National Institute of Justice

Adding Value to Justice Outcome Evaluations

Edwin W. Zedlewski

February 2010

NCJ 230402

Discussion Paper

NIJ

Adding Value to Justice Outcome Evaluations

Edwin W. Zedlewski

February 2010

NCJ 230402

Despite exhortations by policymakers to adopt evidence-based programs, there is little evidence that program managers read — or heed — evaluation findings. Absent evidence that an existing program does not work, managers logically prefer the status quo. One possible reason findings are ignored is that outcome evaluations rarely include information about program costs.

Program directors are making investment decisions in budget-constrained environments. Adopting a new program typically means cutting an existing one. Without the ability to compare two programs side by side in terms of relative effectiveness and relative cost, it is difficult for new programs to supplant existing ones. Cost-benefit analysis and cost-effectiveness analysis help inform these investment decisions. They transform the question of "what works" into a more sophisticated question: "Is it worth it?"

If the merits of cost-benefit analysis are so compelling, why are they not more commonplace? In part, it is because they are not as straightforward as they might seem. This paper illustrates the challenges encountered when incorporating cost analyses in common criminal justice settings. Four specific examples of evaluation and valuation challenges are presented, followed by concluding observations that discuss overarching principles in deciding on methodology.

Some Basic Terminology

Outcome evaluations in criminal justice have focused almost entirely on the issue of relative effectiveness: Does Program A produce better outcomes than Program B? This comparison assumes that the programs as well as the outcomes are reasonably comparable. Cost-benefit analysis and cost-effectiveness analysis do not change the nature of those comparisons, rather they enhance them.

Both cost-benefit and cost-effectiveness analyses shift the focus of the investigation away from the social scientist and toward the financial analyst or the policymaker (i.e., they present information more relevant to an investment decision.)[1] Both analyses measure efficiency — outcomes achieved per dollar invested in cost-effectiveness analysis and monetary returns per dollar invested in cost-benefit analysis.

Cost-effectiveness analysis connects the costs incurred to the program's outcomes. It computes input costs and compares them to outcomes obtained. In medical studies, for example, one commonly sees comparisons such as dollars spent per life saved.[2] In criminal justice, one might see presentations of the crimes averted per dollar invested.

Cost-benefit analysis goes one step further. It estimates or imputes a monetary value of outcomes, a process called valuation, or monetization. It then compares monetary benefits per dollar invested. In short, it determines the societal return on investment. Benefits are not always easy to convert into dollar terms. Cohen's (1988) research on victim costs made major breakthroughs in outcome valuation by estimating the monetary benefits of averted

[1] See for example; Rossi, Lipsey and Freeman (2004); Chapter 11, for discussions of the relative merits of cost-benefit analysis and cost-effectiveness analysis.

[2] Gold and colleagues (1996) document the lengthy experience of the public health sector in raising and answering questions related to choices between health treatment options.

victimizations, including victim pain, psychological suffering and risk of death. Less progress has been made on monetizing other benefits like prevention costs and the value of just outcomes.

The literature often portrays a benefit-to-cost ratio, a simple quotient of monetized benefits to monetized input costs. Zedlewski (1985), using a gross aggregate measure of the average social cost of a crime and self-reports by inmates of their offending rates prior to incarceration (Chaiken and Chaiken, 1982), estimated that the average offender created $430,000 in crime costs per year. At the time, housing an offender in prison cost $25,000 annually. Zedlewski arrived at a benefit-to-cost ratio of 17, or $17 in societal value returned for every dollar invested in incarceration.[3]

Should criminal justice evaluations incorporate valuations of outcomes? Yes, but that in no way implies that such applications are straightforward. The sections that follow illustrate the kinds of problems that can arise when analysts try to incorporate economic valuation into outcome evaluations.

The Cost-Effectiveness of Private Security: An Infrequently Explored Question

Households and businesses invest in locks, alarms, cameras and guards to protect themselves and their property. Annual revenues for private security firms in 2005 reached $37 billion in the United States, according to the U. S. Census Service Annual Survey.[4] In contrast, expenditures for police protection exceeded $98 billion in 2006 (Bureau of Justice Statistics, 2008). Expenditures on private security are no doubt underestimated because they address only expenditures on commercial private goods and services. They miss household purchases on deadbolt locks, dogs and weapons. They also miss important household decisions relating to risk avoidance, such as taking safe paths to work and choosing to pay a housing premium to live in safer neighborhoods.

Very little is known about the overall effectiveness (and cost-effectiveness) of specific private security measures because the analyst must estimate the number of crimes averted as aggregated across all measures in the area of interest. Clotfelter (1977) estimated that the relative demand for public and private security was a function of socio-economic factors and existing crime levels, but he provided no information about the effectiveness of these measures.

Zedlewski (1983), using a simultaneous equation model with both reported crime and victimizations, estimated the relative effectiveness of public and private security investments. This estimation was important because it disentangled the confounding of safety contributions between the two sets of resources. He found that public safety resources affected crime rates but private security resources did not; rather, they appeared to displace crime.

On the surface, evaluations of security investments aimed at protecting property should be well-suited for cost-benefit evaluations. The costs of protection are known, and one could assess the loss histories of households or firms that make the investment relative to those that do not. There

[3] Estimates apply to decisions made at the margin (i.e., the next person incarcerated will yield a 17 to 1 benefit-cost ratio, not that those returns can be counted on for massive increases in incarceration).

[4] Table 7.1 — North American Industry Classification System (NAICS) 5616, Investigation and Security Services.

would be minimal complications with factors such as fear, pain and suffering in the valuation of economic losses.

Such evaluations, however, are quite difficult in natural settings because individuals and firms do not make these investments in a single time period but instead, make them over extended periods of time. One could theoretically design a rigorous experiment by buying enough alarm systems to install randomly in houses throughout a target neighborhood so that differences among other security measures are randomized out of the analysis. Would this design produce a decent assessment of whether the alarms reduced burglary losses?

Probably not, or at least it would not produce a result that could be generalized to other settings. Burglar alarms are a form of target hardening. They try to bring attention to perpetrators so that police respond in time to foil the attempt. Study outcomes depend on the availability of police to respond in a timely manner. So alarms should perform favorably in high-density areas with police presence and poorly in rural settings where police are half an hour or more away from the scene. This co-production of public safety derived from the combined resources of police and private security) (Ostrom, 1973; Ostrom and Parks, 1984) has been a central theme in public-sector performance measurement but has not yet appeared in the evaluation literature.

A second consideration, noted by Hakim, Rengert, and Shachmurove (1996), is that the alarms create moral hazards because investors reduce their own vulnerability while increasing the risks of intrusion in unprotected houses. Suppose, for example, burglars observe physical signs on houses where alarm systems were installed. Other factors being equal, they would logically bypass those houses and target houses that had no signs. The difference in burglary rates observed between protected and unprotected houses is then greater than it would have been if there were no visible signs of protection.

Finally, alarm systems may reduce victimization risks for unprotected houses if they are installed in sufficiently large quantities. Burglary is a form of work and is lucrative only if the risk-adjusted returns exceed the wages that could be obtained from other kinds of labor. As the fraction of dwellings with alarms increases, the amount of time spent finding a vulnerable house increases as well, and the returns per hour of burglary decreases. In fact, the burglar's expected search time will rise in proportion to $p/(1-p)$, where p is the proportion of houses with alarms. Thus, well-guarded communities and enclaves may deter burglars altogether. Hakim, Rengert, and Shachmurove (1996) performed a cost-benefit analysis of burglary alarm systems by modeling these investments within a community. For the reasons given above, these findings should be extrapolated cautiously to other locations.

This example shows that valuation may be fairly straightforward in some settings. Nonetheless, evaluation design issues confound the estimation of effect sizes and consequently misestimate benefits.

Youth Delinquency Prevention Programs: Underestimated Benefits

Delinquency prevention programs are widely publicized as the correct way to control crime — by preventing or minimizing the onset of criminal careers — rather than reacting to criminal events. Much evaluation literature describes delinquency prevention, but few researchers have attempted to convert evaluations into cost-benefit analyses. The attempts that have been made

(Greenwood, 1996; Aos et al., 2004), although they should be applauded for their initiative, contain serious problems in approach and measurement.

Consider as a case in point Greenwood's (1996) cost-benefit analysis of a classic experiment that provided low income families free home nursing visits during the first two years after childbirth, with families randomly assigned to receive the service or not. A long-term evaluation (Olds et al., 1997) demonstrated that mothers in the program had lower levels of criminal activity and fewer incidents of child neglect, and that children in the program had significantly fewer delinquency experiences than children in a control group. These findings are considered to be sound, and Greenwood's monetization of outcomes met the standards that one would have imposed more than a decade ago.

Greenwood's analysis brings out the difficulties in trying to retrofit benefits valuation to evaluations. His analysis failed to capture certain nonsafety benefits, in part because the data had not been collected. By improving parenting and general family functioning, one would expect the intervention to produce significant outcomes such as educational achievement, eventual employment and a variety of pro-social activities. These social benefits were not captured by the analysis and therefore the overall social benefits were understated.

Cohen (1998) developed an alternative valuation estimate of a more comprehensive set of benefits. He calculated the discounted net present value of three anti-social outcomes: lack of a secondary school diploma, lifetime drug addiction and career criminality. He included outcomes such as lost wages and productivity, health costs, victim costs and criminal justice costs. Lack of a diploma led to approximately $300,000 in lost productivity. If one person had all three anti-social outcomes, it would cost society $1.3 million. Estimating these costs from actual programs would require waiting 30 or more years after a program is over; however, one can infer that savings of that magnitude are likely to accrue if a program demonstrates effectiveness during its first few years.

Nonetheless, policymakers should not make the leap of faith that wide-scale expansion — or going to scale — will yield similar benefits. Delinquency prevention programs have typically been small-scale and have usually been shepherded by researchers of considerable talent and expertise. The challenge is to bottle the innovation — to take a program that appears to be very promising in a single setting and expand it to the national level. When programs leave behind the minds, skills and personal attention of their originators and are re-created by less skilled or less committed replicators, they typically perform less effectively.

"Going to scale" challenges analysts to identify and monetize the intangible resources that facilitated the original outcomes. It also requires analysts to predict how program performance would be diminished if the program were expanded to a larger scale. Analysts would benefit from evidence-based guidelines as to what kinds of "scale degradation factors" should be employed within the valuation process when predicting performance outcomes on a national scale.

Police Interventions: MonetizingFear of Crime

Horowitz and Zedlewski (2006) recently completed an assessment of the readiness of 20 rigorously designed police outcome evaluations to support a cost-benefit analysis. They found

that only two of the 20 evaluations reviewed contained any information on program resources. Moreover, many of the key outcome variables could not be monetized. The discussion below brings out some of the more prevalent issues.

The two evaluations cited by Horowitz and Zedlewski (2006) that specified resources adequately (Sherman and Rogan, 1995; Sherman and Weisburd, 1995) were experiments with patrol tactics. Cities were divided into sectors and police implemented the experimental tactic in randomly chosen sectors, while in the remaining sectors police patrolled as usual. The outcome variables were relative changes in crime incidence over time between the experimental sectors and the control sectors. For both the experimental and control sectors, the researchers determined the number of officer hours consumed and the number of patrol car hours consumed. The difference between the experimental and the control resources is called the "dosage" of the intervention.

Depicting the incremental resources consumed by an intervention is quite different from determining the cost of the resources. Nonetheless, applying estimates of direct wage rates to the officer hours is probably even a worse approach than foregoing cost estimations.

Wayson and Funke (1989) showed that full costs of an officer hour are at least five times the cost of direct wages when training investments, vacation and sick leave, health and other fringe benefits, retirement, and direct and indirect supervision are accounted for. Given the enormous disparities in both wages and indirect costs across European Union countries, one could easily see the advantages of stopping at descriptions of the resources consumed in their natural terms (e.g., direct labor hours, management overhead, vehicle use). Fully accounting for all direct and indirect costs is expensive and presents estimates that are less relevant to other European Union members.

The worst examples in terms of assessing resources consumed were studies involving task forces that had been assembled to address certain drug or crime problems. In no instance did evaluators adequately describe the composition of these task forces in terms of numbers of members or the number of hours the team spent in meetings and other activities.

In some studies, evaluators assessed interventions that targeted neighborhoods or neighborhood "hot spots." Conventional wisdom, now documented by empirical results (Sherman and Weisburd, 1995), suggests that concentrating police resources in a troubled area can suppress crime and disorder for a period of time.

Such interventions abound with both valuation and evaluation issues. One is whether the intervention creates a temporal or a real displacement of crime; that is, whether the intervention truly helped to avert crimes or whether it merely postponed or displaced crimes. Another issue is whether temporarily diverting police resources from one area to obtain concentrations of force in the targeted area increases crime in the diverted areas. Using reported crime as evidence, the limited research available seems to support the idea that police should target "hot spots."

In addition to crime outcomes, researchers have used crime correlates as outcome variables. Two commonly used measures are neighborhood disorder and fear of crime. In their famous "broken windows" essay, Kelling and Wilson (1982) argued that signs of disorder such as graffiti, trash in the streets, and prostitutes and drug dealers operating in the open all act to encourage additional unlawful behaviors. By adopting a "zero tolerance" attitude for these behaviors and aggressively attacking them, police can reduce the prevalence of more serious crimes.

Several evaluations have examined this type of strategy, largely by measuring fear and disorder before and after the intervention. Their general approach has involved surveying residents in the neighborhood at two points in time, but the surveys have been fairly unsophisticated and have not used a common set of questions. For example, the questions posed might include, "How afraid are you to walk in your neighborhood at night?" or "Do you think that conditions in the neighborhood have improved?"

Economists can contribute to this sort of measurement plan. For example, an economist might infuse a cost-benefit analysis into the evaluation by measuring the amounts of graffiti, trash and other signs of public disorder and calculating the costs of their removal prior to the intervention and again after implementation of the program or policy. Evaluators can also contribute to such measurements by conducting "drive by" surveys to photograph the entire neighborhood and then tally their observations of various types of disorder. Drive-by surveys are less expensive than door-to-door surveys and give more concrete assessments of disorder.

Monetizing the fear of crime, another frequent measure of program outcomes, is even more complex. Fear is psychological loss. It translates into costs through behavioral changes such as traveling longer routes to avoid risky areas and foregoing exercise like walking and jogging. Analysts can estimate a person's time lost from fear of crime by examining the time wasted in changing travel routes and perhaps by determining changes in the numbers of health club memberships. Policymakers should also try to determine whether the fear manifested is consistent with the dangers actually posed. If it is consistent, then reducing fear through activities like neighborhood goodwill campaigns could actually induce residents to change behaviors in ways that increase victimization risks.

Offender Rehabilitation: Weak Measures of Reoffending

Despite a well-developed literature on outcome evaluations that deal with offender rehabilitation programs, there has been erratic progress on the valuation of rehabilitation benefits. The drug treatment literature offers a robust body of findings (Belenko, Patapis, and French, 2005) that takes into account crimes averted, health care savings and employment gains. Other relevant fields such as offender employment, training and cognitive restructuring offer scores of rigorous evaluation designs and well-conducted outcome evaluations in both institutional and community corrections settings. But they offer very little in the way of valuation results. Perhaps Cohen's cost estimates of antisocial outcomes could be adjusted to provide estimates of benefits to the offender; but the broader question of public safety benefits needs to be addressed, and that requires a changes in how corrections outcomes are measured.

A critical shortcoming of correctional evaluations is the reliance on recidivism as the central outcome measure. Whether measured by rearrest or reconviction, recidivism is simply an event — the recorded outcome of a single crime. Although it tells us that the intervention has not

resulted in total cessation of criminal acts, we are left to guess whether the offender is now more or less active. Survival analysis, or measuring the time to recidivism, is only somewhat better, because times to failure (rearrest) can be compared to times between previous arrests, and researchers can surmise whether an offender has increased or decreased his activity. Even in this type of analysis, however, researchers still fall short of estimating a numerical crime savings.

Using published studies of self-reports by offenders is one approach to estimating crime savings. Chaiken and Chaiken (1982) interviewed more than 2,000 incoming inmates to estimate the mean and median number of crimes that they committed in the year prior to their incarceration. MacKenzie et al. (1999) interviewed probationers to determine whether they decreased their offending rates or desisted altogether while on probation. They found significant declines in offending rates prior to sentencing, and still further declines after the offenders were placed on probation. Thus, crime savings from imprisonment should be compared to crime savings from probation.

Bhati (2007) has developed a promising approach to modeling offender crime trajectories based on their previous offending patterns. By appending an offender's post-intervention arrests to the original history, Bhati is able to compute a revised trajectory. The differences between the two trajectories represent the crime savings attributable to the intervention. More computational experience is needed before this approach can be recommended with confidence. An alternative approach is to develop baseline estimates of convicted offenders' self-reported annual offending rates. These rates can be used to calculate the ratios of admitted offenses to arrest frequencies. The ratios can in turn be used to impute post-intervention offending rates to post-intervention arrest histories.

Readers should keep in mind that crime savings may only be one of several program benefits. Other benefits often encountered included educational attainment and employment, improved health (from drug abstinence) and improved family functioning. Both evaluators and economists should consult with program advocates and policymakers to determine which benefits might be measured and by what methods. Benefits that will not be measured should still be detailed in any report.

Other programs often encountered in corrections settings are those designed to reduce prison populations. Among the strategies observed in practice are early-release programs, house arrest, and intensive community supervision with random drug testing and other forms of surveillance. In developing cost-benefit comparisons between community-based programs and incarceration, per capita program costs — including costs incurred from rearrests, convictions and new crimes — are compared against average per capita prison costs. Small-scale community programs cannot serve enough clients to warrant closing a prison, so the relevant cost comparisons are the marginal costs of confinement (e.g., food, clothing and some modest savings in supervision not average costs (e.g., costs of facility maintenance, support services such as food and health care and prison management). Perhaps such savings could be realized if they were summed over a large number of similar programs across a state. Still, policymakers should view such propositions with appropriate caution.

What Guidance Is Appropriate for Evaluators?

A primary purpose of this paper is to show that there are few overarching principles to apply to the merger of valuations and evaluations. Program settings generate a wide range of unique challenges and considerations, and they must be identified and accounted for during the planning phases.

Economists should participate in evaluations. Universal guidance shows that economists need to be part of the evaluation team at the outset. They need to participate in identifying outcomes and outcome measures. They need to establish expectations before the evaluation begins that state whether there will be a cost-benefit analysis or a cost-effectiveness analysis. Their participation at the planning stage will help economize the data collection and analysis and contribute to identifying appropriate evaluation designs and measurements.

Joining criminologists and economists is a very challenging task. These disciplines follow very different research philosophies and use very different research methods. Criminologists prefer direct observation and measurement; they estimate results through experimental and quasi-experimental evaluation designs. Economists dislike original data collection (and especially opinion surveys); they estimate results through multivariate methods and econometric modeling. Moreover, few economists are knowledgeable about criminal justice processes, let alone the research literature. In coordination, criminologists should lead evaluation design teams and use economists to help develop policy-relevant outcome measures.

Measure proximal rather than distal outcomes. The causal proximity in the model between what the evaluation will measure overall and what social outcome will be monetized should strongly influence the decision to pursue a cost-benefit or a cost-effectiveness study. As a rule, one would want the evaluation results that are measured to show a close causal link to the monetized benefit. In a police intervention, for example, the evaluation may primarily report increases in the numbers of apprehensions. Then the evaluator could simply report the number of additional apprehensions achieved per additional resource consumed, stopping at cost-effectiveness. Alternatively, the evaluator could conduct longer-term evaluations that measured the outcomes in terms of the criminal sanctions applied after the arrests. Then the evaluator could measure crime savings due to short-term deterrence effects and long-term incapacitation effects. Because these crime savings are so causally distant from program resources, analysts could only make tenuous attributions of crime savings to the intervention. In contrast, a correctional rehabilitation program would logically measure and monetize crime savings because there is a short causal chain between program results and crime savings.

Use logic models to inform evaluation design decisions. A corollary piece of guidance is that evaluation planners should employ logic models (See Shadish, Cook and Campbell, 2002, for a thorough discussion of logic models) to inform the evaluation design and the decision between cost-benefit and cost-effectiveness analyses. Logic models are diagrams that depict the causal chains through which program resources will produce program activities that create program outputs and program outcomes. They also enable planners to identify the exogenous contributors to program outputs and outcomes (e.g., support agencies, citizen involvement and private resources) and ensure that these contributors are part of the resource measurement strategy.

Employ appropriate methods. More sophisticated analyses of social investments lead to greater understanding and the potential for better investment decisions. Making important investment decisions absent any cost information will likely cause decision-makers to overlook viable

alternatives and waste public resources. Economic analysis of evaluative studies deserves a resounding endorsement. Nonetheless, researchers must determine how to decide which level of sophistication is appropriate for which decisions.

A few obvious questions to ask include:

How imminent is a decision on program options? Have decision-makers voiced interest in costs and benefits? How large are the potential savings? On balance, the appetite for sophisticated analysis should increase with the level of government — local decision-makers would confront smaller and more immediate issues than national decision-makers — and the sophistication of the decision-making machinery would increase as well.

For those few instances where the decision lies between a cost effectiveness analysis and a cost benefit analysis, the small checklist of questions below may be useful.

- Are there fair and reasonable alternatives in terms of desired outcomes and relevant target populations? It is easy to assess apples against oranges when focusing on benefit-cost ratios. Aos (2004) identified dozens of cost-beneficial programs and portrayed them in broad categories such as delinquency prevention and community corrections. That does not mean that one can simply select one of his categories and pick out the best program by scanning for the highest benefit-to-cost ratio. Programs may target different needs such as employment or drug treatment. They may target different populations such as preschool or elementary school children. Even assessing probation programs against prison programs is hazardous because prison inmates will pose greater risks of serious offending than probationers. Take care to ensure that the treatment populations are comparable in the experimental and comparison groups.

- How completely can resources be identified and estimated? The more comprehensive a program is in scope, the more likely that hidden resources exist. Consider a program that tries to ensure the continuity of care for drug-involved offenders from the point of arrest to departure from probation. These offenders will draw resources from police, prosecution and defense, courts, treatment providers, probation officers, family service workers and possibly jails. If the evaluation does not contain an event-based tracking system, it will surely miss the effects of some resources on these offenders.

- How thoroughly have program outcomes been identified? The preceding example suggests two obvious outcomes: cessation of drug use and rearrest. Both of these are commonly reported on various studies of offender populations. Many other indicators of program performance are less commonly reported, perhaps because of cost considerations, but their omission obscures the interpretation of findings. An evaluation might find, for example, that there was little difference between rearrest rates for the experimental and comparison groups. Rearrest rates are noisy variables, because they only modestly reflect underlying offending. Insightful information regarding many dimensions of program performance can be gleaned from client data about job stability, housing stability and family functioning. A corollary observation is that every outcome measure is likely to have its own constituency or evaluation consumer.

- Can costs and benefits be identified on a short-term and long-term basis? Analysts must read the findings from the evaluation of startup programs with extreme caution. A new program employs staff who are inexperienced in delivering the services and activities available through the program. Links to complementary services within organizations or to external resources will not be in place. Additionally, this kind of exploratory intervention will not have data systems to routinely collect performance information. Nonetheless, short-term costs must be used to perform a cost-benefit analysis. The good news is that any costing performed under these circumstances is likely to overstate costs and provide sensible reasons for permanent adoption.

- How many of the program outcomes can be monetized? This is the most trying question of all. If analysts can only measure crime savings, then the program is a crime-reduction program. If analysts can also monetize health costs from drug abuse, then the program is a drugs and crime program. Any program dealing with sexual assault victims should logically involve police, victim services and the medical profession. It is possible to measure outcomes relevant to constituencies like these, but monetizing such outcomes may not yet be possible. Criminal justice has made great strides in monetizing crime outcomes, but little else. Practitioners do not know how to estimate the various social costs of justice or its associated outcomes. They cannot estimate the cost of a wrongful conviction or the benefit of saving a police officer's life. Until practitioners can estimate these costs and benefits, they should treat cost-benefit analysis cautiously. At a minimum, any cost-benefit analysis should enumerate the costs that were measured and the costs that were not.

Concluding Comments

This paper presents a formidable array of challenges to cost-benefit analysis in program evaluations. It does not attempt to defer future activity on the topic. Rather, it attempts to gauge whether the cost-benefit glass is half empty or half full. My assessment is that the glass is about a quarter full — full of validated promise but largely empty on wide-scale delivery. I remain optimistic that clever people in many fields can find ways to value the intangible and unravel the intractable. Hopefully, this paper highlights some of the most important areas for exploration by criminologists and economists.

Acknowledgments

I would like to acknowledge the able assistance of Lisa Russell, NCJRS International Center Liaison, in producing research materials for this report.

References

Aos, S., R. Lieb, J. Mayfield, M. Miller and A. Penucci. *Benefits and Costs of Prevention and Early Intervention Programs for Youth,* Washington State Institute for Public Policy, Report 04-07-3901, July 2004.

Belenko, S., N. Patapis and M.T. French, *Economic Benefits of Drug Treatment: A Critical Review of the Evidence for Policy Makers*. St. Charles: MO: Alcohol and Drug Problem Association of North America, 2005.

Bhati, A. "Estimating the Number of Crimes Averted by Incapacitation: An Information Theoretic Approach." *Journal of Quantitative Criminology* 23 (4) (2007): 355-375.

Bureau of Justice Statistics, *Justice Expenditure and Employment Extracts, 2006.* Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics, 2008. Available online at http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=1022.

Chaiken, J. and M. Chaiken, *Varieties of Criminal Behavior*. Santa Monica, CA: Rand Corporation, 1982.

Clotfelter, C.T. "Public Services, Private Substitutes, and the Demand for Protection Against Crime." *American Economic Review* 67 (1977): 867-877.

Cohen, M.A. "Pain, Suffering, and Jury Awards: A Study of the Cost of Crime to Victims." *Law and Society Review* 22 (1988): 537-555.

Cohen, M.A."The Monetary Value of Saving a High Risk Youth." *Journal of Quantitative Criminology* 14 (1998): 5-33.

Gold, M.R., J.E. Siegel, L.B. Russell and M.C. Weinstein, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.

Greenwood, P.W., K. Model, C.P. Rydell and J. Chiesa. "Diverting Children from A Life of Crime: What Are the Costs and Benefits?" Santa Monica, CA:,RAND Corporation, 1996.

Hakim, S., G.F. Rengert and Y. Shachmurove. "Estimation of the Net Benefits of Residential Electronic Security." *Justice Quarterly* 13 (1996): 153-170.

Horowitz, J. and E. Zedlewski. "Applying Cost Benefit Analysis to Policing Evaluations." *Justice Research and Policy* 8 (2006): 51-66.

Kelling, G. and J.Q. Wilson. “Broken Windows: The Police and Neighborhood Safety.” *Atlantic Monthly* (1982): 29-37.

MacKenzie, D. K. Browning, S.B. Skroban and D.A. Smith. “The Impact of Probation on the Criminal Activity of Offenders.” *Journal of Research in Crime and Delinquency* 36 (4) (1999): 423-453.

Olds, D.L., J. Eckenrode, C.R. Henderson, Jr., H, Kitzman, J. Powers, R. Cole, K. Sidora, P. Morris, L.M. Pettitt and D. Luckey. “Long-Term Effects of Home Visitation on Maternal Life Course and Child Abuse and Neglect: Fifteen-Year Follow-Up of a Randomized Trial.” *Journal of the American Medical Association* 278 (1997): 637-643.

Ostrom, E. “On the Meaning and Measurement of Output and Efficiency in the Provision of Urban Police Services.” *Journal of Criminal Justice* 1 (1973): 93-111.

Ostrom, E. and R. Parks. “Policing as a Multi-Firm Industry.” In *Understanding Police Agency Performance,* ed. G. Whitaker, Washington, DC: U.S. Department of Justice, National Institute of Justice, 1984, pp. 7-22.

Rossi, P.H., M.W. Lipsey and H.E. Freeman. *Evaluation: A Systematic Approach,* 7th ed. Thousand Oaks, CA: Sage Publications, 2004.

Shadish, W., T. Cook and D. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.*. New York: Houghton Mifflin Company, 2002.

Sherman, L. and D. Rogan. “Effects of Gun Seizures on Gun Violence: ‘Hot Spots’ in Kansas City.” *Justice Quarterly* 12 (1995): 673-694.

Sherman, L. and D. Weisburd. “General Deterrent Effects in Crime ‘Hot Spots’: A Randomized Controlled Trial.” *Justice Quarterly* 12 (1995): 625-648.

Wayson, B. and G. Funke. *What Price Justice?; A Handbook for the Analysis of Criminal Justice Costs*. Washington, DC: U.S. Department of Justice, National Institute of Justice, 1989.

Zedlewski, E. “Deterrence Findings and Data Sources: A Comparison of the Uniform Crime Reports and the National Crime Surveys.” *Journal of Research in Crime and Delinquency* 20 (1983): 262-276.

Zedlewski, E. “When Have We Punished Enough?” *Public Administration Review* 45 (1985): 771-779.