically, a discussion of deterrence belongs under that heading. Observe here, however, that the definition of the “need” for a response to crime is intrinsically tied to theory. If, for example, society did not anticipate some reduction in crime from punishment, it would presumably define the “need” for punishment in a different, perhaps purely legal, way.)

Against this backdrop, consider again the question of bed space “need.” How many prison beds are “needed” when there are 1,000 assaults? Certainly, one can anticipate that not all assaulters should be incarcerated. The question is, which ones should be? Prosecutors might focus on the most serious cases, those in which substantial injury was involved, but that suggests a criterion of need—assaults in which only major injuries were sustained. Legislators have enacted laws that endorse greater punishment for weapons-related offenses, so perhaps “need” could be defined as assaults in which a weapon was used. In both cases, to establish the need for bed space, one first must identify how many assaults occurred that met the definition of “serious.” Few jurisdictions or states in fact provide such operational definitions of need, or base their implicit or explicit definitions on the amount of crime reduction that can be expected from using specific quantities of incarceration for particular types of offenders.

This issue pervades criminal justice policy. Consider several prominent examples. Thirty years ago, no states had super-maximum security prisons (“supermaxes”), and now over forty have them (Mears, 2005); yet there is no evidence that states empirically assessed the need for

these high-cost facilities (Kurki & Morris, 2001). One common justification is that they serve to house the “worst of the worst” inmates (Toch, 2003). States typically do not, however, provide clear, measurable criteria for identifying such inmates; consequently, they cannot say how many supermax beds they need or, in turn, whether existing beds are too few, adequate, or too much.

A second example: private prisons are now common—roughly 6 percent of state prison inmates are held in privately operated facilities (Bales, Bedard, Quinn, Ensley, & Holley, 2005, p. 58)—but evidence for their need is scant. Advocates justify privatization on the grounds that correctional agencies spend more than needed to incarcerate offenders. By contrast, private companies may be able to incarcerate offenders at less cost while achieving comparable levels and quality of supervision and services. Few states can quantify, though, how efficient they are in providing, say, X units of supervision and services or in producing a certain recidivism rate. Not surprisingly, then, they cannot assess the extent of need for privatization. Of course, private agencies can enter into contracts and promise to provide “business as usual” supervision and services at less cost. Whether the promise can be upheld, or whether the difference in costs is sufficient to constitute a need for privatization, is, however, an empirical question that remains largely unanswered (Gaes, 2005; Lanza-Kaduce, Parker, & Thomas, 1999; Pratt & Maahs, 1999; Thomas, 2005).

Similar problems arise when one examines other popular crime policies. Sex offender laws, for example, generally assume that recidivism is greater among sex offenders, that such offenders are less amenable to treatment, and, more generally, that widely publicized cases are evidence of a pervasive problem. Each assumption is incorrect (Sample & Bray, 2003). As a result, states have little credible foundation on which to assert a “need” for tougher sanctioning of sex offenders. Similarly, drug sentencing and other “get tough” laws aimed at stiffening penalties for drug-related offending have proliferated. Here, again, little attention has been given to establishing the need for such laws. Certainly, drugs are implicated in much crime, but that is different from defining and quantifying the need for a law. In addition, the causal relationship between drugs and crime remains in question (White & Gorman, 2000). To the extent it is not causal, one might well argue that a need exists, but that it is for treatment, not punishment

(Nadelmann, 2004). Finally, in the juvenile justice system, almost all states have recently embraced a wide array of laws enabling youth to be transferred to the criminal justice system for processing and sanctioning (Feld, 2004). Again, empirical evidence in support of the need for such laws is lacking. Few states, for example, have systematically evaluated how much punishment youth receive in their juvenile justice systems and how these youth likely would be punished in the adult system. An empirical analysis might well reveal that youth would receive less severe sanctions. Were that true, the “need” for transfer laws would be eliminated; notably, studies suggest that in many cases youth indeed receive less severe sanctions in adult court (Fagan & Zimring, 2000).

Such examples, while illustrative, appear to be representative of a general failure of federal, state, and local jurisdictions to identify the need for their policies. Blomberg and Waldo (2002, p. 341) have argued as much, noting that “juvenile justice policy has rarely been guided or significantly influenced by research,” an assessment equally relevant to criminal justice. The consequences of this failure are at once simple and critical—resources are likely to be expended on problems that may not warrant attention and other problems may go largely unaddressed.

2. Design issues: Crime theory and the pursuit of silver bullet “solutions”

The limited use of crime theory in developing crime policies

Effective policies typically require solid theoretical foundations—plans that describe how they are “supposed to work”—without which failure is likely (Rossi et al., 2004, p. 55). Surprisingly, however, many crime policies draw loosely if at all on mainstream criminological theories. That may result in no small part from the reluctance of scholars to tease out the implications of their theories for policies (Barlow, 1995). Or it may simply result from a lack of policymaker or practitioner familiarity with crime theory. Whatever the reason, programs that lack a solid theoretical foundation are more likely to fail, and so a prominent problem with many current crime policies is that they are unlikely to succeed because of poor conceptualization.

Recent meta-analyses reflect that assessment, showing that programs targeting criminogenic factors are more effective in reducing recidivism than those that do not (Cullen & Gendreau,

2000; Howell, 2003; Lipsey, 1999). The analyses are heartening in demonstrating that, contrary to the conclusion of the Martinson (1974) report, effective interventions exist. At the same time, they are disheartening in revealing that many policies fail to draw on established crime theories.

Even so, almost all policies are grounded in theory, whether articulated or not. The problem is that the theories may not be well-supported, or may not be easily put into practice in a particular context. Consider “scared straight” prison visitation programs and boot camps, both of which have been enormously popular. Visitation programs are premised on the idea that would-be youthful offenders who see what prison is like will be deterred from future offending—put differently, they will be “scared straight.” Boot camps operate under a similar logic. Although such programs have a certain intuitive appeal, there is little research to suggest that the logic in fact is sound. Certainly, a large body of work has focused on deterrence, but its implications for visitation programs are unclear. Among other things, visitation programs and boot camps typically emphasize the severity, rather than the certainty and swiftness, of punishment, thus neglecting two other components of deterrence theory. They also do little to address other factors known to cause crime. Notably, rigorous evaluations have found that these programs either have no effect or are harmful (Akers & Sellers, 2004; Farabee, 2005; Petrosino, Turpin-Petrosino, & Buehler, 2003).

The problem in these cases is not simply a failure to draw on theory, it is a failure to clearly show that the theory can be implemented in a manner that would logically contribute to a desired outcome. Consider, again, supermax prisons. The theory underlying them assumes that placing violent and disruptive inmates in one place where they experience twenty-three-hour-per-day single-cell confinement with few services produces a deterrent effect among general population inmates. It is, however, “unlikely that supermax facilities serve as a deterrent because . . . placement in these facilities is relatively rare and often based on administrative decisions using risk factors over which inmates have little control” (Pizarro & Stenius, 2004, p. 258).

Similar problems arise upon closer inspection of many popular crime policies. Drug sentencing laws, for example, typically rest on dubious logic. To illustrate, the laws often target drug dealers. As Kleiman (2004, pp. 436-437) has argued, however, “the logic of replacement—

detering or incapacitating one dealer opens up a market niche for another dealer—suggests that deterrence and incapacitation have limited capacity to shrink the drug markets.” That ultimately is an empirical issue, yet there is little evidence that “incarceration levels are important determinants of drug prices,” and so, “if incarceration cannot substantially raise prices or limit their physical availability, then it cannot reduce drug consumption, which means that it cannot reduce the side-effects of drug consumption in terms of . . . crimes” (p. 437). An equally problematic issue is that the causal role of drugs in offending needs to be elaborated, and the possibility remains that drug use and criminal behavior result from similar underlying factors (White & Gorman, 2000).

Faith-based criminal justice programs increasingly are popular (O’Connor & Pallone, 2002), but also lack a strong theoretical foundation (Mears, Roman, Wolff, & Buck, 2006). One design issue is that no consistent definition of “faith-based” exists, and so a diversity of programs come under the “faith-based” umbrella even when they share little to no characteristics (Hodge & Pittman, 2003; Stern, 2006). Such programs include prayer groups, personal development and parenting classes, peer mentors, revivals, life skills Bible study, anger management, and many others (Corrections Compendium, 2003; O’Connor, 2004, 2005). In some cases, faith or religion are not even the focus of the program. For instance, a study of prison chaplains found that 40 percent “did not select religion as the best method of treatment” and instead felt that “secular methods are better suited to bringing about inmate change” (Sundt, Dammer, & Cullen, 2002, p. 72). As importantly, criminological research provides little foundation for anticipating a strong “faith” effect (Mears et al., 2006). Most studies point to a nominal to modest correlation between religiosity and offending, and few have examined how *changes* in religiosity are associated with changes in offending (Clear & Sumter, 2002; Regnerus, 2003).

Recent sex offender laws also illustrate the problems of flawed program design. Many of these laws call for the creation of sex offender registries that in part are to be used by victims or the public to track where sex offenders reside. Citing research that suggests that the vast majority of women who are raped know the offender, Wright (2003, p. 101) has, however, aptly asked, “How useful is the information contained in sex offender registries where the victim

knows and has an ongoing relationship with the perpetrator?” Perhaps the public may benefit, but little theory or evidence suggest that the location of sex offenders will influence individuals’ decisions about where to reside or that they will know about or use the registries. Even if they do, what can be done with the information that would have an appreciable effect on reducing crime rates?

Many other examples exist, and ultimately an empirical study would be needed to assess what percentage of all crime policies are guided by a solid theoretical logic consistent with extant research. Nonetheless, the examples here are of prominent policies. Although many examples of effective policies, as well as principles of effective intervention, can be cited (see, e.g., Cullen & Gendreau, 2000; Lipsey, 1999; Welsh & Farrington, 2006), many also exist that are ineffective and rest on questionable theoretical foundations (Eskridge, 2005).

The problem of silver bullet “solutions”

Before proceeding, the emphasis by policymakers on silver bullet crime policies bears brief discussion, if only because of the ubiquity of such “solutions” (Marx, 1995) and because they so frequently rest on flawed theory and logic. Regardless of the factors that contribute to the political penchant for silver bullet “solutions”—or, more generally, the politicization of crime policy and research (see, e.g., Blumstein, 1997; Marion, 1997; Miller, 1973; Stolz, 2002)—it is important to be clear about exactly why such solutions are ineffective and can be harmful.

A silver bullet solution is premised on four critical assumptions. First, a particular problem, such as drug use, causes criminal behavior. Second, the problem is widespread, so that any efforts to focus on it will necessarily have a large-scale impact. Third, effective solutions to the problem exist. Fourth, these solutions can appreciably prevent or reduce crime.

These are high bars to meet and, indeed, not many putatively criminogenic factors—including those commonly cited by policymakers, such as drugs (White & Gorman, 2000), mental illness (Hodgins, 1993), family dysfunction (Farrington, 2002), and immigration (Mears, 2001)—meet them. For example, few criminological studies definitively show that a particular factor causes crime; rather, they establish correlations between them. In many instances, the

correlations, as with mental illness and immigration, are not especially strong or are non-existent. Although some criminogenic factors are widespread, frequently the prevalence is overstated, as when “drug problems” are equated with “drug use” (Butts & Roman, 2004). Not least, effective interventions undoubtedly exist and have been documented, but all-too-often the effectiveness is purely statistical—that is, a study identifies a statistically significant effect of treatment but the magnitude of the effect is minimal (Lipsey, 1998).

Juxtaposed against these challenges lies a more important consideration—namely, research increasingly suggests that multi-faceted policies that target multiple causes of crimes using principles of effective intervention are more likely to reduce crime than those that target fewer crimes and that do not rely on these principles (Cullen & Gendreau, 2000; Howell, 2003; Lipsey, 1999). Accordingly, the best and most recent research reviews suggest that silver bullet solutions can not substantially reduce crime.

3. Implementation issues: Disjunctures between theory and practice

To be effective, a policy must be implemented in the way it was intended to be carried out—that is, the specific activities and services envisioned for the policy must actually occur (Rossi et al., 2004, p. 170). That well-established axiom is unproblematic for areas in which implementation is simple and fidelity to a policy’s design can be assumed. Few crime policies, however, enjoy that status. Instead, they typically entail relatively complex designs that require the coordination and cooperation of different stakeholders and a steady supply of resources. The result often is incomplete or poor implementation—that is, a disjunction between theory and practice—which ultimately undermines policy effectiveness. Evaluations of pilot initiatives, where considerable attention is given to implementation, reinforce this point, identifying substantially larger treatment effects than do evaluations of subsequent efforts to implement the policies elsewhere (Lipsey, 1999). Many factors influence this finding, but fidelity to policy design—through funding, staffing, training, etc.—is a central, if not the primary, contributor.

As discussed earlier, legislatures frequently feel pressured to enact new laws to demonstrate to the public their responsiveness to crime (Dawson, 1988). Similarly, state and local agencies

feel pressured to demonstrate that they are at the cutting edge, that they are pioneers, and that they are proactively addressing crime, and so create or adopt new policies. In the meantime, most of the prior policies remain. The situation results in an ever-expanding set of crime-fighting initiatives, each competing with the others for scarce resources and each receiving an ever-diminishing amount of research attention, if any is given at all.

In the meantime, minimal support for research exists. As Farabee (2005, p. 38), summarizing studies on the use of evaluations in criminal justice, has emphasized, “the actual operation of any given correctional program is rarely subjected to . . . evaluation.” Indeed, federal, state, and local funding for criminal justice research is often trivial relative to the range of policies that exist. Consider, for example, the minimal staffing in most justice system research divisions and the fact that the staff frequently must spend most of their time providing highly descriptive empirical descriptions of system operations. Moreover, despite calls for more governmental accountability, there is little indication that research, a source of accountability, is being given greater priority. Recently, for example, Texas eliminated the Criminal Justice Policy Council, which, for over twenty years, served as an independent crime policy evaluation agency (Perry, 2003).

Organizationally, then, few states or jurisdictions have the capacity to monitor the quality of policy implementation. Instead, implementation (also termed “process”) evaluations occur only for select policies that have received particular attention, whether from policymakers, funders, or the public. At best, a few “success stories” can be touted, while other policies, as well as the range of everyday activities comprising the “doing” of criminal justice (Logan, 1993), go unexamined. If one could trust that other policies were implemented well, that would not be a problem. As a large evaluation literature attests, however, implementation issues pervade criminal justice policy. Indeed, it is a truism to say that the “law on the books” rarely resembles the “law in practice” (Feld, 1999). (Implementation issues associated with crime policy *evaluations* also are common—General Accounting Office, 2003; Lipsey et al., 2005.)

Numerous examples illustrate the point. There is the long-standing issue of “justice by geography,” in which specific laws are applied with different intensity in particular counties or

regions (Feld, 1999). That raises issues of fairness, but just as importantly highlights that a law's effectiveness cannot be assumed given inconsistent implementation. How effective, at least on a state level, can a law aimed at harshly punishing sex offenders be if some counties ignore the law, others only partially use it, and still others only selectively use it for particular sex crimes (Sample & Bray, 2003; Wright, 2003)? Replace "sex" with "juvenile," "first-time felony," "drug," or any other type of crime, and the same question arises.

Just as common a problem is implementation of a policy in a situation where it cannot succeed or where contextual factors impede any potential impact. Davis, Smith, and Taylor (2003), for example, have described how a Milwaukee prosecutor's policy of increasing domestic violence case filings, regardless of victim desires, contributed to delays in case processing, reduced convictions, and decreased victim satisfaction. These outcomes stemmed primarily from the fact that staffing increases were not commensurate with increased caseloads.

Self-adjusting systems changes are perhaps just as common. Here, a policy is implemented, but because of adjustments in other parts of the criminal justice system, no change in outcomes occurs. For example, in 1996, Pennsylvania enacted a law that, instead of relying solely on judicial discretion to send some youth to the adult system, required youth age fifteen and older who were charged with certain violent offenses, or who met other criteria, to be processed in adult court. The law targeted populations that previously would be less likely to be transferred to the criminal justice system (e.g., younger offenders, females, youth with few prior offenses). Snyder, Sickmund, and Poe-Yamagata (2000) found that, due to de-certification provisions allowing certain cases to be rejected by prosecutors or to be sent to the juvenile justice system, the number and profile of youth actually sanctioned in adult court did not differ appreciably from what occurred prior to the new policy.

Such examples constitute far from unique or isolated cases. Indeed, one rarely reads evaluations of crime policies, especially of non-pilot initiatives, where less funding and scrutiny typically are involved, and not find substantial implementation problems. This situation reflects precisely what one would anticipate when crime policies burgeon, criminal justice systems grow exponentially, and funding for research is limited or non-existent.

A final point bears emphasis. In recent years, states have emphasized performance monitoring (Logan, 1993; Rossi et al., 2004). This shift may lead to greater attention to quality implementation, especially if the monitoring, a type of implementation evaluation, is required by law. Mandates, even legal ones, are no guarantee that more and better data will be collected. The larger issue, though, consists of whether such monitoring, as practiced, improves performance. Equally important is the relevance of the data being collected and interpretation of the statistics that are produced. Imagine a prosecutor keeping track of the number of cases prosecuted or the percentage resulting in conviction. What do these numbers mean? What if a small number of important cases rather than a large number of trivial cases were prosecuted? What if difficult cases led to lower conviction rates? In theory, one might devise a monitoring system that took such qualifications into account. Doing so, however, requires exceptional and rarely exercised care and attention, drawing on information about past practice, established standards in the field, expert opinion, and other sources that can be used to establish appropriate measures and criteria for establishing good or bad performance (Rossi et al., 2004, p. 75).

4. Limited relevance of extant impact evaluations for crime policy

Rigorous impact evaluations are needed

Whatever the specific activities targeted by a crime policy—enhancing prosecutorial discretion, increasing community policing, implementing a specialized court, etc.—the ultimate goal, with few exceptions, is crime prevention or reduction. Perhaps the most notable exception concerns justice, which may require special procedures and rules that do not necessarily affect crime but nonetheless create or provide for just and fair processing. That exception aside, society typically and reasonably expects policies to reduce crime.

A fundamental and oft-made criticism of much crime policy research is, however, that it lacks sufficient rigor to identify effective policies (Berk, 2005; Farabee, 2005; Lipsey et al., 2005; Petersilia, 1991; Sherman, 2003, 2004; Sherman et al., 2002). Frequently, no control groups are used, or, if they are, the groups differ, making comparisons suspect. In quasi-experimental designs, for example, researchers may identify treatment effects and control for

factors that might bias estimates. If, however, they fail to control for factors, such as the motivation to participate in treatment or levels of community-level collective efficacy that may influence relevant outcomes, the estimated treatment effects will still be biased.

The sum result? Policymakers lack a large body of well-evaluated options—ones shown to be effective across many populations, places, and time—from which to select. More rigorous impact evaluations are needed, especially those that rely on experimental designs and, more generally, that take advantage of naturally occurring experiments or employ quasi-experimental designs that have higher “internal validity” (Farrington, 2003; Sherman, 2003). In these cases, one can trust more that the effects of a policy indeed are due to the policy and not other factors. Internal validity is important because studies have found that weaker research designs tend to be more likely to produce favorable findings of policies, in turn creating the risk that policies build off of treatment “effects” that are artifacts of study design (Weisburd, Lum, & Petrosino, 2001).

Problems with experimental designs

Although experimental designs have substantial strengths, they also have limitations that bear emphasis because of the widespread belief that they constitute the “gold standard” and the corresponding emphasis that has been given to them in recent years (see, e.g., Sherman, 2003). Some limitations arise from how such evaluations are conducted, but some stem from the nature of crime and how social context influences evaluations (Berk, 2005). Certainly, more experimental studies are needed, but their limitations should be recognized, as should the merits of non-experimental designs, which often can be conducted at far less cost (Rossi et al., 2004).

First, impact evaluations frequently are undertaken in the absence of process evaluations. For simple interventions that consist of a single activity (e.g., attending a one-hour class), that may not present a problem. Many criminal justice interventions, however, consist of multiple dimensions, including a variety of supervision, sanction, and treatment options sequenced over many months. Here, should an impact evaluation identify a “treatment” effect, it is unclear—in the absence of more evaluation—what exactly led to the effect. Is it the supervision? Treatment? Some combination of the two? Such questions remain endemic in crime policy

research. For example, despite the widespread prevalence of drug courts, as well as separately conducted process and impact evaluations of them, researchers lack firm evidence about the independent or interactive contributions of supervision and treatment, two pillars of these courts (Gottfredson, Najaka, & Kearley, 2003). The problem of “black box” experimental evaluations, more generally, is that “the evidence from experiments does not cumulate. The end result of a research program based on experiments is just a list of programs that ‘work’ and ‘don’t work,’ but no understanding of why they succeed or fail” (Heckman & Smith, 1995, p. 108).

Second, sample sizes frequently are too small for impact evaluations to identify substantively significant policy effects (Lipsey, 1998). The problem lies in the fact that real effects may exist, but the studies cannot detect them. In a concluding section of articles, the assessment in these cases tends to be “we found no treatment effects” or the treatment “doesn’t work” (Weisburd, Lum, & Yang, 2003). Technically, the description should be that “the study identified no statistically significant treatment effects.” Even this phrasing is misleading because the studies could only have detected effects of a large magnitude. Thus, to be both technically correct and not misleading, the descriptions should be that “the study identified no statistically significant treatment effects and could not have so identified any effect size of less than X percent.” Such phrasing, however, will not likely offset initial impressions of a policy’s ineffectiveness.

The ubiquity of this problem cannot be overstated. Summarizing the results of a meta-analysis of 556 delinquency intervention studies, Rossi et al. (2004, p. 314) found that “many effects of a magnitude that might well represent important program benefits were not found to be statistically significant . . . mainly because of small sample sizes and underutilization of control variables.” Their discussion of studies involving an effect size of .30 is sobering:

An effect size of .30 in standard deviation units for such a measure [recidivism rates of juvenile offenders] corresponds to the difference between an intervention group with a 38 percent recidivism rate in the six months after intervention and a control group with a 50 percent recidivism rate. This represents a 24 percent reduction in the number of juveniles reoffending (.12/.50), which certainly seems like a worthwhile program effect. Despite that, effects of this magnitude were not found to be statistically significant *for more than half the*

instances in which they were reported. Such results reveal more about the failure of evaluators to design adequately powered studies than the failure of the programs to produce meaningful results. (P. 314; emphasis added; see also Weisburd et al., 2003)

The problems attendant to underpowered studies include not just the failure to identify a meaningful effect. They also include the possibility that a policy will be voided (e.g., a program may close or a law may be repealed), that a less effective alternative will be adopted in its place, and, of course, the reality that substantial funds were expended to support the study.

A third problem concerns the generalizability of impact evaluations to other populations, conditions, and places, a concept known as external validity (Farrington, 2003). The best studies are thought to be experimental designs, or “efficacy” studies; these evaluations occur under ideal settings where conditions that might influence the treatment and control groups are strictly controlled. Such designs, however, tell us little about whether a policy will have an effect outside these ideal settings. To make such an assessment, one needs “effectiveness” studies—that is, evaluations in real world settings (Mrazek & Haggerty, 1994, p. 372). Sherman, Schmidt, and Rogan’s (1992) Minneapolis experiment stands as a cautionary tale. The study, involving a randomized design, found that arresting abusers reduced reoffending. Despite the authors’ caveats concerning limitations of the study and the need for more research, many states enacted laws mandating arrest of abusers. Subsequent studies of these laws found that arrest produced a similar effect in some places, but in others did not have any effect or made matters worse, a situation possibly resulting from differences in the “real world” conditions in the different places, including the composition of the targeted population (Schmidt & Sherman, 1996).

In short, the external validity of many efficacy studies is unknown. The same holds true for many effectiveness studies. Suppose, for example, an evaluation of a drug court that operates under less-than-ideal situations nonetheless is found to produce sizable reductions in offending (see, e.g., Gottfredson et al., 2003). That does not mean the court will be effective in other places, especially if the effectiveness was a function of the characteristics of the clients or staff in the particular location. Ultimately, only large-scale, multi-site randomized impact evaluation designs are capable of addressing such issues, yet remain rare in criminological research.

A rarely discussed fourth problem involves substitution bias, which “arises when members of an experimental group gain access to close substitutes for the experimental treatment”—in this case, “control group outcomes no longer correspond to the untreated state” (Heckman & Smith, 1995, p. 105). The issue is the appropriate comparison. As Sherman (2003, p. 15) has noted, “the problem of differing control groups is basic to experimental design. To say that a program increases literacy by 50 percent is to imply a comparison: a 50 percent increase relative to what?” Consider two experimental designs involving drug courts. In one community, there are numerous treatment services available for offenders in the control group and in another there are none. In the first, the comparison is between a group that receives treatment services from the drug court and a group that receives a range of services in the community. In the second, the comparison is between a group that receives treatment services from the drug court and a group that receives no services. Perhaps the evaluation of the first drug court finds no treatment effect but the evaluation of the second does. Can one conclude that drug courts work? It depends on the comparison. In short, as helpful as experimental designs can be, they, too, like evaluations involving weaker designs, suffer from external validity problems.

Bardach (2004) has captured the essence of the above two problems by noting that many replication studies—in which the effectiveness of a particular policy is tested in another place—actually are not replications but extrapolations. That is, the policy essentially is different, due to variations in the environments where the policy is introduced, and so by default one is attempting to extrapolate the idea behind the policy. In Bardach’s (2004, p. 209) phrasing, policymakers attempt to transfer the “basic mechanism”—a causal force of some kind—to a new area. For example, if specific deterrence is the basic mechanism through which mandatory arrest policies produce (if they do) decreased domestic violence recidivism, then the extrapolation to a new place should emphasize specific deterrence, not mandatory arrest. Consider that arrest, as a particular activity, might reduce recidivism among employed people and increase it among the unemployed (the example is not hypothetical—see Schmidt & Sherman, 1996). In this case, in areas with large concentrations of unemployed people, one would want a substitute for arrest that produces a specific deterrent effect among such populations.

Fifth, a greater emphasis on methodological rigor biases studies towards interventions that are more narrow in scope, such as programs that serve a specific at-risk population or that consist of a single-factor intervention. The reason is simple—these interventions lend themselves more easily to randomization and in turn the production of less equivocal results. In contrast, interventions such as community policing, which typically target large populations and areas and consist of multiple strategies to reduce crime, cannot easily be studied through randomization. That situation arises because of the difficulty in identifying appropriate comparison sites, the multi-factor nature of the intervention, and the prohibitive costs of large-scale, multi-site experimental evaluations. Even if an experiment is done, it likely will suffer from the issues described above and have little ability to identify the parts of the “black box” that caused the observed outcomes.

The problem consists of a trade-off in returns—although easier to study, policies that focus on a narrowly circumscribed population or area may not produce the magnitude of return that less effective policies focusing on diverse populations or areas do. The risk, then, lies in creating a body of evidence about effective policies that collectively may not produce nearly the amount of crime reduction that large-scale, community-focused policies could. On the other hand, the latter policies can be extremely costly and may not be effective. In addition, their complexity makes them difficult to replicate or extrapolate. The large body of work on community policing is instructive. Despite the popularity of this crime prevention policy, “sustained evidence of such positive community impacts [as have occurred in Chicago] remains scant” (Greene, 2000, p. 327; see also Rosenbaum, Lurigio, & Davis, 1998; Skogan & Frydl, 2004).

Finally, a related issue, elaborated below, is that of determining whether the costs associated with achieving a specific magnitude of impact constitutes a worthwhile investment relative to other options. Should a city invest \$1 million in a domestic violence abuser counseling program that one can expect has high probability of reducing by 10 percent the recidivism among the 100 participants it will serve? Or should it invest the same money in a domestic violence educational awareness campaign that has a low probability of reducing by 1 percent all domestic violence occurrences in the city? If one restricts attention primarily to the high-quality experimental

designs, the result, on average, will be to privilege the former program. In the world of personal finances, that is akin to investing retirement funds in low-risk, low-yield savings accounts rather than high-risk, high-yield stocks—a risky approach over the long-term (Carlson, 2005).

Impact evaluations often are not possible or needed

Contrary to conventional wisdom, it is unreasonable to expect that virtually every implemented policy be shown to be effective. To conduct rigorous impact evaluations of policies is costly and time-consuming. Virtually no state or local jurisdiction can afford to undertake impact evaluations of more than a handful of their crime policies. Does this fact provide license for policies to be enacted without fear of criticism, since no study will show that they work or not? No—there are many policies that research shows can be effective, and studies can be done at relatively minimal cost to monitor and improve implementation to ensure that policies have the best chance possible to be effective.

Consider an analogy to medicine. Society expects that doctors will be trained in and use treatments that science has shown to be effective with particular diseases (Institute of Medicine, 2001). The expectation in fact differs from practice—Sherman (2003, p. 7) has observed that “in medicine as in government, much of what is done proceeds from theory, conjecture, and untested new ideas,” and, citing Millenson’s (1997) discussion of a 1983 U.S. Office of Technology Assessment study, “that 85 percent of everyday medical treatments had never been scientifically tested.” Nonetheless, one does not typically expect that doctors will reinvent the wheel by showing that their particular use of a treatment is effective. Rather, one expects doctors to implement effective treatments appropriately, and the hope is that the treatments work in particular cases. If they do not, it is understood that treatment is effective “on average,” not necessarily in each and every case. If doctors refused to or did not want to be monitored, one of course might be concerned.

By the same token, one should expect crime policies to have a solid theoretical and empirical, research-based foundation and that assessments be undertaken to ensure that these policies are well-implemented. Just as it is unrealistic, as well as unnecessary, to require impact

evaluations of each and every doctor's implementation of an established treatment, however, it is unrealistic and unnecessary to require impact evaluations of every crime policy.

5. Few cost-efficiency analyses

Consistent with national demands for greater government accountability has been the increased emphasis on efficiency—that is, showing that resources are expended in the most productive manner possible. A basic question confronts those who wish to select from among possible crime policies: what is the relative cost-efficiency of crime policies that aim to prevent a diverse array of crime as well as those that aim to achieve additional benefits (e.g., improved health, employment, and housing outcomes)? Unfortunately, despite increased calls for them, crime policy cost-efficiency analyses remain rare (Cohen, 2000; Dhiri & Brand, 1999; Nagin, 2001; Rossi et al., 2004; Welsh & Farrington, 2000). Those that exist often suffer from methodological problems that undermine their usefulness. In the meantime, crime creates substantial costs for society—in 1993, the last year for which a comprehensive assessment was conducted, the costs were estimated to be \$450 billion (Miller, Cohen, & Wiersema, 1996).

Typically, an efficiency assessment involves cost-effectiveness or cost-benefit analyses (Kee, 1994; Rossi et al., 2004). The former analysis identifies the cost per outcome (e.g., for every dollar spent, one crime was averted) and is useful when comparing policies targeting the same outcomes. The latter identifies, in monetary terms, a policy's costs and benefits (e.g., for every dollar spent, two dollars were saved) and is useful when comparing policies targeting different outcomes (e.g., one aiming to prevent teen pregnancies and another aiming to reduce crime).

Cost-benefit analysis is especially useful because it enables one to make “apples to apples” comparisons. If, for example, one has to choose between a policy that prevents one rape for every dollar spent and another that prevents one robbery for every dollar spent, it is difficult to determine which investment is the most efficient. By converting the outcomes to monetary values, that comparison can more easily be made. Although monetizing outcomes is difficult (and some would argue that it is impossible to assign appropriate monetary values to the cost of such crimes as murders and rapes), society invariably makes such judgments when it funds

specific amounts of money for some crime prevention efforts targeting, say, domestic violence and provides little for other types of crimes (Cohen, 2000, p. 271). The implicit valuation is that the money was better spent preventing a certain number of domestic violence incidents rather than some number of burglaries or white-collar crimes.

Perhaps the most prominent problem with crime policy cost-efficiency analyses—apart from the limited number of them—centers around the fact that, as Welsh and Farrington (2000, p. 307) have emphasized in their review of them, the adequacy of the analyses “depends crucially on the methodological rigor of the underlying evaluation design.” As the prior sections indicate, many evaluations suffer from flaws that cast into doubt the accuracy of their assessments of impact. Indeed, Welsh and Farrington (2000, p. 345) hoped to restrict their review to “programs with high-quality evaluation designs, but the small number of analyses militated against this.” Notably, they found only one community crime prevention study that met their inclusion criteria.

A related problem is that many cost-benefit analyses produce estimates that are not appropriate for comparison. As Welsh and Farrington’s (2000, p. 308) review uncovered:

Researchers have used different methods for calculating costs and benefits (e.g., marginal or average), included different program costs (operating, capital, or both), monetized different program effects (e.g., on crime, employment, education failure, etc.), and aggregated program costs and benefits on different theoretical assumptions.

In addition to this issue are misunderstandings and disagreements about central concepts in cost-benefit analysis. For example, Cohen (2000, p. 271) has observed that few researchers distinguish between social and external costs; in fact, “there is no real agreement on which crime costs are social costs.” (“An external cost is a cost imposed by one person onto another. . . . Social costs are costs that reduce the aggregate well-being of society”—Cohen, 2000, p. 272.) There also are disagreements about the appropriate perspective of analysis. The perspective—for example, that of taxpayers, victims, government agencies—is critical because it dictates which costs and benefits to include. To illustrate, a government agency charged with public health might ignore the crime reduction benefits of drug treatment, while a criminal justice agency might ignore the public health benefits of reduced drug abuse (Cohen, 2000, p. 277).

Such variation is cause for concern. On top of the many caveats and concerns that policymakers must take into account when considering policy impacts, they then confront the challenge of decoding the cost-benefit estimates presented to them. In short, for rational, evidence-based crime policies to be adopted, a foundation—currently lacking—is needed for deciding which ones constitute the best, most efficient investments.

Towards rational evidence-based crime policy

Consideration of the above observations leads to several recommendations. First, each of the five major types of evaluations—needs, theory and design, process or implementation, impact, and cost-efficiency—should be made a central activity of criminal justice systems. More pointedly, needs, theory, and process evaluations should be required of *all* policies. The almost invariable consequence of failing to do so is that substantial resources will be wasted.

Requiring needs, theory, and process evaluations would, of course, entail costs, but the costs need not necessarily be exorbitant. For example, jurisdictions might conduct annual crime surveys at relatively low expense to obtain independent, and generally more valid, assessments of crime and thus the “need” for a given policy or practice. Similarly, state legislatures or local governmental authorities might require criminal justice system agencies to allocate, say, 5 percent of agency or policy funds for research, or any new legislatively-funded crime initiatives might be required to be preceded by needs and theory evaluations and, if implemented, to have implementations be part of the funded activities. Doing so would substantially increase the ability of agencies and policymakers to target the most important needs, to do so using policies that have the greatest likelihood of success, and to monitor implementation to ensure the greatest possible impact. Although such funding would constitute a dramatic increase in typical research allocations, it also has the considerable potential to help avoid adoption of unnecessary, costly, and ineffective policies.

It bears emphasizing that needs and theory evaluations alone have the potential to help criminal justice systems avoid expending resources on questionable investments. To illustrate, consider that a single supermax might cost \$1 billion to build and operate over a thirty-year

period of time and that four-fifths of states now have at least one. Their widespread adoption has occurred despite questions about the need for them and a theoretical foundation premised on questionable assumptions (Mears & Reisig, 2006). Prior to building a supermax, expending, say, \$1 million for a needs evaluation would be trivial compared to the resources saved if it turned out an insufficient “need” for a supermax existed, that their theoretical foundation was weak, or that cheaper alternatives existed.

Similarly, implementation evaluations are critical—especially because impact evaluations frequently are not possible—and should be integrated into policies in a manner similar to the standards of medical practice. That is, policies should be monitored to ensure that they are applied at appropriate times for appropriate populations and in an appropriate manner (Rossi et al., 2004). Systems-level monitoring also is needed to identify redundancies, inefficiencies, and possible improvements (Logan, 1993). Such activities should be ongoing parts of criminal justice systems and policies, not one-time events.

Second, although impact evaluations cannot realistically be conducted on all policies, where possible, they should be pursued, especially for policies that have the greatest likelihood of generating large-scale benefits (Farabee, 2005; Farrington, 2003; Sherman, 2003). As Lipsey et al. (2005, p. 2) have emphasized, “Resources should be directed mainly toward evaluations with the greatest potential for practical and policy significance from expected evaluation results and for which the program circumstances are amenable to productive research.” Efforts should also be made to ensure that researchers evaluate a *range* of policies, including not only those that target individuals but also those that attempt to improve and affect entire communities, justice system operations, or state crime rates. In so doing, researchers must recognize the limitations of experimental designs and that levels of rigor exist (Farrington, 2003; Lipsey et al., 2005; Mrazek & Haggerty, 1994; Rossi et al., 2004; Sherman, 2003; Sherman et al., 2002). As Heckman and Smith (1995) have emphasized, much can be learned from non-experimental research. In addition, impact evaluations should be coupled with process evaluations to identify the features of the policy that produce the outcomes (Rossi et al., 2004).

Third, not least, cost-efficiency analyses should be conducted and should monetize all policy

effects (Welsh & Farrington, 2000, p. 352). Given their costs, such analyses should be pursued, as with impact evaluations, on a selective basis. For policies of any substantial cost, however, break-even analyses should be undertaken to identify precisely how much impact is required for the benefits to at least equal the costs. In many instances, this exercise may reveal that the needed impact well exceeds what is realistic to expect (Lawrence and Mears, 2004).

A final recommendation is that researcher, practitioner, and policymaker communities should create stronger links to one another (Blumstein, 1997; Eskridge, 2005; Marion & Oliver, 2006; Petersilia, 1991; Welsh & Harris, 2004). One strategy for achieving this goal is for states to establish criminal justice policy councils that serve both a research and policy deliberation function. Such councils should be autonomous, to the extent possible, and charged with conducting research and bringing together scholars, practitioners, and policymakers together to discuss crime problems, proposed or existing solutions, and the research that should be undertaken. Although full autonomy is likely impossible, examples of such organizational arrangements exist, including the now disbanded Texas Criminal Justice Policy Council and the U.S. Government Accountability Office. A related approach is to develop, as Blumstein (1997, p. 360) has advocated, “forums that will be more closely allied with the policy process, such as sentencing commissions or Presidential commissions.” Such entities, which such national organizations as the American Society of Criminology might spearhead, have the potential to ensure that research informs the development, implementation, and evaluation of policy. Still another option, one advocated by Cullen (2005, p. 28), is the development of institutes or centers, possibly within universities, “devoted to the dissemination of research knowledge in an accessible form.” Some universities, colleges, schools, and departments may resist such efforts, but such resistance might be tempered through attempt to “define our [criminology’s] research mission to include assistance in the field” (Petersilia, 1991, p. 12).

Ultimately, an effective criminal justice system, including the range of policies aimed at preventing crime, must be ensure that all efforts are systematically integrated and coordinated with one another to increase system effectiveness and efficiency (Marion & Oliver, 2006; Welsh & Harris, 2004). Such a system ideally will be built on the increasingly large body of research

identifying “best” practices, models, and principles (e.g., Cullen & Gendreau, 2000; Howell, 2003; Lipsey, 1999; Mihalic, Irwin, Elliott, Fagan, & Hansen, 2001; Mrazek & Haggerty, 1994; Perry, McDougall, & Farrington, 2006; Welsh & Farrington, 2006). (See also the work emerging from the Crime and Justice Group of the Campbell Collaboration—www.campbellcollaboration.org—which, on an ongoing basis, is reviewing research on the effectiveness of crime and justice programs.) Of course, “politics” will inevitably govern the selection of policies (Garland, 2001), but such influences are not necessarily harmful, especially in a democracy. In situations where political factors may likely lead to unnecessary, ineffective, or inefficient policies, empirical research can serve a potentially tempering influence (Blumstein, 1997). Even in these cases, though, quality research is not a substitute for a deliberate and careful policymaking process, one in which research is but one part (Eskridge, 2005; Marion & Oliver, 2006; Robinson, 2003; Stolz, 2002).

Conclusion

In the United States, crime results in roughly one-half trillion dollars in costs each year (Cohen, 2000) and annual criminal justice system expenditures now exceed \$160 billion (Bauer & Owens, 2004). Juxtaposed against those costs is a paucity of systematic attention to conducting any of the five major kinds of evaluations—needs, theory and design, process or implementation, impact, and cost-efficiency—typically used to justify, design, implement, and assess policies (Rossi et al., 2004). The situation stems in part from the influence of politics, as well as the lack of infrastructure to support evaluations (Cullen, 2005; Eskridge, 2005; Lipsey et al., 2005; Marion & Oliver, 2006; Sherman, 2004). Criminal justice systems, for example, typically have limited research capacity, and few criminology and criminal justice programs formally introduce students to evaluation research. More broadly, criminology as a discipline historically has not always embraced or valued public policy research (Blumstein, 1997; Eskridge, 2005; Petersilia, 1991).

The sum result is that crime policy on the whole lacks a rational foundation in the sense of resting on an evidence-based, evaluation research platform that would help ensure that it is

effective and efficient (Eskridge, 2005). Consider the questions posed at the beginning: are current crime policies needed, based on sound theory and design, implemented well, effective in contributing to a range of outcomes, and cost-effective or cost-beneficial? In each instance, reviews of the literature lead to the same conclusion—the answers are largely unknown.

As disheartening as that assessment may be, some solutions exist. Research, for example, could be accorded a prominent role in policy development and implementation. Institutionalized support for and mandates to conduct each of the types of evaluation for all crime policies, at least those that cost the most, would facilitate this role. Colleges, schools, and departments of criminology and criminal justice could give greater emphasis to teaching and conducting evaluation research. At the same time, stronger links between researchers and policymaking community could be forged—whether through policy councils, commissions, or other organizational arrangements (e.g., crime policy centers and institutes)—to better inform not only debates about policy options but also the design and implementation of specific policies. Collectively, such efforts may begin to institutionalize the development and greater use of evidence-based crime policies and, more generally, contribute to improved policymaking efforts (Marion & Oliver, 2006; Welsh & Harris, 2004).

Acknowledgements

The author thanks Emily Leventhal and the anonymous reviewers for their helpful comments and suggestions.

References

- Akers, R. L., & Sellers, C. (2004). *Criminological theories: Introduction, evaluation, and application* (2nd ed.). Los Angeles, CA: Roxbury.
- Bales, W. D., Bedard, L. E., Quinn, S. T., Ensley, D. T., & Holley, G. P. (2005). Recidivism of public and private state prison inmates in Florida. *Criminology and Public Policy*, 4, 57-82.
- Bardach, E. (2004). Presidential address—the extrapolation problem: How can we learn from the experience of others? *Journal of Policy Analysis and Management*, 23, 205-220.
- Barlow, H. D. (Ed.). (1995). *Crime and public policy: Putting theory to work*. Boulder, CO: Westview Press.
- Bauer, L., & Owens, S. D. (2004). *Justice expenditure and employment in the United States, 2001*. Washington, DC: Bureau of Justice Statistics.
- Beckett, K. (1997). *Making crime pay: Law and order in contemporary American politics*. New York: Oxford University Press.
- Berk, R. A. (2005). Knowing when to fold 'em: An essay on evaluating the impact of Ceasefire, Compstat, and Exile. *Criminology and Public Policy*, 4, 451-466.
- Blomberg, T. G., & Waldo, G. P. (2002). Evaluation research, policy, and politics. *Evaluation Review*, 26, 340-351.
- Blumstein, A. (1997). Interaction of criminological research and public policy. *Journal of Quantitative Criminology*, 12, 349-362.
- Blumstein, A., & Wallman, J. (Eds.). (2000). *The crime drop in America*. New York: Cambridge University Press.
- Bureau of Justice Statistics. (2000). *Correctional populations in the United States, 1997*. Washington, DC: Bureau of Justice Statistics.
- Butts, J. A., & Adams, W. (2001). *Anticipating space needs in juvenile detention and correctional facilities*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.
- Butts, J. A., & Roman, J. (2004). *Juvenile drug courts and teen substance abuse*. Washington,

- DC: Urban Institute Press.
- Campbell, R. (2003). *Dollars and sentences: Legislators' views on prisons, punishment, and the budget crisis*. New York: Vera Institute of Justice.
- Carlson, R. C. (2005). *The new rules of retirement: Strategies for a secure future*. Hoboken, NJ: Wiley.
- Catalano, S. M. (2005). *Criminal victimization, 2004*. Washington, DC: Bureau of Justice Statistics.
- Clear, T. R., & Sumter, M. T. (2002). Prisoners, prison, and religion: Religion and adjustment to prison. In T. P. O'Connor & N. J. Pallone (Eds.), *Religion, the community, and the rehabilitation of criminal offenders* (pp. 127-159). New York: Haworth Press.
- Cohen, M. A. (2000). Measuring the costs and benefits of crime and justice. In D. Duffee (Ed.), *Measurement and analysis of crime and justice* (pp. 263-315). Washington, DC: National Institute of Justice.
- Corrections Compendium. (2003). Faith-based programming. *Corrections Compendium*, 28(8), 8-20.
- Cullen, F. T. (2005). The twelve people who saved rehabilitation: How the science of criminology made a difference. *Criminology*, 43, 1-42.
- Cullen, F. T., & Gendreau, P. (2000). Assessing correctional rehabilitation: Policy, practice, and prospects. In J. Horney (Ed.), *Policies, processes, and decisions of the criminal justice system* (pp. 109-175). Washington, DC: National Institute of Justice.
- Davis, R. C., Smith, B. E., & Taylor, B. (2003). Increasing the proportion of domestic violence arrests that are prosecuted: A natural experiment in Milwaukee. *Criminology and Public Policy*, 2, 263-282.
- Dawson, R. O. (1988). The third justice system: The new juvenile-criminal system of determinate sentencing for the youthful violent offender in Texas. *St. Mary's Law Journal*, 19, 943-1016.
- Dhiri, S., & Brand, S. (1999). *Analysis of costs and benefits: Guidance for evaluators*. London, England: Home Office.

- Donohue, J. J., & Wolfers, J. (2006). Uses and abuses of empirical evidence in the death penalty debate. *Stanford Law Review*, 58, 791-846.
- Eskridge, C. W. (2005). The state of the field of criminology. *Journal of Contemporary Criminal Justice*, 21, 296-308.
- Fagan, J., & Zimring, F. E. (Eds.). (2000). *The changing borders of juvenile justice: Waiver of juveniles to the criminal court*. Chicago: University of Chicago Press.
- Farabee, D. (2005). *Rethinking rehabilitation: Why can't we reform our criminals?* Washington, DC: AEI Press.
- Farrington, D. P. (2002). Families and crime. In J. Q. Wilson & J. Petersilia (Eds.), *Crime: Public policies for crime control* (pp. 129-148). Oakland, CA: ICS Press.
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *Annals*, 587, 49-68.
- Feld, B. C. (1999). *Bad kids: Race and the transformation of the juvenile court*. New York: Oxford University Press.
- Feld, B. C. (2004). Juvenile transfer. *Criminology and Public Policy*, 3, 599-604.
- Fox, J. A. (2000). Demographics and U.S. homicide. In A. Blumstein & J. Wallman (Eds.), *The crime drop in America* (pp. 288-317). New York: Cambridge University Press.
- Gaes, G. (2005). Prison privatization in Florida: Promise, premise, and performance. *Criminology and Public Policy*, 4, 83-88.
- Garland, D. (2001). *The culture of control: Crime and social order in contemporary society*. Chicago: University of Chicago Press.
- General Accounting Office. (2003). *Justice outcome evaluations: Design and implementation of studies require more NIJ attention*. Washington, DC: Author.
- Glaze, L. E., & Palla, S. (2005). *Probation and parole in the United States, 2004*. Washington, DC: Bureau of Justice Statistics.
- Gottfredson, D. C., Najaka, S. S., & Kearley, B. (2003). Effectiveness of drug treatment courts: Evidence from a randomized trial. *Criminology and Public Policy*, 2, 171-196.
- Greene, J. R. (2000). Community policing in America: Changing the nature, structure, and

- function of the police. In J. Horney (Ed.), *Policies, processes, and decisions of the criminal justice system* (pp. 299-370). Washington, DC: National Institute of Justice.
- Harrison, P. M., & Beck, A. J. (2005). *Prisoners in 2004*. Washington, DC: Bureau of Justice Statistics.
- Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, 9, 85-110.
- Hodge, D. R., & Pittman, J. (2003). Faith-based drug and alcohol treatment providers: An exploratory study of Texan providers. *Journal of Social Service Research*, 30, 19-40.
- Hodgins, S. (Ed.). (1993). *Mental disorder and crime*. Newbury Park, CA: Sage.
- Howell, J. C. (2003). *Preventing and reducing juvenile delinquency*. Thousand Oaks, CA: Sage.
- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
- Kee, J. E. (1994). Cost-effectiveness and cost-benefit analysis. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (2nd ed.; pp. 506-541). San Francisco: Jossey-Bass.
- Kleiman, M. A. R. (2004). Toward (more nearly) optimal sentencing for drug offenders. *Criminology and Public Policy*, 3, 435-440.
- Kurki, L., & Morris, N. (2001). Supermax prisons. *Crime and Justice*, 28, 385-424.
- Lanza-Kaduce, L., Parker, K. F., & Thomas, C. W. (1999). A comparative recidivism analysis of releasees from private and public prisons. *Crime and Delinquency*, 45, 28-47.
- Lawrence, S., & Mears, D. P. (2004). *Benefit-cost analysis of supermax prisons: Critical steps and considerations*. Washington, DC: Urban Institute.
- Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 39-68). Thousand Oaks, CA: Sage.
- Lipsey, M. W. (1999). Can rehabilitative programs reduce the recidivism of juvenile offenders? An inquiry into the effectiveness of practical programs. *Virginia Journal of Social Policy*

and Law, 6, 611-641.

Lipsey, M. W., Adams, J. L., Gottfredson, D. C., Pepper, J. V., Weisburd, D. (Eds.). (2005).

Improving evaluation of anticrime programs. Washington, DC: National Academies Press.

Logan, C. (1993). Criminal justice performance measures for prisons. In *Performance measures for the criminal justice system* (pp. 15-59). Washington DC: Bureau of Justice Statistics.

Marion, N. E. (1997). Symbolic politics in Clinton's crime control agenda. *Buffalo Criminal Law Review*, 1, 67-108.

Marion, N. E., & Oliver, W. M. (2006). *The public policy of crime and criminal justice*. Upper Saddle River, NJ: Pearson.

Martinson, R. (1974). What works? Questions and answers about prison reform. *Public Interest*, 35, 22-54.

Marx, G. T. (1995). The engineering of social control: The search for the silver bullet. In J. Hagan & R. Peterson (Eds.), *Crime and inequality* (pp. 225-246). Palo Alto, CA: Stanford University Press.

Mears, D. P. (2001). The immigration-crime nexus: Toward an analytic framework for assessing and guiding theory, research, and policy. *Sociological Perspectives*, 44, 1-19.

Mears, D. P. (2005). A critical look at supermax prisons. *Corrections Compendium*, 30(5), 6-7, 45-49.

Mears, D. P., & Reisig, M. D. (2006). The theory and practice of supermax prisons. *Punishment and Society*, 8, 33-57.

Mears, D. P., Roman, C. G., Wolff, A., & Buck, J. (2006). Faith-based efforts to improve prisoner reentry: Assessing the logic and evidence. *Journal of Criminal Justice*, 34, 351-367.

Mihalic, S., Irwin, K., Elliott, D., Fagan, A., & Hansen, D. (2001). *Blueprints for violence prevention*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Millenson, M. (1997). *Demanding medical excellence: Doctors and accountability in the information age*. Chicago: University of Chicago Press.

- Miller, T. R., Cohen, M. A., & Wiersema, B. (1996). *Victim costs and consequences: A new look*. Washington, DC: National Institute of Justice.
- Miller, W. B. (1973). Ideology and criminal justice policy: Some current issues. *Journal of Criminal Law and Criminology*, 64, 141-162.
- Mrazek, P. J., & Haggerty, R. J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- Nadelmann, E. A. (2004). Criminologists and punitive drug prohibition: To serve or to challenge? *Criminology and Public Policy*, 3, 441-450.
- Nagin, D. (2001). Measuring the economic benefits of developmental prevention programs. *Crime and Justice*, 28, 347-384.
- O'Brien, R. M. (1995). Crime and victimization data. In J. F. Sheley (Ed.), *Criminology* (2nd ed.; pp. 57-81). New York: Wadsworth.
- O'Connor, T. P. (2004). What works, religion as a correctional intervention: Part I. *Journal of Community Corrections*, 14(1), 11-27.
- O'Connor, T. P. (2005). What works, religion as a correctional intervention: Part II. *Journal of Community Corrections*, 14(2), 4-26.
- O'Connor, T. P., & Pallone, N. J. (Eds.). (2002). *Religion, the community, and the rehabilitation of criminal offenders*. New York: Haworth Press.
- Penner, R. G. (2001). *Errors in budget forecasting*. Washington, DC: Urban Institute Press.
- Perry, A., McDougall, C., & Farrington, D. P. (Eds.). (2006). *Reducing crime: The effectiveness of criminal justice interventions*. Hoboken, NJ: Wiley.
- Perry, R. (2003, June 22). *Gov. Perry signs \$117 billion state budget: Governor uses line-item veto power to eliminate \$81 million in spending* [press release]. Austin, TX: Office of the Governor.
- Petersilia, J. (1991). Policy relevance and the future of criminology. *Criminology*, 29, 1-16.
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of randomized experimental evidence. *Annals*, 589, 41-62.

- Pizarro, J., & Stenius, V. M. K. (2004). Supermax prisons: Their rise, current practices, and effect on inmates. *Prison Journal*, *84*, 248-264.
- Pratt, T. C., & Maahs, J. (1999). Are private prisons more cost-effective than public prisons? *Crime and Delinquency*, *45*, 28-47.
- Regnerus, M. D. (2003). Linked lives, faith, and behavior: Intergenerational religious influence on adolescent delinquency. *Journal for the Scientific Study of Religion*, *42*, 189-203.
- Riveland, C. (1999). Prison management trends, 1975–2025. In M. H. Tonry & J. Petersilia (Eds.), *Prisons* (pp. 163-203). Chicago: University of Chicago Press.
- Roberts, J. V., & Hough, M. (2005). The state of the prisons: Exploring public knowledge and opinion. *Howard Journal*, *44*, 286-306.
- Roberts, J. V., & Stalans, L. J. (1998). Crime, criminal justice, and public opinion. In M. H. Tonry (Ed.), *The handbook of crime and punishment* (pp. 31-57). New York: Oxford University Press.
- Robinson, M. B. (2003). An obligation to make a difference in the real world? Thoughts on the proper role of criminologists and critical criminologists in the 21st century. *Western Criminology Review*, *4*, 226-238.
- Rosenbaum, D. P., Lurigio, A. J., & Davis, R. C. (1998). *The prevention of crime: Social and situational strategies*. Belmont, CA: West/Wadsworth.
- Rossi, P. H. (1980). The Presidential address: The challenge and opportunities of applied social research. *American Sociological Review*, *45*, 889-904.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Ruth, H., & Reitz, K. R. (2003). *The challenge of crime: Rethinking our response*. Cambridge, MA: Harvard University Press.
- Sabol, W. J. (1999). *Prison population projection and forecasting: Managing capacity*. Unpublished report, Washington, DC, U.S. Department of Justice, Office of Justice Programs.
- Sample, L. L., & Bray, T. M. (2003). Are sex offenders dangerous? *Criminology and Public*

Policy, 3, 59-82.

- Schmidt, J. D., & Sherman, L. W. (1996). Does arrest deter domestic violence? In E. S. Buzawa & C. G. Buzawa (Eds.), *Do arrests and restraining orders work?* (pp. 43-53). Thousand Oaks, CA: Sage.
- Sevigny, E. L., & Caulkins, J. P. (2004). Kingpins or mules: An analysis of drug offenders incarcerated in federal and state prisons. *Criminology and Public Policy*, 3, 401-434.
- Sherman, L. W. (2003). Misleading evidence and evidence-led policy: Making social science more experimental. *Annals*, 589, 6-19.
- Sherman, L. W. (2004). Research and policing: The infrastructure and political economy of federal funding. *Annals*, 593, 156-178.
- Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (Eds.). (2002). *Evidence based crime prevention*. London: Routledge.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. (Eds.). (1997). *Preventing crime: What works, what doesn't, what's promising*. Washington, DC: Office of Justice Programs.
- Sherman, L. W., Schmidt, J. D., & Rogan, D. P. (1992). *Policing domestic violence: Experiments and dilemmas*. New York: Free Press.
- Skogan, W., & Frydl, K. (Eds.). (2004). *Fairness and effectiveness in policing: The evidence*. Washington, DC: National Academies Press.
- Snyder, H. N., Sickmund, M., & Poe-Yamagata, E. (2000). *Juvenile transfers to criminal court in the 1990's: Lessons learned from four studies*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Stern, G. (2006, January 8). Faith-based confusion. *The Journal News*, pp. A1, A8.
- Stolz, B. A. (2002). *Criminal justice policy making: Federal roles and processes*. Westport, CT: Praeger.
- Sundt, J. L., Dammer, H. R., & Cullen, F. T. (2002). The role of the prison chaplain in rehabilitation. *Journal of Offender Rehabilitation*, 35, 59-86.
- Thomas, C. W. (2005). Recidivism of public and private state prison inmates in Florida: Issues

- and unanswered questions. *Criminology and Public Policy*, 4, 89-100.
- Toch, H. (2003). The contemporary relevance of early experiments with supermax reform. *Prison Journal*, 83, 221-228.
- von Hirsch, A., & Ashworth, A. (Eds.). (1992). *Principled sentencing*. Boston, MA: Northeastern University Press.
- Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals*, 578, 50-70.
- Weisburd, D., Lum, C. M., & Yang, S. (2003). When can we conclude that treatments or programs “don’t work”? *Annals*, 587, 31-48.
- Welsh, B. C., & Farrington, D. P. (2000). Monetary costs and benefits of crime prevention programs. *Crime and Justice*, 27, 305-361.
- Welsh, B. C., & Farrington, D. P. (Eds.). (2006). *Preventing crime: What works for children, offenders, victims and places*. New York: Springer-Verlag.
- Welsh, W. N., & Harris, P. W. (2004). *Criminal justice policy and planning* (2nd ed.). Dayton, OH: LexisNexis, Anderson.
- White, H. R., & Gorman, D. M. (2000). Dynamics of the drug-crime relationship. In G. LaFree (Ed.), *The nature of crime: Continuity and change* (pp. 151-218). Washington, DC: National Institute of Justice.
- Wilson, J. Q., & Petersilia, J. (Eds.). (2002). *Crime: Public policies for crime control*. Oakland, CA: ICS Press.
- Wright, R. G. (2003). Sex offender registration and notification: Public attention, political emphasis, and fear. *Criminology and Public Policy*, 3, 97-104.