

The obstacles that hamper multisite evaluations in criminal justice settings can be overcome through an understanding of how the criminal justice system is organized and how it operates.

Multisite Evaluations in Criminal Justice Settings: Structural Obstacles to Success

Wesley G. Skogan, Arthur J. Lurigio

This chapter examines some structural impediments to multisite evaluations (MSEs) in the criminal justice domain. Besides the usual methodological considerations involved in conducting credible evaluations, these structural features impose social, political, and organizational constraints on evaluation research, which make MSEs difficult and risky. In this chapter we describe the obstacles to MSEs and then review two projects that succeeded despite these impediments. A third, notable criminal justice MSE, the Spouse-Assault Replication Program, is described by Reiss and Boruch (this volume). These cases illustrate how evaluators might overcome—or learn to live with—the roadblocks to MSEs within the criminal justice system.

There are several characteristics of the criminal justice system that hamper the implementation and evaluation of multisite programs. To begin, the system is extremely *decentralized*. Police departments, for example, can operate within the province of municipalities, counties, campuses, public housing, mass transit, and the states. There are nearly 23,000 independent policing agencies within the United States. The criminal justice system is also highly *fragmented*. Cities administer police departments and jails; counties administer sheriffs' and prosecutors' offices, jails, and probation agencies; state governments run the prisons. Agencies are embedded in disparate political settings, each with its own priorities for taxing and spending. The federal government plays a minor role in financing local systems (only 8 percent of the criminal justice system budget is federal) and in establishing standards and sparking innovation.

Criminal justice agencies are traditionally very *insular*. Limited mobility or interchange occurs across jurisdictional boundaries. This is coupled

with a strong sense that local situations are unique and a "not invented here" attitude toward innovation, which is especially prevalent in policing. Many practitioners are not very sophisticated with respect to research; only a high school diploma is required for entry-level positions, and in most places a university degree has not been seen as a prerequisite for promotion into the ranks of management. The research and planning divisions of criminal justice operating agencies typically perform little of either activity; usually their primary task is to draft new forms and internal regulations, and their objectives are narrowly focused. Successes and (especially) failures are not shared among agencies, and information regarding new developments seldom moves beyond department perimeters. In the absence of a tradition of research or even professional communication, no real mechanism exists for broadly communicating about programs or policies.

Because much of their work is *labor intensive*, criminal justice agencies have not faced the prod of technological change. For example, few changes have appeared in the technology of police work since the introduction of car radios in the 1930s, and, if anything, prisons have regressed, becoming more overcrowded and less able to provide services to inmates. The prominent exception to this rule has been the recent development of computer and telephone automation supporting more effective systems of home detention (for example, see Lily, Ball, and Wright, 1987).

Criminal justice agencies also have great difficulty in devising *meaningful performance measures*. This has impeded any substantive focus on enhancement of productivity, a source of innovation in many organizations; criminal justice agencies generate many activity counts but do not have a good handle on their effectiveness at solving their target problems. The lack of standardized performance criteria has made it very easy to manipulate indicators of agency effectiveness; statistics such as a prosecutor's "conviction rate" cannot wisely be taken at face value. Some indices, such as revocation rates in probation, can be double-edged. Violations may be interpreted either as successes (probation officers are agents of the court and their job is to monitor infractions and report them summarily) or as failures (officers are also agents of change and their job is to help offenders relinquish criminal activity).

Criminal justice agencies also foster a *subculture of secrecy* concerning their work. This has several functions: It obfuscates the tremendous discretion that many functionaries in the system enjoy, it shields the general public from the role that violence plays in their daily work, and it masks from view the corruption problems that plague law enforcement. This subculture has serious consequences for evaluators, who are seen readily as "snoops" for management, the courts, or individuals with political agendas. Line staff easily adopt an "us against them" mentality toward outside evaluators. Furthermore, police and correctional administrators work in a heavily unionized environment that sometimes interferes with efforts to redeploy personnel, their only flexible resource.

Criminal justice agencies generally exist in highly charged political environments. They are the most visible components of local government, as the media, who historically have assumed a watchdog or adversarial posture toward the system (Lavrakas, Rosenbaum, and Lurigio, 1990). Evaluators may document poor or inadequate performance; worse, they may stumble upon evidence of impropriety or malfeasance. Judges, prosecutors, and sheriffs generally are elected officials, and police chiefs and state corrections administrators stand close to their mayors and governors and have a vested interest in seeing them remain in office. Criminal justice policy-making is both emotionally laden and costly, leading everyone involved to try to look good, err on the side of caution, and when in doubt take the conservative course. Researchers easily can seem unappreciative of the sensitive and precarious political environments in which criminal justice agencies exist (Lurigio and Skogan, 1990). Indeed, evaluators would be wise to work in communities with stable political and agency leadership, for a turnover at the top can easily bring in new leaders who are not interested in the prior administration's pet projects.

Finally, the criminal justice system operates within a *context of individual rights*. The participants have a strong sense of legal constraint in procedural issues, an unwillingness to risk injustice in individual cases, and a stated (although not actually delivered) commitment to providing individualized treatment. This translates, for example, into a general aversion to the concept of random or unbiased assignment, which is the hallmark of the best designs for yielding interpretable information about programs.

Two Examples of Multisite Evaluations

These impediments to evaluation in the criminal justice arena can be seen in two multisite programs. The first was an evaluation of new policing programs, which was conducted in multiple neighborhoods in two cities. The second was an evaluation of intensive probation supervision that was conducted in eleven sites across the country. Both the programs and their evaluations were shaped in important ways by the decentralized, fragmented, insular, politicized character of the system in which they were embedded, and by its rights-oriented environment.

The policing evaluation illustrates how innovative program ideas flowed from jurisdiction to jurisdiction, the process by which the original program construct was transformed by local politics and leadership factors, why the agencies wanted to be involved in the MSEs, especially in the development of performance measures, and the politics of disseminating research findings. The probation evaluation illustrates how programs diverged from place to place, how the discretionary nature of the probation process shaped everything from the client pool for the evaluation to the out-

comes measures, and the problems involved in coordinating random assignment and collecting data in a rights-oriented environment.

Community Policing in Houston and Newark. Late in 1982, the National Institute of Justice (NIJ) (the research arm of the U.S. Department of Justice) sponsored an evaluation of what has come to be known as "community policing." Community policing involves organizational and programmatic adaptations that promise to make law enforcement more responsive to the needs of communities. This responsiveness is accomplished through organizational changes that open departments to public input concerning their priorities and procedures, and that adopt a broad, problem-solving orientation toward local issues.

To find out how well these changes might work, new policing strategies were implemented in neighborhoods in Houston, Texas, and Newark, New Jersey (for a description of the programs, see Skogan, 1990; Skolnick and Bayley, 1986). The two cities are extremely different. Houston is a sprawling, low-density, low-rise, garden-apartment city laid out for easy access to freeway interchanges. The 380,000 residents of Newark are packed into nineteen square miles of old wood-frame homes, apartments, and public housing blocks, arrayed along narrow, crowded streets.

The policing strategies were locally planned, and they reflected differences in the problems targeted for action and in the approaches used to solve them. The Houston planning task force focused on the lack of contact police had with ordinary citizens, and on the city's almost nonexistent neighborhood life. The members of the task force thought that Houston lacked any sense of local community or tradition of collective self-help, making it difficult for neighborhoods to respond to disorder and crime. They also believed that people came into contact with police only under stressful circumstances that forestalled any informal communication between them. Newark, on the other hand, ranks near the top on almost every indicator of big-city problems, and the planning task force decided that its first objective was to demonstrate that the police still controlled the streets and could exert their authority to discipline those they considered out of line. Thus, the list of problems that community policing was to address turned out to be quite different in each city.

The planning and implementation process also worked quite differently in the two cities, reflecting differences in the organizations and the local leadership styles. In Newark, the planning and implementation process was "top-down," that is, the command staff of the police department and outside experts planned their strategies, which were then implemented by teams of rank-and-file patrol officers under the command of their sergeants and district commanders. In Houston, on the other hand, most elements of the program were designed by a group of patrol officers headed by a sergeant from the Planning Division of the police department; the same officers then took charge of individual strategies, recruited a few

more officers from the districts where the program was targeted, and carried out the program themselves. Newark's approach exemplified its traditional, paramilitary management style, while Houston's reflected the police chief's masterplan for departmental decentralization.

These two planning processes produced quite disparate versions of community policing. Houston's program stressed local problem solving and granted the patrol officers in charge a great deal of autonomy. Police opened a storefront office, walked door-to-door gathering information about local problems, fostered local community organizations, wrote and distributed newsletters, and initiated a program serving crime victims. Newark used its preexisting management structure and incorporated more traditional enforcement tactics into its programs. As in Houston, police opened a storefront office and conducted door-step interviews. In addition, they organized harsh crackdowns on street gatherings and public drinking, intensified levels of foot patrol, set up roadblocks to screen drivers for drinking and outstanding warrants, and ejected from buses people who were drinking, eating, or smoking. These activities brought police into direct confrontations with more people than ever before, and the approach was considered very controversial. The crackdowns on street congregations were probably unconstitutional, and the roadblocks were at best heavy-handed (Skolnick and Bayley, 1986).

There were no federal funds to pay for the strategies, so each police chief had to find internal resources, principally personnel, to conduct the programs. In each case it was extremely difficult to justify the disconnecting of officers from the constant stream of incoming emergency or "911" calls. Newark's department shrank by almost 30 percent during the 1970s because the city was virtually bankrupt. To garner enough officers to conduct the program there, it was necessary to reassign administrative personnel to street duty, make staff schedule changes that were in violation of the city's contract with the police union, and push the department's resources to the limit. Newark also tried to involve the school district in a youth recreation component for the program, but in a year the department was unable to overcome bureaucratic obstacles to opening a local school building after dark. Houston's department had not expanded to match the explosive growth of the city during the 1970s, and it was having difficulty answering emergency calls throughout its vast territory. Houston put only a few officers into the pilot projects and ran them virtually without supervision because the chief did not yet have sufficient political and organizational capital to invest more in the effort.

These personnel shortages did make it easy to sell a quasi-experimental design to the police because there was no hope of taking the programs city-wide. Areas, not individuals, were allocated to treatment or control status, and as policing is organized along area lines, the fielding of demonstration projects and setting aside of comparison areas fit their mode of thinking.

How did the two planning task forces come up with innovative program plans? The programs in Houston and Newark were formulated with the assistance of the Police Foundation, a research organization based in Washington, D.C., that fosters police innovation. The foundation provided technical assistance to the two departments during their planning effort, which explains some of the similarity of the two programs. Foundation staff "hothoused" the planning process by bringing in outside experts and by having task force members visit successful programs in other cities. Some task force members were trained by professional community organizers and victim services administrators, while others visited cities like Detroit and Santa Ana, California, to talk with officers already involved in community policing. Senior foundation advisers virtually set up residence in each of the cities for almost six months during the planning phase.

The Police Foundation also conducted the evaluation of the programs, taking on a dual role that raised eyebrows in some circles. To facilitate this evaluation, elements of the program were implemented in different areas of Houston and Newark, and one comparison area was identified in each to mark citywide changes. The comparison areas were matched with program areas on several key demographic variables. The evaluation involved observations of the strategies in action, monitoring of events in the cities and in the police departments, interviews with participating officers, and the collection of large amounts of administrative data. (The evaluators, however, were not allowed to read key sections of offense reports made out when people were victimized in the program areas; such openness was judged to be "contrary to departmental policy.")

The success of the projects also needed to be judged along dimensions for which it is difficult to capture routine performance measures. The program was expected to increase police visibility in the target neighborhoods, enhance public confidence in the police, encourage community cohesion and self-help, and reduce fear of crime. Large and expensive sample surveys were conducted in all of the program and comparison areas to measure these outcomes. Since it was not clear at the outset what community policing officers would actually *do* during their year in the field, the surveys included a grab bag of potential process and outcomes measures; at the end of the field period the evaluators then had to negotiate the measures for which each individual team was to be held "responsible" in the report, and which activities should be described but could not be systematically assessed. This process of negotiation greatly increased the program's chances of looking good.

What did the evaluation conclude? Community policing, as implemented in Houston and Newark, showed some success at responding to neighborhood needs. The teams found distinct ways of reaching out for community input and support, and they all succeeded on at least a few of the evaluation measures. The programs in both cities achieved a surprising level of visibility, and in almost every instance disorder went down and

citizen satisfaction with their neighborhoods and with the police went up. The traditional enforcement efforts mounted in Newark had few measurable benefits, however, other than driving down levels of visible street disorder.

The worrisome finding of the evaluation was that the benefits of the Houston program largely were confined to whites and home owners of the target areas (see Skogan, 1990). Houston's police chief insisted that selected program neighborhoods have black, Hispanic, and non-Hispanic white residents. This ensured that no ethnic or racial group was "left out" of the program. Analyses of the data from this heterogeneous sample suggested that the program had differential effects. Also for political reasons, the Newark evaluations were conducted only in uniformly poor, black areas, and as a consequence the evaluation had no opportunity to detect treatment interactions by race or class.

Intensive Probation Supervision Programs. The field of corrections is in crisis because of prison overcrowding. The push to alleviate institutional crowding has resulted in the sentencing of serious offenders to a number of "intermediate" treatments short of prison, ranging from harsh "boot camps" (MacKenzie and Shaw, 1990) to fines based on income (Hillsman, 1990) and home confinement (McCarthy, 1987).

One of the most visible intermediate sanctions is intensive probation supervision (IPS). As its name implies, IPS is more strenuous than routine probation. Typically, IPS offenders have multiple weekly contacts with their probation officers; they also are held strictly to curfews and other conditions of release, subjected to unscheduled drug tests, and ordered to perform community service activities (Clear and Hardyman, 1990; Byrne, Lurigio, and Baird, 1989; Lurigio, 1987a). At the beginning of 1990, forty states and Washington, D.C., were administering IPS programs (Byrne, 1990).

The structure of the criminal justice system makes it difficult to conduct MSEs of IPS programs. The greatest difficulties stem from the tremendous diversity of the programs, which is fostered by the decentralization and insularity of that system and by its need to respond to diverse local political concerns (Byrne, Lurigio, and Baird, 1989). IPS programs encompass widely varying goals. Some are "front-end" release mechanisms for prison-sentenced offenders who look eligible for less expensive treatment (Erwin, 1987). Others are "back-end" or early release mechanisms for persons who are already in prison but can be let out at low risk in order to free up prison bed space (Pearson and Harper, 1990). Still others are for felony offenders already sentenced to regular probation who score high on a risk scale that assesses their potential to commit future crimes (Byrne and Kelly, 1989).

IPS programs also employ divergent definitions of "intensive" with respect to supervision (Byrne, Lurigio, and Baird, 1989; Byrne, 1990; Peter-silia, 1987). For some programs, intensive only means multiple face-to-face visits with probation officers; for others it means electronically monitored

home confinement, curfew checks, periodic imprisonment, visits to employers and family members by investigators, and the payment of restitution to victims. Caseload sizes can range from twelve per officer to forty or more per officer (Tonry and Will, 1988). Programs that appear similar on the surface may emphasize certain elements of casework practices (for example, treatment) over others (for example, surveillance) and may initiate divergent responses to probation violators (offenders who fail to adhere to the conditions of release) (Byrne, 1990; Tonry and Will, 1988).

Another source of differences among IPS programs is their disparate target populations. The programs can involve a broad range of participants, including violent and nonviolent offenders, those at low and at high risk of reoffending, probation and parole violators, and drug offenders (Byrne, 1986). MSEs of IPS programs also are encumbered by a lack of uniform record-keeping practices, a problem that plagues probation and criminal justice agencies in general. In the case of IPS, record keeping often varies with program goals and functions. Back-end early release programs typically involve surveillance and monitor recidivism, whereas front-end programs typically involve treatment and focus on keeping clients out of trouble.

The problem of "net widening" further complicates MSEs of IPS programs. IPS has become such an attractive option to judges that they are sentencing offenders to IPS who would have ordinarily been placed on routine supervision (Lurigio, 1987b). Judges are very sensitive to the adverse repercussions that may ensue from placing prison-bound or high-risk offenders on regular probation (Petersilia and Turner, 1990). For the judiciary, IPS has much more political currency than does traditional probation. The saturation of IPS programs with candidates for regular probation confounds comparisons between intensive and routine probation programs within the same jurisdiction, and it muddles comparisons between IPS programs in different jurisdictions unless evaluators are knowledgeable about the nature and extent of net widening.

IPS programs also serve hidden or latent goals, which are difficult to evaluate. Tonry (1990) identified three unstated functions of IPS. First, it has helped to bolster probation's flagging public relations image by claiming to be tough on offenders. Second, the adoption of IPS has garnered greater resources and esteem for probation officers, who "get to do probation work the way it ought to be done" and "to work closely with just a few people so [they] can make a difference in their lives" (Pearson, 1987, p. 105). Third, IPS strikes a responsive chord with the public's harsher attitudes toward crime and criminals and it does so while attending to fiscal constraints; hence, IPS is infinitely sellable to both politicians and the public.

In short, IPS programs are highly variable with respect to policies and practices. Differences in program goals, operations, personnel, and target populations result in low construct validity (that is, the "intensive" in IPS is defined differently across sites) and in low external validity (that is, the

findings from one evaluation may have no relevance or applicability to other programs). Disparities between programs make cross-site comparisons unwieldy and uninterpretable. So far, there has only been one MSE of IPS programs that was able to overcome some of these difficulties. It was not only the first multisite IPS study but also the first to employ randomized experimentation.

In 1986, the Bureau of Justice Assistance (BJA) funded several IPS demonstration projects. To be eligible for funding, sites were required to (1) design their programs in accordance with Georgia's IPS program, which involves reduced caseloads, curfews, employment preparation, drug testing, and community service work; (2) accept only adult offenders; (3) participate in training conferences and technical assistance workshops presented by outside experts; and (4) cooperate in a study designed to evaluate the impact of their programs, which meant collecting a core of data and randomly assigning offenders to the program or routine probation. Eleven sites in seven states were selected for participation. BJA funded evaluations of the programs, which were conducted by the RAND Corporation under the stewardship of Joan Petersilia, director of RAND's Criminal Justice Program (Petersilia and Turner, 1990).

Petersilia (1989) provides a highly informative description of the implementation of the evaluation and useful caveats for future researchers. BJA allocated only \$25,000 per site to underwrite the evaluation at each location, which presented serious obstacles to the study; one consequence of the financial constraints was that the programs, independently of each other, had to collect the data and carry out random assignment. RAND researchers were aware that this decision potentially threatened the integrity of the study but concluded that such trade-offs were necessary to undertake the evaluations. In Petersilia's (1989, p. 442) words, "practical realities took precedence over ideal scientific methods."

To bolster their confidence in the soundness of the data and implementation of the research design, RAND staff performed validity checks on the data collected at each site. Petersilia (1989) notes that agency resistance to randomization was overcome primarily by appealing to their self-interest. They were told that random assignment was really the only way to learn about the impact of their efforts and to generate information that they could use persuasively to support the continuation of the program and to establish its credibility. The evaluators also convinced practitioners that random assignment was not capricious but rather an equitable way to make treatment decisions, as there were more eligible offenders than IPS program slots.

To minimize case attrition following random assignment, RAND encouraged the agencies to select offenders late in the process. After reviewing the Federal Judicial Center's (1981) stance on the issue and other legal arguments, the evaluators did not see a compelling obligation to ask clients if they

wanted to participate in the experiment. This eliminated attrition associated with involuntariness. Finally, RAND only allowed overrides to randomization when judges made special requests to place offenders on IPS; Petersilia (1989) reports that there were only a handful of such cases.

Diffusion of treatment is a serious threat to the internal validity of randomized experiments (Cook and Campbell, 1979). Two strategies were employed to avoid the problem of control group contamination. The first involved explaining to officers the importance of keeping IPS and control interventions as divergent as possible, and the second involved disguising the control cases so that officers would not treat randomly assigned offenders any differently than their other clients. However, a preliminary analysis of program records revealed that treatment and control groups were significantly different on several program-defining variables, such as number of contacts, collateral checks, counseling sessions, and drug tests (Petersilia and Turner, 1990).

RAND researchers could not preclude the emergence of practical differences between the various programs, which undermined their comparability, a major problem in MSEs. Participating agencies were given the latitude to develop IPS along lines compatible with their resources, caseloads, and political contexts. This led to considerable variation across the sites in the operational definition of intensive, procedures, and program clients. For example, only some of the programs had special features such as electronic monitoring and on-site testing. Also, some were designed as prison diversion efforts, whereas others focused on high-risk probationers. At some sites, offenders with any violent convictions or sex offense histories were barred from participation, whereas at others they were included. Because of these differences, Petersilia (1989, p. 440) conceded that "in essence, RAND was conducting 11 different evaluations."

Most of the sites overestimated size of their target populations and had trouble meeting subject quotas during the study period. To compensate for the torpid flow of new clients into IPS, agencies began to downgrade their acceptance criteria so that offenders originally excluded from IPS (for example, drug and violent offenders) later became acceptable. At different sites, there were different shifts in eligibility guidelines, which also commonly resulted in operational changes. For example, when caseloads were light, supervision of those involved was stricter. Researchers at RAND had difficulty keeping pace with the rapidly changing nature of the target populations of the programs. This problem was coupled with uneven screening and referral practices. Program staff were responsible for assessing offenders to determine whether they were eligible for IPS. The evaluators found many instances of failure to refer eligible cases, as well as referrals of cases that were ineligible according to the selection criteria employed.

Despite these obstacles, RAND researchers quickly completed their evaluations of three of the eleven sites. In the California IPS program,

random assignment was successful: There were no significant differences between treatment and control groups on demographic measures or indicators of clients' prior records (Petersilia and Turner, 1990). IPS was indeed more intensive than routine probation, as measured by the number and type of contacts and the services provided clients, but IPS offenders had higher rates of technical violations (that is, rule infractions other than new arrests) than did non-IPS offenders. There were no significant differences in rearrest rates between IPS and routine probationers.

Lessons

The community policing and IPS evaluations illustrate a number of the barriers to MSEs erected by the criminal justice system. The decentralization of the system had similar consequences for the community policing and the IPS experiments—although the two sets of programs fell under the same rubric, the actual programs varied widely. Local political, leadership, and resource considerations played as large a role as the program construct in determining what services were actually delivered. IPS was either a substitute for incarceration or a replacement for it, and perhaps it was a punishment for some offenders who otherwise would not have been punished at all. Community policing took on widely divergent targets in different neighborhoods and involved everything from holding meetings in private homes to discussing crime prevention, to cracking down on public drinking by street people.

The fragmented nature of criminal justice policy was reflected in the corrections crisis that brought IPS and other alternatives to incarceration into focus. Legislatures instituted stiffer sentences without funding new prisons; the political push for narcotics enforcement generated millions of new arrests but did not provide judges, jails, or treatment beds. Judges and prosecutors used IPS for varying reasons, producing different client profiles in different jurisdictions. And the lesson of Newark's attempt to coordinate its efforts with that of the school system is a reminder that interagency programs can be very difficult to assemble.

The insular character of the police (in particular) called for expensive technical assistance before they could envision a new way of doing their job. In order to have a program to evaluate, the evaluators had to help organize, educate, and motivate the planning task forces, and to plug them into newly formulated networks of innovative police agencies. The evaluators spent a great deal of time and money ferrying program personnel back and forth between Houston and Newark so that they could reinforce each others' commitment to the program concept and maintain some comparability in projects across the two sites. The evaluators' investment in the resulting programs might have threatened the objectivity with which they were evaluated if the evaluators had not had impartial outsiders monitoring data collection and analyses.

in both cases, the evaluation findings ran afoul of politics. Evidence of success in Houston was sufficiently tainted by suggestions of class and race bias to preclude wide dissemination of results. The community policing MSE could not assess how general this problem of bias might be, however, for political considerations in Newark confined the programs there to poor and homogeneously black neighborhoods. Similarly, RAND released preliminary site-by-site reports of the success of IPS, and there was an uproar because the programs often did not look very successful. Agencies may believe they will be anonymous in an MSE, so in order to avoid this sort of rancor it is important to negotiate in advance the ways in which evaluation data will be analyzed and presented.

New performance measures had to be developed to evaluate the community policing programs, and the data were not the sort that could be collected routinely by any agency. The only data on community policing that were comparable across neighborhoods and cities were gathered in expensive pre- and post-program sample surveys. A change in Houston's computer system meant that even its own crime data collected at the time of the MSE were not comparable to those collected in the immediate past. In addition, the iterative and exploratory nature of the program meant that no one was sure what the appropriate measures of process and outcomes would be until well after the pretest survey had been completed. IPS could utilize inexpensive counts of program failures by the agencies involved, but how failure was defined varied from site to site, and in individual cases the definition often hinged on highly discretionary decisions.

Finally, both MSEs faced civil rights issues. In Houston, any interpretation of the evaluation's findings hinges on the weight that is given to the program's differential benefits by race and class. Perhaps as a result, the project's sponsors never widely published the results of the evaluation and gave short shrift to the benefits issue when they did discuss the project. Independent observers doubted not only the constitutionality of some of Newark's efforts but also the propriety of others (Skolnick and Bayley, 1986). Because the program dealt with individual offenders, IPS faced complex legal and professional issues; case workers were to mete out different punishments using decision rules that took into account neither individuals' offenses nor their treatment needs. This practice did not fit any prevailing correctional ideology, and RAND found it difficult to make the case assignment rules stick. To be sure, the criminal justice system itself mostly gives just lip service to its formal norms, which is a source of great frustration to evaluators when those standards are then presented as barriers to their research. In democratic political systems most evaluations examine competing models of care and assess their benefits for the recipients; but in the criminal justice arena it is often the effectiveness of different forms of pain that is under scrutiny, and it is often the apparent benefit of that pain for other people, not the subjects of treatment, that determines

what shall be done with offenders. Even in their capacity as citizens, evaluators in the criminal justice arena therefore have special responsibilities with regard to how individuals are treated.

References

- Byrne, J. M. "The Control Controversy: A Preliminary Examination of Intensive Probation Supervision Programs in the United States." *Federal Probation*, 1986, 50 (3), 4-16.
- Byrne, J. M. "Assessing What Works in the Adult Community Corrections System." Paper presented at the annual meeting of the Academy of Criminal Justice Sciences, Denver, Colorado, March 1990.
- Byrne, J. M., and Kelly, M. *Restructuring Probation as an Intermediate Sanction: An Evaluation of the Massachusetts Intensive Probation Supervision Program*. Lowell: Center for Criminal Justice Research, University of Massachusetts, 1989.
- Byrne, J. M., Lurigio, A. J., and Baird, S. C. "The Effectiveness of the New Intensive Supervision Programs." *Research in Corrections*, 1989, 2 (3), 1-48.
- Clear, T. R., and Hardyman, P. L. "The New Intensive Supervision Movement." *Crime and Delinquency*, 1990, 36 (1), 42-61.
- Cook, T. D., and Campbell, D. T. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin, 1979.
- Erwin, B. S. *Evaluation of Intensive Probation Supervision in Georgia*. Atlanta: Georgia Department of Corrections, 1987.
- Federal Judicial Center. *Experimentation in the Law: Report of the Federal Judicial Center Advisory Committee on Experimentation in the Law*. Washington, D.C.: Federal Judicial Center, 1981.
- Hillsman, S. I. "Fines and Day Fines." In M. Tonry and N. Morris (eds.), *Crime and Justice: A Review of Research*. Chicago: University of Chicago Press, 1990.
- Lavrakas, P. J., Rosenbaum, D. P., and Lurigio, A. J. "Media Cooperation with the Police: The Case of Crime Stoppers." In R. Surette (ed.), *The Media and Criminal Justice Policy*. Springfield, Ill.: Thomas, 1990.
- Lily, J. R., Ball, R. A., and Wright, J. "Home Incarceration with Electronic Monitoring in Kenon County, Kentucky: An Evaluation." In B. R. McCarthy (ed.), *Intermediate Punishments: Intensive Supervision, Home Confinement, and Electronic Surveillance*. Monsey, N.Y.: Criminal Justice Press, 1987.
- Lurigio, A. J. "Evaluating Intensive Probation Supervision." *Perspectives*, 1987a, 11 (1), 17-19.
- Lurigio, A. J. "The Perceptions and Attitudes of Judges and Attorneys Toward Intensive Probation Supervision." *Federal Probation*, 1987b, 51 (4), 16-29.
- Lurigio, A. J., and Skogan, W. G. "Doing Policy Relevant Research in Criminal Justice Settings: A Tough Road to Hoe." Paper presented at the Illinois Criminal Justice Information Authority's Trends and Issues Conference, Chicago, July 1990.
- McCarthy, B. R. (ed.). *Intermediate Punishments: Intensive Supervision, Home Confinement, and Electronic Surveillance*. Monsey, N.Y.: Criminal Justice Press, 1987.
- MacKenzie, D. L., and Shaw, J. W. "Inmate Adjustment and Change During Shock Incarceration: The Impact of Correctional Boot Camp Programs." *Justice Quarterly*, 1990, 7 (2), 125-150.
- Pearson, F. S. *Research on New Jersey's Intensive Supervision Program*. New Brunswick, N.J.: Dept. of Sociology, Rutgers University, 1987.
- Pearson, F. S., and Harper, A. G. "Contingent Intermediate Sentences: New Jersey's Intensive Supervision Program." *Crime and Delinquency*, 1990, 36 (1), 75-86.

- Petersilia, J. M. *Expanding Options for Criminal Sentencing*. Santa Monica, Calif.: RAND Corporation, 1987.
- Petersilia, J. M. "Implementing Randomized Experiments: Lesson from BJA's Intensive Supervision Project." *Evaluation Review*, 1989, 13 (3), 435-458.
- Petersilia, J. M., and Turner, S. "Comparing Intensive and Regular Supervision for High-Risk Probationers: Early Results from an Experiment in California." *Crime and Delinquency*, 1990, 36 (1), 87-111.
- Skogan, W. G. *Disorder and Decline: Crime and the Spiral of Decay in American Cities*. New York: Free Press, 1990.
- Skolnick, J., and Bayley, D. *The New Blue Line*. New York: Free Press, 1986.
- Tony, M. "Stated and Latent Features of IPS." *Crime and Delinquency*, 1990, 36 (1), 174-191.
- Tony, M., and Will, R. *Intermediate Sanctions*. Castine, Maine: Castine Research Corporation, 1988.

Wesley G. Skogan is professor of political science and a member of the research faculty at the Center for Urban Affairs and Policy Research at Northwestern University, Evanston, Illinois.

Arthur J. Lurigio is a social psychologist. He is currently assistant professor of criminal justice at Loyola University Chicago and research associate at the Center for Urban Affairs and Policy Research, Northwestern University.