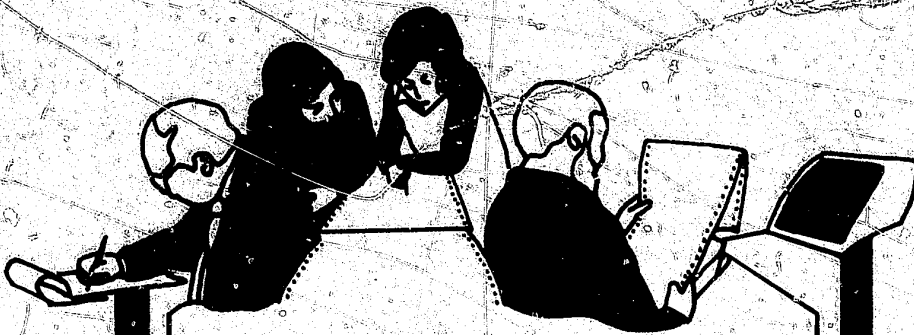


SNI



**HOW WELL
DOES IT WORK?
Review of
Criminal Justice
Evaluation,
1978**

64112

**U. S. Department of Justice
Law Enforcement Assistance Administration
National Institute of Law Enforcement and Criminal Justice**



Prepared for the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice by Aspen Systems Corp., under contract number J-LEAA-023-77.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock Number G77-000-0082-8

Due out of Acc.
Acc. 4-16-80

RUSH

SNI

**HOW WELL
DOES IT WORK?**

**Review of
Criminal Justice
Evaluation,
1978**

Commissioned and compiled by
National Criminal Justice Reference Service

June 1979

U. S. Department of Justice
Law Enforcement Assistance Administration
National Institute of Law Enforcement and Criminal Justice



Law Enforcement Assistance Administration
Henry S. Dogin, Administrator
Homer F. Broome, Jr., Deputy Administrator for Administration

National Institute of Law Enforcement
and Criminal Justice
Harry M. Bratt, Acting Director

Among contributors to this volume are the following:

Authors:

Lawrence Bennett
Michael F. Cahn,
Edward H. Kaplan,
and John G. Peters, Jr.
Albert P. Cardarelli
and Charles P. Smith
William C. Cunningham
and Todd H. Taylor
Malcolm M. Feeley
Roger B. Parks
Lee Sechrest
and Robin Redner
Wesley G. Skogan
James M. Tien
and Kent M. Colton
Robert K. Yin

NCJRS Staff:

Georgette Semick,
Program Director
Justus Freimund, Manager,
Criminal Justice Services
Joe Holt Anderson,
Coordinating Editor
David M. Horton,
Mark Shanley, and
Roy W. Reese,
Reference Specialists
Susan E. Carpenter,
Design Artist
Veronica Martin,
Administrative Assistant

Editorial Review Board:

Ilene Bernstein,
Indiana University
Ward Edwards,
University of
Southern California
Robert L. Emrich,
Pacific Institute for
Research and Evaluation
Don M. Gottfredson,
Rutgers University
Richard C. Larson,
Massachusetts Institute
of Technology

LEAA Representatives:

Morton Goren
William Heenan
Paul Lineberry

Other Reviewers

Pierce R. Brooks
Jan Chaiken
Charles M. Friel
Jay David Jameson
Glen A. Kercher
George G. Killinger
Ben A. Menke
Merlyn D. Moore
Gregory Riede
Frank Schmalleger
Sam Souryal

Articles in this Review are individually copyrighted by their authors. All rights are reserved to the Law Enforcement Assistance Administration or to the authors. For information, address Program Director, National Criminal Justice Reference Service, P.O. Box 6000, Rockville, MD 20850.

Opinions expressed are those of the authors and do not necessarily reflect those of the Law Enforcement Assistance Administration, U.S. Department of Justice.

TABLE OF CONTENTS

PART I	The World of Criminal Justice Evaluation	1
PART II	A Review of Evaluation Results	17
	Corrections	
	Lee Sechrest and Robin Redner: Strength and Integrity of Treatments in Evaluation Studies	19
	Lawrence A. Bennett: Probation, Parole and Correctional Programs--Review of Evaluation	63
	Community Crime Prevention	
	Robert K. Yin: What Is Citizen Crime Prevention?.. . . .	107
	Wesley G. Skogan: Community Crime Prevention Programs--Measurement Issues in Evaluation	135
	Courts	
	Malcolm M. Feeley: The Impact of Innovation	171
	Police	
	Michael F. Cahn, Edward H. Kaplan, and John G. Peters, Jr.: Police Field Studies Evaluation.	205
	Roger B. Parks: Police Reorganization	237
	William C. Cunningham and Todd H. Taylor: Police Personnel and Training Evaluation	257
	James M. Tien and Kent W. Colton: Police Command, Control, and Communications.	293
	Juvenile Delinquency	
	Albert P. Cardarelli and Charles P. Smith: Delinquency Prevention and Control Program	337
PART III	Where Do We Go From Here?	363

**PART I:
HOW WELL DOES IT WORK?**

**The World of
Criminal Justice Evaluation**

"Does it work?" and "How well does it work?" These are the defining questions for evaluators in any discipline, especially for those in such a fast-developing, relatively new field as that of the evaluation of programs in law enforcement and criminal justice. And answers to these questions are essential for the policymakers who write the budgets for criminal justice practitioners.

A relatively new field is criminal justice evaluation. Records of arrests, more or less accurate, date back to the 18th century beginnings of modern-day police forces, and studies of criminal behavior date almost to the 16th century beginning of the quantification of behavioral sciences data. Nevertheless, one would not far miss the mark by dating the modern-day concern about evaluating law enforcement and criminal justice programs to Congress' creation of the Law Enforcement Assistance Administration (LEAA) under the Department of Justice in 1968. (This statement is not intended to downgrade the evocative work of the Wickersham Commission in 1931 or the Katzenbach Commission and D. C. Crime Commission of 1965, or other earlier advances. However, as one of our authors notes, the creation of LEAA nationalized the war against crime.)

In successive legislation since 1968, Congress has continually increased its emphasis on the importance of evaluating the relative efficiency of law enforcement programs, new and old. The National Institute of Law Enforcement and Criminal Justice (NILECJ), as the main research, development, and evaluation arm of LEAA, has operated the National Evaluation Program (NEP) since 1974. Once the evaluated program is shown to be of merit, it can be replicated in other jurisdictions through the assistance of NILECJ's Office of Development, Testing and Dissemination (ODTD).

For scholars, practitioners, students, Congress and other policymakers, and interested laymen, NILECJ is publishing Review of Criminal Justice Evaluation, 1978. It is only the first of a series and does not pretend to be exhaustive of the far-reaching topic of criminal justice evaluation. It does partially fulfill NILECJ's responsibility to report annually to the President, Congress, State and local officials, and the public on the state of the art in criminal justice practice.

All the members of the audience discussed above have an interest in finding answers to the basic question, "What works in criminal justice?" The challenge is that of finding answers which will improve the results of criminal justice programs; answers which, when a program works in Iowa, show its likelihood of working in Delaware or Nevada.

However, the Review is a thick book. On many pages, it is written in the vocabulary of statisticians and social scientists, a particularized and necessary technical language with which

scholars and practitioners of other disciplines are not necessarily familiar. Indeed, it is axiomatic that those who make policy in criminal justice, such as legislators and council members, come from other disciplines. Therefore this summary is provided not only to explore briefly evaluation work of the National Institute but to summarize quickly the findings of the 16 distinguished scholars and practitioners who are the authors of the 10 chapters of Part II of the Review. Finally we will take a fast look at the findings of a survey of criminal justice researchers and evaluators, a survey which is more fully reported in the Review's Part III.

A REVIEW OF RESULTS IN FIVE FIELDS

The papers in Part II may be roughly divided among the fields of Corrections, Community Crime Prevention, Courts, Police, and Juvenile Delinquency Control evaluations. The authors are not always in agreement with each other, but their views and findings are presented, with only limited editing and no censorship, as the views of persons highly regarded in their fields.

Corrections

Lee Sechrest and Robin Redner explore existing studies of "treatments," or programs to reduce criminal behavior through the corrections system. They find that most such studies fail to measure either the "strength" of the treatment, how intensive it was, or the "integrity," how consistently the planned treatment was administered.

Lawrence Bennett indicates that present research into probation, parole, and determinate sentencing offers no panacea for recidivism, but does offer hope for a lower "failure" rate, increased efficiency, and a better opportunity for reintegration of past offenders within their communities.

Community Crime Prevention

Robert K. Yin argues that the efficacy of citizen crime prevention programs has been vastly underrated, that they are a primary tool in crime prevention, and that the lack of effective evaluation of their results is the reason they are not more highly regarded.

Wesley Skogan suggests that one should think of evaluations of community crime prevention programs as a contest between a program and the noise created by inadequate measures. "Programs appear to be 'winners' only when they can outshoot the opposition"; that is, when their effects are so pronounced as to be detectable in spite of deficiencies of measurement.

Courts

Malcolm M. Feeley discusses innovations in court procedures, organization, sentencing, and the like. Although he finds few scientific evaluations of these innovations, he cites many less formal assessments of them as indicative of certain apparent trends.

Police

Michael F. Cahn, Edward H. Kaplan, and John G. Peters, Jr., explore experimental findings on techniques of police patrol and other field activities.

Roger Parks finds little evaluative evidence of the efficacy of recent police reorganizations, particularly those in which smaller forces merge into larger units. He suggests that police mergers are not necessarily at fault, but that the paucity of adequate evaluation leaves doubt.

William C. Cunningham and Todd H. Taylor look at police recruitment and training programs and sharply criticize existing support programs and the associated evaluation of them.

James M. Tien and Kent W. Colton explain many sophisticated new computer-aided types of police control and communications, and discuss the lack of effective evaluation of most of them.

Juvenile Delinquency

Albert P. Cardarelli and Charles P. Smith investigate strategies toward delinquency control and prevention; they conclude that "Linkages between the school, the world of work, and the family need to be encouraged and strengthened."

Conclusions

Thus our Part II authors seem in consensus that, despite progress, improving the state of the art in assessing criminal justice studies will, in Bennett's words, "keep evaluators busy for some time to come."

WHICH QUESTIONS TO ANSWER?

Most criminal justice evaluation conducted thus far answers only the first of four evaluation questions which may be posed as follows:

- Are the programs working?
- Are the programs producing the desired results?

- Are the results of programs commensurate with the funds they consume?
- Are there better ways to attack and solve society's problems? In addition, criminal justice evaluation to date has focused on the first two of the following four areas which it might be expected to measure:
 - **Input:** program resources, responsibilities, and constraints.
 - **Process:** program implementation and operation, taking the effects of concurrent programs into account.
 - **Outcome:** program impacts, e.g., changes effected in attitudes, behavior, or criminal activity.
 - **Systemic Impact:** the program's impact as viewed from a total systems perspective, e.g., generalizability.

Tien writes:

In general, the input and process measures serve to 'explain' the resultant outcome measures. Input measures alone are of limited usefulness since they only indicate a program's potential--not actual--performance. On the other hand the process measures do identify the program's performance but do not consider the impact of that performance. Finally, the outcome measures are the most meaningful observations since they reflect the ultimate results of the program. In practice, as might be expected, most of the available evaluations are fairly explicit about the input measures, less explicit about the process measures, and somewhat fragmentary about the outcome measures. The fourth set of evaluation measures--the systemic measures--can also be regarded as impact measures but have been overlooked to a large extent in the evaluation literature. The systemic measures allow the program's impact to be viewed from a total systems perspective.*

* James M. Tien, "Toward a Systematic Approach to Program Evaluation Design," Special Issue of the IEEE Transaction on Systems, Man and Cybernetics entitled Public Systems Methodology, September, 1979. This work was partially supported by the National Institute of Law Enforcement and Criminal Justice, Grant No. 77-NI-99-0058.

The National Institute (NILECJ) has attempted, increasingly, to fund evaluations which would view the relative merit and cost benefit of projects and programs from a total criminal justice system perspective.

From its inception, the Institute has funded a number of major evaluations or research projects with an evaluation component. In 1973 Congress directed the Institute to evaluate, where possible, criminal justice assistance programs supported by the Law Enforcement Assistance Administration, including those funded through the block grant program. Today, two offices within the National Institute are responsible for carrying out various aspects of the Institute's evaluation effort. In addition, the Office of Research Programs also supports evaluative research. For example, the Office funded an evaluation of the LEAA restitution program implemented in several jurisdictions throughout the country.

The Office of Program Evaluation focuses upon assessing a limited number of specific criminal justice programs and innovations supported at the Federal, State and local levels. Although State and local criminal justice initiatives, such as the New York Drug Law, Massachusetts Gun Law, and Vera Evaluation of Court Employment Program, have been evaluated with NILECJ support, the major portion of this office's work focuses on programs created and supported with LEAA funds.

The Office of Research and Evaluation Methods focuses on development of new and more effective methods and more standardized data collection and maintenance techniques. The Institute also helps State and local evaluators improve their capabilities by continual assessment, improvement, and dissemination of the best available information on the state of the art.

OFFICE OF PROGRAM EVALUATION

The Office of Program Evaluation designs, funds, and administers evaluations of national-level LEAA programs, innovative and experimental State and local projects, criminal justice techniques and procedures, and significant State and local legislative and administrative reforms. The Office is responsible for the National Evaluation Program, Demonstration Test Evaluations, Field Test Evaluations, and other nationally identified evaluations. It also supports the development of evaluation guides and handbooks for State and local use.

National Evaluation Program

Created as an efficient mechanism for evaluating the vast number of LEAA block grant programs, the National Evaluation Program has sponsored a series of phased evaluations in 38

"topic areas" (see Figure 1). Each topic area includes ongoing criminal justice projects or functions with similar objectives and strategies, such as Employment Services for Ex-offenders, Street Lighting Projects or Family Counseling. In this way, it is possible to gauge the impact of broad categories of programs supported by LEAA.

The initial step is a "Phase I" study that identifies key issues and measurement points, assesses what is currently known about them, and outlines approaches or methodological needs for more intensive national or local evaluation. Each Phase I results in an assessment of a project type or function area based on available data, site visit documentation, and limited pre-testing of possible designs for more intensive Phase II evaluation. A Summary report from each NEP study is widely disseminated in order to provide feedback to a variety of criminal justice planners and evaluators, especially at State and local levels. To date, 27 Phase I studies have been completed on topics ranging from Juvenile Diversion to Street Lighting Projects.

Phase II programs funded include Pretrial Release and Treatment Alternatives to Street Crime (TASC). A summary report of the TASC Evaluation is available from NCJRS. A Phase II evaluation of probation is undertaking a controlled experiment to compare the effectiveness of different probation techniques in order to determine effective levels of supervision and the utility of different treatments for different types of offenders. A description of some general NEP findings follows this summary of additional NILECJ evaluation activities.

Demonstration Program Evaluations

LEAA supports a variety of national programs designed to demonstrate the effectiveness of various concepts and methods of reducing crime and improving criminal justice. The Office of Program Evaluation supplies funds for evaluations of a select number of these programs each year. One effort currently underway is an evaluation of the LEAA Community Anticrime Program, which has awarded approximately \$30 million to support community organizations operating independently of State and local governments and agencies. A related effort involves the evaluation of another major community initiative launched by LEAA in 1978, the Comprehensive Areawide Crime Prevention Program. This program focuses on the coordinated efforts of various agencies and community groups. Other evaluations undertaken include the LEAA Career Criminal Program, the Standards and Goals Program, Statistical Analysis Centers, Major Criminal Conspiracies, and Anti-fencing (Operation Sting) Programs.

Field Test Evaluations

Another major program is conducted in conjunction with the National Institute's Office of Development, Testing, and Dissemi-

nation. ODTD designs experimental programs based on research findings and implements them in a limited number of sites on a test basis. The field tests are designed to identify needs for future research and to experiment with promising programs that may warrant wider demonstration by LEAA.

Evaluation begins with the commencement of testing, thus ensuring thorough and comprehensive assessment of the program's experience. Currently, the Office of Program Evaluation is evaluating the field test of the LEAA Neighborhood Justice Center Program, which uses dispute resolution techniques to divert certain types of criminal and civil cases from the courts. Other tests under evaluation include Managing Patrol Operations, Pre-release Centers, Multicounty Sentencing Guidelines, Neighborhood Action to Reduce Burglary, and Commercial Security.

Other Evaluations

Many innovations originate with State and local governments. Where these have national implications, the Institute supports examination of their effects. Most of these projects are managed by the Office of Program Evaluation, although other Institute offices also fund evaluative studies.

One recently completed evaluation assessed the impact of the New York State's revised drug laws, enacted in 1973. Overall, the evaluators found that the new legislation had not achieved its objectives in its first 3 years: heroin use, for example, was as widespread in New York City in 1976 as it had been in 1973, when the "get-tough" laws were passed. Neither was there any significant impact on the patterns of drug abuse or drug-related crimes in New York City, as compared with other large East Coast cities. Court backlogs and other effects on the criminal justice system did tend to decrease over the 3-year period, but even in this respect the new legislation does not appear to have been costeffective. A final report on this study was published under the title of "The Nation's Toughest Drug Law: Evaluating the New York Experience." The researchers also presented their findings in testimony before the New York State Standing Committee on Codes, which considered changes in the State's drug laws. They also testified before legislative committees in Pennsylvania and Michigan, which were considering mandatory sentencing laws and increased penalties for drug abuse.

Another study is assessing Alaska's experience with abolition of plea bargaining. The evaluation is still in progress, but the misdemeanor portion has uncovered some striking and unexpected findings. The researchers found that, on the average, misdemeanor defendants who actually went to trial under the new policy received dramatically longer sentences than those who pleaded guilty to similar charges. Before the new policy was instituted, no difference was detected in the sentencing for these

NEP PHASE I ASSESSMENTS

September, 1979

TOPIC AREA	FUNDING STATUS				PROGRESS STATUS		PUBLICATION STATUS
	UNDER CONSIDERATION	AWARDED	ACTIVE	COMPLETED	REPORTS UNDER REVIEW	FULL REPORTS AVAILABLE FROM NCJRS *	SUMMARY PUBLISHED
FY 1975							
Operation Identification Projects		X		X		X	X
Treatment Alternatives to Street Crime (TASC)		X		X		X	X
Pretrial Screening Projects		X		X		X	X
Select Patrol Strategies: Specialized Patrol Operations		X		X		X	X
Early Warning Robbery Reduction Projects		X		X		X	X
Citizen Crime Reporting Programs		X		X		X	X
Pretrial Release Programs		X		X		X	X
Prevention of Juvenile Delinquency		X		X		X	
Alternatives to Juvenile Incarceration		X		X		X	X
Juvenile Diversion		X		X		X	X
Traditional Preventive Patrol		X		X		X	X
Youth Service Bureaus		X		X		X	+
Team Policing Projects		X		X		X	X
Citizen Patrol Projects		X		X		X	X
Patrol Support Systems: Crime Analysis Units		X		X		X	X
Detention of Juveniles and Alternatives to Its Use		X		X			X
Security Survey: Community Crime Prevention Programs		X		X		X	X
FY 1976							
Residential Inmate Aftercare Projects (Halfway Houses) for Adult Offenders		X		X		X	X
Court Information Systems		X		X		X	X
Institutional Furlough Programs		X		X		X	+
Intensive Special Probation		X		X		X	X
Employment Services for Releases in the Community		X		X		X	X
Street Lighting Projects		X		X		X	X
Policing Urban Mass Transit Systems		X		X		X	
Institutional Education Programs for Inmates		X		X		X	X
FY 1977							
Police Juvenile Units		X		X		X	
Coeducational Correctional Institutions		X		X		X	X

*Full reports can be obtained on a loan basis by writing the National Criminal Justice Reference Service, Box 8000, Rockville, Maryland 20850, Attention: Loan Librarian.

+ Available on a loan basis only.

TOPIC AREA	FUNDING STATUS			PROGRESS STATUS		PUBLICATION STATUS	
	UNDER CONSIDERATION	AWARDED	ACTIVE	COMPLETED	REPORTS UNDER REVIEW	FULL REPORTS AVAILABLE FROM NCJRS*	SUMMARY PUBLISHED
FY 1978							
PROMIS Projects		X	X				
Correctional Data Systems		X	X				
Police Training Programs		X	X				
Victim/Witness Assistance Projects		X	X				
Correctional Personnel Training Programs		X	X				
Shoplifting and Employee Theft Programs		X	X				
Restitution Programs		X	X				
Police Command and Control Programs		X	X				
FY 1979							
Screening and Evaluation for Mental Health Services		X	X				
Family Counseling		X	X				
Police Liaison Activities		X	X				
FY 1980							
Minority Employment	X						
State and Local Utilization of Evaluation Information	X						
Police Service Aides	X						

*Full reports can be obtained on a loan basis by writing the National Criminal Justice Reference Service, Box 8000, Rockville, Maryland 20850, Attention: Loan Librarian.

*Available on a loan basis only.

groups. Further study will now examine several possibilities: (1) that implicit plea bargaining still exists in Alaska, despite the perceptions of courts personnel; or (2) that under the new policy judges may be penalizing misdemeanants who exercise their right to trial; or (3) that defendants with more serious records are now going to trial.

The effects of the Massachusetts Gun Law are also being studied. Using stochastic modeling techniques, the researchers found evidence that the gun law has reduced the rate of gun-related crimes in Boston. In the case of armed robberies the Gun Control Law appears to have had a lasting effect. The armed robbery rate began to decline in February of 1975. Two years later it had stabilized on a lower trend-line.

A major assessment of governmental responses to crime during the period 1968-1978 is investigating criminal justice, social, and economic responses to the problems of rising crime. The study is examining 10 tri-city sites across the country and will reassess major LEAA efforts, such as the Impact Cities Program and the Environmental Design Program in these areas.

OFFICE OF RESEARCH AND EVALUATION METHODS

The Office of Research and Evaluation Methods supports projects that explore the methodological and measurement problems facing criminal justice researchers and evaluators. The research usually entails the development or adaptation of advanced analytical techniques to problems in crime analysis or crime control initiatives. Emphasis is placed on those measurement and methodological and measurements problems facing criminal justice researchers and evaluators. The research usually entails the development or adaptation of advanced analytical techniques to problems in crime analysis or crime control initiatives. Emphasis is placed on those measurement and methodological issues that recur in a wide variety of criminal justice settings so that approaches developed can be extended to system-wide research and evaluation.

This Office also administers two of the Institute's long range priorities--Deterrence and Criminal Justice Performance Standards and Measures--and has mounted major programs in these areas.

Deterrence

General deterrence is one of the most fundamental and controversial concepts underlying the design of crime control policy. It is based on the idea that the risk of suffering undesired consequences for a contemplated criminal act has a crime-suppressive effect on the potential offender. Which sanctioning policies offer the best promise of deterrence have long been the subject

of public debate. By 1975, general deterrence had also generated controversy among scientists as seemingly contradictory findings were reported in the literature. Available evidence thus far provides little useful guidance.

To clarify the issues involved, NILECJ commissioned the National Academy of Sciences to make a detailed technical inquiry into the analytic merits and shortcomings of research on deterrence and incapacitation. In its report, the NAS panel concluded that it "could not yet assert that the evidence warrants an affirmative conclusion regarding deterrence." But the panel noted that its "reluctance to draw a stronger conclusion does not imply that deterrence does not exist, since the evidence certainly favors a proposition supporting deterrence more than... one asserting that deterrence is absent." The major challenge for future research, the Panel said, is to "estimate the magnitude of the effects of different sanctions on various crime types, an issue on which none of the evidence available thus far provides very useful guidance."

Therefore, the National Institute began a research program on the "Theory of General Deterrence." The projects funded fall under two broad headings:

- Modeling Advances--In constructing conceptual models of how sanctions affect human behavior, researchers inevitably make a host of tacit and often complex assumptions to arrive at conclusions about whether a particular sanction indeed caused a reduction in crime. Projects in this area are examining how inferences of a deterrence effect are drawn from analysis and how objectivity might be enhanced.
- Quasiexperiments--The deterrence hypothesis suggests that a credible and significant change in sanctioning level would produce a consequent and oppositely directed change in crime rates. (The more severe the penalty, the lower the level of crime, and vice versa). The validity of this theory can be tested in "quasiexperimental" settings such as a well-advertised change in penalties or police crackdown on certain classes of offenders. Projects in this category are investigating the effects of what happens in real-life situations when criminal justice sanctioning policy is changed.

Performance Standards and Measures

Research in this priority is concerned with the development and validation of performance measures based on an overall theory or set of theories about the purposes of the criminal justice system and its components. The program emphasizes cumulative knowledge-building over an 8-to 10-year period, starting in the first phase with articulating better definitions of the issues

and problems in performance measures and identification of future research needs.

During FY 1978, funds were awarded to create a consortium of five research institutions to develop a coherent conceptualization of performance in the criminal justice system and its subsystems. They are attempting to elucidate the many-faceted, and sometimes conflicting, nature of the measures used to assess the performance of specific parts of the system. For example, a police performance measurement system that unduly rewards sheer numbers of arrests affects the performance of the prosecutor's office which must deal with the large number of cases police turn over to them. By breaking down the generic "performance" into separate aspects, by exploring the influences of organizational style on these aspects, and by providing a cohesive measurement context, the consortium should produce a platform for further empirical research.

The Office is also studying the state of the art in State and local evaluation under its Operational Performance Measurement Systems Program. This effort complements the longer-range theory-based program by studying existing exemplary processes initiated at the State and local levels and seeks to transfer these concepts to other local, as well as Federal, programs.

Methodology Research

The Office of Research and Evaluation Methods also sponsors a modest program of support for a range of studies of high technical merit aimed at research, development and testing of methodological innovations of potential significance to criminal justice.

Some studies are investigating the validity of crucial assumptions that underlie certain commonly-used elements of criminal justice evaluation design. Some explore sophisticated techniques that have demonstrated their utility in other social research and evaluation to determine their applicability to criminal justice research. Finally, some of the projects seek a theoretical breakthrough on critical measurement and methodological issues.

Other Research

The Office of Research and Evaluation Methods is also managing a project stemming from the National Academy of Sciences Panel on Research in Rehabilitative Techniques. The effort is examining successful treatment programs used in fields such as education and mental health and assessing the feasibility of transferring the approaches to a correctional setting.

NEP GENERAL FINDINGS

The 1973 Crime Control Act directed NILECJ to evaluate LEAA-funded projects wherever possible and to share the findings with State and local officials and planners. Given the large number of projects, full-scale evaluations of each would have been enormously expensive, far beyond the Institute's annual budget. The apparent alternative was to fund a limited number of expensive evaluations each year, with the payoff far in the future. Thus the National Evaluation Program evolved.

A series of phased studies was devised to collect relevant information in an orderly sequence. This would lower the initial cost of evaluation and increase the number of programs that could be examined. Equally important, it would accelerate the process of providing valuable information to policymakers.

Phase I assessments have been conducted without extensive data collection and analysis by reviewing completed evaluations of the projects being studied and by conducting a limited number of site visits. Phase I assessments give the Institute a sound basis for deciding whether intensive evaluation is warranted. Where appropriate, the design for intensive evaluation is implemented as a Phase II evaluation. Where the available data and prior research are insufficient for assessing potential effects, the Phase I field work lays the groundwork for development of strategies for obtaining the necessary data. Included are definition of data requirements, recommendation of measurement techniques, and identification of measurement points.

The two-phase approach also allowed NILECJ to draw some general conclusions about the state of criminal justice evaluation. The NEP assessments revealed a compelling need for a clearer understanding of evaluation and for training in evaluation methods. Frequently, NEP evaluators find that many researchers are unable to design or conduct field evaluations. In addition, policymakers often have difficulty in understanding the tasks involved in such evaluations and the confidence they can place in evaluation findings. Another general finding of the program is that the availability and usefulness of data varies greatly. Although relevant information does exist in many cases, it often is not satisfactorily related to performance and usually does not lend itself to comparative analyses.

An assessment of pre-trial screening, for example, found abundant data. However, much of the data on case dispositions were misinterpreted simply because the prosecutor's charging policy was unknown or misunderstood.

When performance measures are selected locally or nationally without a valid evaluation framework, meaning and interpretation will necessarily vary from project to project, making overall

comparisons impossible. The results of the Phase I studies should help to create a foundation for developing standard measurement models.

A number of the assessments have uncovered discrepancies between what the project was theoretically designed to do and what it actually does. Juvenile diversion projects, for example, were created to reduce the flow of juveniles through the juvenile justice system but, in some cases, they have the opposite effect. Instead of "softening the blow" these projects often bring many more juveniles into contact with the criminal justice system.

In cases such as these, the Phase I study can pinpoint gaps between policy and implementation. In 1977, the Phase I assessments were restructured to include a limited pretest of the Phase II evaluation design in a small number of test sites. The results are then submitted along with any necessary revision of the Phase II design. The pretest gives more conclusive results at the Phase I stage, a better estimate of the feasibility of the Phase II evaluation, and an opportunity to improve the Phase II design. The field tests are designed to identify needs for future research and to experiment with promising programs that may warrant wider demonstration by LEAA.

WHERE DO WE GO FROM HERE?

In the final steps of preparing the Review, the editors of the National Criminal Justice Reference Service polled a small but distinguished sample of scholars and practitioners as to where we are and where we must go next in criminal justice evaluation. Part III details their answers to these four questions:

- o What are the most important evaluation findings?
- o What questions should evaluators next answer?
- o What are the critical issues in criminal justice evaluation?
- o What are the dangerous pitfalls in evaluation?

As earlier noted, Part III also discusses coming needs for expansion of the evaluators' art. We hope those who have read thus far will want to finish the whole volume which we have to this point summarized.

Joe Holt Anderson,
Technical Editor
David M. Horton,
Senior Reference Specialist
Georgette Semick,
Program Director

National Criminal Justice Reference Service

**PART II:
HOW WELL DOES IT WORK?**

**A Review of
Evaluation Results**

Strength and Integrity of Treatments in Evaluation Studies¹

Lee Sechrest is professor of psychology at The Florida State University and a frequent author and editor in social psychology and evaluation methodology.

Robin Redner is a doctoral candidate at Florida State. She is a research assistant to the National Academy of Sciences on Rehabilitation of Criminal Offenders.

INTRODUCTION

When interpreting the results of an evaluation study, the question whether the study was a valid one on which to base the interpretations is of paramount importance. If the validity of the study is suspect, any interpretations will be unwarranted and potentially misleading. It is now widely recognized that there are at least four types of validity which a study must possess if it is to be interpreted properly and completely although partial interpretations may be accurate and satisfactory in some circumstances. The analysis which follows will adapt, in only an approximate way, the formulations of the four types of validity derived by Cook and Campbell,² since there is room for some disagreements about just what type of validity is jeopardized by which problem.

Failures of validity of experiments may lead either to an erroneous conclusion that a treatment is effective when it is not, or to the erroneous conclusion that a treatment is not effective when, in fact, it is. These erroneous assertions are also commonly known as Type I and Type II errors. A Type I error is one in which a result of a treatment is characterized as an experimental effect when, in fact, it occurred by chance alone; a Type II error is one in which valid experimental effect is ignored because it appears to be insignificant--the result of chance alone. Historically, the prospect of making a Type I error rather than a Type II error has been more carefully guarded against by the use of a conventional level of significance (variation equals 0.05).

However, it should be noted that the usually forgotten Type II error is also important especially in applied research where human welfare is more directly affected by research results. Science is probably not much impeded if an occasional Type II error is made, but considerable social harm may result if effective treatments are erroneously abandoned as worthless.

Internal validity is the sine qua non for interpreting the results of an evaluation study, and refers to the confidence with which one can assert that the outcome of an experiment is attributable to the intervention (treatment) and to no other variable. Obviously, internal validity may be thought to involve a continuum

Copyright (c) 1979 Lee Sechrest and Robin Redner.

so that conclusions may be reached with more or less confidence, and thresholds for levels of confidence may differ from one researcher or policymaker to another. Still, one must with some confidence be able to assert that the experimental treatment produced an effect before one may legitimately proceed to other issues.

Conclusion validity refers to what are primarily statistical issues relating to the adequacy of the analyses to detect effects which have occurred. As for internal validity, a lack of conclusion validity may lead either to the erroneous assertion that the experimental treatment had an effect or to the equally erroneous assertion that it did not. Common sources of problems with conclusion validity include: (1) sample sizes too small to permit detection of effects of the size intended; (2) inefficient statistical tests that reduce the likelihood of detecting effects for the sample size employed; (3) multiple comparison without correction for the probability of obtaining some significant values by chance alone; (4) careless experimentation resulting in inflation of error terms used for statistical tests, again with the result of reducing the likelihood of detecting effects; (5) use of unreliable outcome measures, with the result that effects may not be detected; (6) poor quality control over data management and processing so that errors may be introduced; (7) outright errors in statistical calculations, which may produce biases of any nature; and (8) erroneous interpretations of statistical findings.

The very best statistical work cannot compensate for a lack of internal validity in an experiment, but even the best experiment can be spoiled by poor statistical considerations or analysis.

The third type of validity, the focus of this paper, is construct validity, referring to the adequacy of our understanding of the nature of the experimental variable(s) being studied. Specifically, the experimental variable(s) must be clearly and completely defined in an operational sense as well as adequately representing the theoretical construct on which based. For instance, the results of any study may be misinterpreted by the original experimenter or subsequent investigators because the experimental treatment is assumed to be something it is not. Some of the reasons a treatment may be misunderstood include: (1) confounding of the intended independent variable with an associated variable, e.g., ignoring the fact that the experimental treatment is delivered by better trained personnel than is the comparison treatment; (2) nonspecific treatment effects such as expectancy or placebo effects, which may be mistaken for experimental treatment effects; (3) adequacy of description of the experimental or comparison treatment, which may determine whether the nature of the treatment is well understood; and (4) adequacy of theoretical formulation of the treatment, which may determine whether generalizations of the treatment are well-founded. There are many other possible factors involved in construct validity, two of which, strength and integrity of treatment, will be addressed at length in this paper.

The final type of validity mentioned in the context of interpreting results of experiments is external validity, referring to the legitimacy of various generalizations of research findings. Even if an experiment is well-designed and implemented, properly analyzed, and if the experimental treatment is carefully and accurately described, it may be unwarranted to assume that the obtained experimental effects would be found with other samples, in other locations, or with other variations of the particular treatment employed. Treatments that seem effective with male offenders may not work with females, treatments that have an effect in small institutions may not work in large ones, and treatments may have different effects if even small changes have to be made to meet local conditions. Establishing external validity for treatments may be the most difficult task of all since there are so many different parameters along which one might wish or need to generalize about treatment effects and since careful and separate tests might be needed for many variations of sample, locus, and treatment type before one could well understand the generalizability of a treatment outcome and feel safe in recommending its implementation in a wide variety of specific instances.

STRENGTH AND INTEGRITY OF TREATMENT AS PROBLEMS IN CONSTRUCT VALIDITY

The essence of construct validity is that one has good understanding of the real or true meaning of a response-eliciting mechanism, e.g., a stimulus. In the field of psychological testing the response-eliciting mechanism is the set of items or other stimuli that constitute the test; the construct is the process or characteristic which is presumed to underlie the response. A test of anxiety has construct validity to the extent that theory and evidence converge on the conclusion that anxiety is reflected in or determines the response to the test. A test of anxiety would not have construct validity for anxiety if some factor other than anxiety determined the responses, but the test might have good construct validity for that other factor, if one only knew what that factor was.

Extended to the context of treatments, construct validity applies to assertions about the effective ingredient(s) in a treatment: treatments do not have construct validity; rather our interpretations of them do. For example, performance contracting might be associated with a reduction in recidivism in a juvenile sample, but the truly effective component of performance contracting might be increased attention consequent to developing and monitoring the contracts. In such a case "performance contracting" would not have construct validity as an explanation of the reduction in recidivism. Note,

however, that "increased attention," if it occurred as an explanation, would have construct validity. Failure of construct validity can also occur when a treatment is assumed to be taking place when, in fact, nothing is happening, or when a treatment is delivered in some diluted or even intensified way. In each instance, the "explanation," the label attached to the treatment, is incorrect or misleading.

Strength of Treatment

If we may draw an analogy from medicine, a drug may be delivered in varying doses, the dose being a function of the quantity and purity of the drug, the frequency of its delivery, and the length of time over which it is given. A physician will ordinarily plan, to the best of his ability, to deliver the right amount of the drug over the appropriate amount of time. Extending the analogy to another type of treatment, one would think that the strength of a counseling intervention might be reflected in, among other things, the number and length of sessions, their frequency, and the period of time over which they are attended. The strength of a system of postrelease financial support might be expressed in terms of the amount of money given, the frequency of its delivery, and the length of time over which the support is extended. In general it is supposed in medicine that weak treatments lead to weak or nonexistent results and that they are to be avoided. The same assumption might be made about the treatments proposed to rehabilitate criminal offenders.

To continue the medical analogy, it is also the case in medicine that a treatment plan, e.g., for delivery of a drug, takes into account the nature and seriousness of the problem and the characteristics of the patient. A longer course of antibiotics will be prescribed for some conditions, stubborn ones, than for less serious ones, and the dosage levels will be adjusted for such factors as drug sensitivities and body weight. What is a strong treatment for one problem or person may not be so strong for another. Again, extending the analogy, what is "a lot" of group counseling for one problem may not be for another problem and, what is sufficient for one person may not be sufficient for another. In medicine attempts are usually made to keep treatments at optimal levels rather than simply to plan the strongest treatment possible for every case, it being recognized that treatments stronger than necessary are not only needlessly expensive but often dangerous.

In medicine there is a fair degree of uniformity of response of patients to treatments owing to substantial homogeneity in anatomy, physiological processes, and the like. That is not to say that patients with the same problems all respond in the same way to a given treatment, but out of 100 patients with a typical medical problem, most will respond in more or less the same way to any

treatment chosen. Moreover, when there is some variation in response, it is often a relatively simple matter to adapt the treatment to the individual, e.g., lengthen the course of treatment, perform a second operation to remove scar tissue. Moreover, there is a great deal of empirical work on standard treatments which provides a basis for determining an optimal or strong treatment. If a physician wants to give a strong dose of tetracycline to a patient, he or she has a pretty good idea of what would be considered a strong dose for a given patient and what the likely response will be.

Here the analogy clearly breaks down. What is a "strong dose" of group counseling for a first conviction felon charged with robbery? What is an optimal level of financial support for a released ex-burglar that will make it possible for him to avoid going back to crime and still make it worthwhile to try to find a job? Uniformity of response of criminal offenders to treatments can be assumed to be improbable, and there is no body of empirical research on which a plan for treatment at any level, let alone optimal, can be based with confidence. Unfortunately, there is no counterpart to animal studies and laboratory physiological assays to guide those planning rehabilitative efforts. Still, to admit the problem exists does not make it go away. In this paper we contend that one of the sorest needs in criminal offender rehabilitation is a way of assessing strength of treatments, both before and after they are delivered.

Strength of treatment is an issue in construct validity because it is so easy when evaluating the results of a treatment to assume that something has happened that has not. If a decision about the effectiveness of a drug were made without knowing whether it was delivered in a strong enough dose, the scientific medical community would raise vociferous objections. A case in point is the University Group Diabetes Program (UGDP) clinical trial of an oral hypoglycemic in the treatment of diabetes.³ When the study directors concluded that the oral agents were ineffective and probably even dangerous, there was an immediate objection that every patient had been given the same dose,⁴ a seemingly reasonable scientific procedure, but one potentially objectionable on clinical grounds since, ordinarily, dosages would be adjusted according to patient characteristics. Yet, when it has been concluded that some form of treatment does not work for rehabilitation of criminal offenders, few objections have been raised that the treatment may not have been strong enough. Lipton, Martinson, Wilks⁵ and Martinson,⁶ to take an example, raised no such questions about the treatments they surveyed, and Kassebaum, Ward, and Wilner⁷ raised no such question about their own study, even though, as Quay⁸ has so well demonstrated, there was ample reason in their own account to do so. Any conclusions about whether a treatment is effective or not must be reached with full knowledge of just how strong the treatment was.

Integrity of Treatment

We make a conceptual distinction between the strength and integrity of treatment although after the fact they often amount to the same thing. Once again, using the medical analogy of a drug treatment, we may think of the strength of the treatment in terms of the treatment as planned and prescribed. Thus, 8 tablets per day of tetracycline for 6 days would be considered a strong treatment for urinary tract infection in the case of an adult. Half that dosage would be considered weak on a priori grounds.

Integrity of treatment refers to the fidelity with which the treatment plan is carried out. Thus, unwitting substitution of a drug of lower strength, use of an adulterated or impure drug, and failure of the patient to take all the medication would all threaten the integrity of what would otherwise have been a good treatment plan.

Similarly, there might be some reason to suppose that for a certain subset of prisoners 2 individual counseling sessions per week for 1 year would be a fairly strong treatment. However, if the therapists available were untrained or did not believe in the treatment, if the sessions were never held, or if during the course of treatment the mode of therapy were switched, the integrity of what might have been a good plan would have been destroyed.

Strength of treatment is something that is planned for and which may be assessed at any time, whereas integrity can be fully assessed only after the treatment has been completed. Integrity of a treatment usually requires a specific set of mechanisms for accomplishing that task and provision for continuous monitoring. In order to maximize integrity of treatment the treatment must be well defined so that standards of judgment can be applied, and at any sign of deviation from those standards, there must be provision for a corrective action. When treatment is completed, it should be possible to make a good estimate of the degree to which the planned intervention was carried out. Results obtained in a test of the intervention then must be interpreted in light of the achieved integrity of treatment.

STRENGTH OF TREATMENT IN THE CRIMINAL OFFENDER REHABILITATION LITERATURE

The strengths of treatments tested in attempts to rehabilitate criminals should be an obvious issue, but it may be even more important than usually recognized if one considers the task to be accomplished. It is not at all clear from reading the lit-

erature in the field that the difficulty of the task of changing criminal behavior has been uniformly and fully appreciated. As will be shown later, many of the treatments tested do not seem especially powerful on a priori grounds, whatever the behavior to be affected, and it seems probable that much criminal behavior should be resistant to change.

The Target Population

In estimating the strengths of treatments necessary to change criminal behavior, one may begin by considering the nature of the populations whose behavior is to be changed. To begin with, and for fairly obvious reasons, the focus of most efforts at rehabilitation has been on inmates of institutions. There is good reason to believe that by the time they are recognized and formally identified by the criminal justice system as offenders (even if they are diverted), the persons involved are a highly select group. They are likely to be unemployed or only partially employed, disproportionately of minority group status, undereducated, adrift from their families or other socially centripetal groups, and to have many friends much like themselves who in one way or another provide support for their criminal activities. These individuals are not good prospects for rehabilitation under any circumstances. Then to encounter tests of such treatments as group counseling, training for probably nonexistent jobs, and wilderness experience does not impress one with the likelihood of change.

It may be argued that the persons involved are already receiving treatment, e.g., incarceration, and that what is being tested is merely an add-on, but even as an add-on, a treatment such as any one of the above would seem likely to be rather insignificant. On a priori grounds it would seem that persons identified as criminal offenders usually would need really massive interventions if they were to be successfully changed for the better.

The Target Behaviors

The rehabilitation of criminal offenders requires that the dual task be undertaken of eliminating a set of undesirable responses and submitting a set of more desirable ones. The difficulty of either task should not be underestimated. Few inmates of institutions are first-time criminal offenders although more of them are serving their first sentences. Even among those identified as offenders and diverted from the system at any point, it seems likely that relatively few are truly first offenders. A very large proportion of inmates have long criminal records and even longer lists of crimes for which they are responsible but for which they have never been apprehended. This is by way of saying that for many criminal offenders the criminal behaviors which must be eliminated are habitual, well-practiced acts. Even if the particular

act involved, e.g., homicide or rape, has not occurred many times, the act is very often merely the most extreme in a series of violent responses going back many years and representing a well-practiced, customary response to stress.

Moreover, criminal behavior does not go unrewarded. Criminal behavior is often highly rewarding, at least in the short run, and rewarding in ways not readily substituted for by noncriminal activity. For example, few noncriminal occupations produce even occasional large infusions of money such as may come from a good heist. Few noncriminal activities can inspire the same respect--of sorts--that comes from being the toughest guy on the block. Criminal behavior is also often inherently rewarding in other respects, e.g., for the excitement it brings to otherwise dull and mundane lives. Criminal behavior may even be virtually normal in the groups and societies from which many offenders come so that rehabilitation involves adoption of a deviant way of life. Again, many of the efforts to change criminal behavior that have been tested appear to be woefully improbable of producing the change desired. Even the argument that prisoners are already being treated for criminality by being incarcerated may not be particularly persuasive if one considers that for many criminals arrest is not especially shameful and imprisonment may be merely an occupational hazard akin to that of duty in remote and unpleasant places for a soldier or confinement to a ship for long periods for a sailor.

The above considerations lead these authors to conclude that rehabilitation of criminal offenders should be regarded from the outset as an extraordinarily difficult task requiring, in all probability, heroic efforts in order to achieve dependable and extensive change. It is instructive to examine the efforts that have been made to produce change. One discovers few, if any, instances of efforts that seem impressively strong, and for those that have been made, it is difficult to assess strength of treatment even within the system of change proposed.

Examples of Problems in Assessing Strength of Treatment

Work Release. On first thought the issue of strength of treatment might not seem especially problematic and perhaps not even relevant in a program such as work release. After all, prisoners are either on work release or they are not. The case, in fact, is not so simple. One must first determine the purpose for work release as a rehabilitative effort. (If it has no rehabilitative intent, but is only to reduce costs of incarceration, no strength issues arise.)

Waldo and Chiricos⁹ suggest that there are a number of theoretical reasons for supposing that work release might have some rehabilitative effect. Work release might have a positive impact on postrelease employment prospects, reducing the pressures to

return to criminal activity; it might introduce the inmate to a new and law-abiding set of potential associates; it might change his own and others' image of himself as a criminal, etc. When one begins to consider the possibilities for effects of work release and answers to the question of whether work release might have a rehabilitative effect, questions such as the following arise:

- When should work release begin?
- For how long a time should a prisoner be on work release?
- How good a job at what kind of pay would be required?
- Is work release likely to be effective if the prisoner returns to a community other than the one in which he did his work release?

All these questions pertain to the strength of work release as a treatment. To reduce the argument to a near absurdity, it would seem evident that a few days of work release at an undesirable job in a strange community would not constitute much of a treatment; 2 years of work release in an extraordinarily good job at high pay in a very desirable community might ought to have at least some effect. There is a lot of room in between.

Let us consider two examples which differ substantially in their conclusions. Witte¹⁰ has evaluated a work release program in North Carolina which appeared to have at least some effect of reducing criminal behavior. Waldo and Chiricos¹¹ evaluated a work release in Florida and concluded it had no effect at all. Admittedly there were differences between the studies, and both analyses seemed to show that work release was at least not harmful and was probably cheaper than continuous incarceration. However, what is of interest here is the nature of the work release program.

Without being precise about it, but rather keeping in mind some general image of what a work release program probably should be, consider the program described by Witte. Prisoners served an average of 7 months, or 24 percent of their sentences, before being put on work release, but 15 percent of them had served over 1 year before being put on work release, and inmates with sentences longer than 5 years have to serve at least 15 percent of it before going on work release. Inmates must have had jobs before going on work release, and a shortage of prison personnel made it difficult for inmates to look for jobs outside the prison; most of them went to jobs they held prior to imprisonment. Work release jobs generally required lower skills and paid lower wages than jobs the men held

either before imprisonment or after release. Prisoners had first to pay the State for their room and board, had to pay State and Federal income taxes, and had to make payment to dependents. Any money left could be saved, and, on the average, men on work release had saved \$428 upon release; however, 28 percent of them had saved \$100 or less.

The article by Waldo and Chiricos illustrates a particular problem in the evaluation literature: their paper contains so little detail on work release in Florida that it is impossible even to begin to compare it with work release in North Carolina, let alone make a judgment about the absolute strength of the treatment. However, Waldo and Chiricos do imply that strength of treatment might be considered an issue since they divide their sample according to "extent of participation" in work release, i.e., amount of time in the program. Dividing the sample at the median, the lower level participants had had up to 82 days in work release; there is no indication whether that is 82 work days or 82 calendar days. Unaccountably, Witte does not provide information on the amount of time that North Carolina prisoners spent on work release.

Now, how strong are these two programs as implementations of work release? Are either clearly strong enough to warrant a clear statement about whether work release reduces recidivism? These writers think not. Witte believes that work release reduced seriousness of postrelease offenses and increased time to first offense although she did not have a strong enough design to permit such an unequivocal conclusion, at least not in our opinion. It is clear, however, that if one set out to design an optimal work release program, the North Carolina program would fall short, in all probability by a considerable degree. The Florida program is simply not described other than to note that the median time in work release was 82 days. Whether it is comparable to North Carolina in other ways, e.g., type of job available and rate of pay, cannot be said. Probably, however, it fell short of a program which would be judged strong in terms of likelihood of rehabilitating offenders, as opposed to reducing costs to the State. No conclusions about the potential effects of well-designed and carefully implemented work release programs should be drawn from the available evidence.

Job training. Again, on a purely commonsense basis, if one were designing a strong job training program it would have several features:

- Identification of job opportunities in the community;
- Selection of well-trained and highly motivated instructors;
- Provision of training materials identical to those used in work settings in the community;

- Development and implementation of a carefully devised, systematic training program;
- Training in work habit skills as well as in technical skills;
- Matching of interests and skills of inmates to the opportunities available in training;
- Assurance that trainees reach reasonable criteria of performance before termination of training;
- Postrelease job placement services.

If one had such a program and if it proved to have no effect on recidivism rates, one might then despair of the prospects for job training. Some writers have concluded that job training does not affect recidivism rates, but how good was the job training?

Two recent commentators believe that job training programs have not been well implemented.^{12,13} They comment on such features of the the programs as the necessity to work job training around stringent and sometimes erratic prison schedules, about job training not coordinated with release dates, obsolete equipment, poorly trained teachers, lack of training in good work habits, inmates forced into unwanted trades, and competition between better prison assignments such as minimum security work details and job training for the better prospects among the inmates. How, then, is one to conclude that job training "doesn't work"?

The study of vocational training in California institutions by Dickover, Maynard, and Painter¹⁴ illustrates the problems of treatment strength in evaluating vocational training. These investigators found that number of hours of training, grades in training, and proximity of training to release date were all associated with successful job placement. On the other hand, lack of staff with responsibility for placement restricted the success of placements. Such findings indicate rather clearly that evaluation of vocational training needs to take into account such factors as amount of training, adequacy of placement services, and the like. It was also evident in the California study that many trainees had poor attitudes toward work or lacked good work habits, so that a really strong vocational training program would have to deal with that problem. The California study also indicated that inmates were sometimes being trained for trades in which there was no demand, a circumstance not obviously inevitable.

The writers would, on the basis of reports and other information currently available, conclude that vocational training, like

so many other treatments, has not been tried in an optimal form and that to impugn its effectiveness in reducing recidivism is unwarranted unless the imputation is carefully circumscribed.

Counseling. By this time readers will surely have developed for themselves the capacity to formulate what would probably be regarded as strong treatments for most types of interventions that might be proposed, and there is no need to detail the requirements for a strong counseling program; they would have to do with training, intensity of contracts, motivation of counselees, etc. Have counseling treatments in strong form even been tried? We believe not, although we have not read all the literature on that topic. Perhaps two examples might suffice to illustrate the problems with and limitations on conclusions that should have, but have not necessarily, resulted.

On February 25, 1963, the Chicago Sun-Times began its lead editorial with a quotation from what was then a very recent research report: "Individual counseling of potentially delinquent high school students is ineffective so far as improving their school behavior or in reducing the number of drop-outs." That quote was, by implication (although we have not been able to find it in the text), from Girls at Vocational High.¹⁵ That Meyer et al. study has since been cited in numerous places as evidence for the inefficacy of counseling. Even though the study was directed toward predelinquent rather than delinquent youth and involved only female counselees, it provides a useful illustration of problems in evaluating strength of treatment.

In the study, predelinquent ("potential problem") girls in a vocational school were identified and enlisted into a counseling program. Experienced social workers in a volunteer youth service agency served as counselors. In many respects the treatment described might be thought to be nearly optimal. Upon closer reading, however, problems are raised that suggest that the utterly pessimistic conclusions drawn from the study may be exaggerated. In several places the authors note that the agency and the social workers involved in the study were trying something new, that despite their previous experience, they had not worked with a population quite like the one involved in the study. They began by taking 53 girls into individual treatment, and 27 of those 53 were ultimately judged not to have been "appreciably involved" (p.112) in treatment. Of the 27 girls not appreciably involved, 12 were seen for 4 or fewer sessions and none for more than 19 sessions. After some period of time spent attempting to treat the referred girls in individual treatment, the therapists concluded that group therapies were probably preferable since individual therapy seemed to make many of the girls unduly anxious. Therefore, the treatment plan was revised to establish a number of different kinds of therapy groups although some girls received additional individual treatment during the course of the project. The authors note in several places, however, that

many of the girls were anxious about having been picked out for referral to the agency since they were unaware of the basis on which they had been selected. In the view of the agency "the levels and goals of treatment should be based upon a psychosocial evaluation of the client. The groups in this project had to proceed without the benefit of what was considered significant diagnostic material since only the more general facts were known" (p.118). Although the social workers had had experience counseling girls from the same general population as the one involved in the study, their experience had been mostly confined to girls who had already met failure, e.g., had already become pregnant, and the authors refer continually to the fact that the project was a learning experience for the counselors. Finally, the authors conclude that a "serious professional task from social work is to develop intervention procedures" (p. 214) to deal with the social systems of which the girls are a part since the type of treatment studied in the project did not appear to be "the most efficacious."

There are two points to be made with respect to the above material. First, although the treatment studied may have been the best that could have been devised at the time, in retrospect one might well improve on it and might produce somewhat more positive findings, e.g., one would probably begin right off with group treatments and do a better initial job of diagnostic study. Of special importance is the fact that despite the evident differences in judgments of the satisfactoriness of treatment, in the analysis all cases, individual or group, involved or uninvolved, are combined. That analytical procedure is admirable from the standpoint of minimizing prospects for a Type I error, but it surely increases the chances for a Type II error. The second point to be made is that the pessimistic conclusion, even if warranted, should be carefully circumscribed. The treatment studied was a fairly narrow variety of counseling that, in particular, did not involve any sort of active intervention of the sort that would ordinarily be a part of social casework. The study has been much more broadly interpreted as an illustration of the more general ineffectiveness of social work and counseling.

The second example of a counseling study widely cited for its pessimistic findings is the Kassebaum et al¹⁶ investigation of the effects of group counseling on inmate behavior and recidivism on parole. The study is widely and, for the most part, justly admired for its experimental design and for the length of the followup period. However, as has been shown especially by Quay, far too little attention has been paid to the actual nature of the treatment about which the very pessimistic conclusions have been drawn.

Without going into great detail (or merely repeating the points made by Quay), the present writers would assert that the treatment tested by Kassebaum et al. does not appear to have been particularly

strong. Counselors were mostly chosen from the ranks of correctional officers and were not professionally trained. They did attend training sessions, but the effectiveness of those sessions is not documented in any way. The authors state that counselor training sessions were poorly attended and that many of the men who attended were apathetic and disinterested. The authors developed a special training program for half the counselors since they recognized the possibility that the eventual findings of the study might be dismissed on grounds of inadequate treatment. However, they do not present any evidence concerning the effectiveness of their training. There is, in fact, no evidence of any kind bearing on the competence of the counselors to do their job. Only one-third of those in the special training program had had any previous counseling experience.

There is evidence that a rather sizable portion of the counselors, perhaps 40 percent or so (p. 64), had very little faith in the efficacy of the counseling program. The authors themselves state that the theoretical basis for the counseling was expressed only in a very general way. In the authors' own terms, "we are not sure just what it is about the group sessions that promote changes in attitudes and behavior," (p. 59) but they then go on to assert that this limitation does not prevent studying the effects of group counseling if an appropriate design is employed. The limitation certainly restricts severely the generalizations one might make on the basis of the findings since it is difficult to state clearly just what was tried.

Nor were the inmates particularly well-disposed toward the counseling. The authors admit that many of the inmates probably attended counseling sessions because they thought it might influence their chances for parole, and three-fourths of them agreed that men did not talk frankly in counseling. Virtually half of them also believed that things brought out in the counseling session might be used against them. Sixty percent believed that correctional officers were not competent to run groups.

There is an additional issue concerning just how much of the poorly specified and possibly poorly implemented treatment was actually delivered. The group sessions were run either in "small" groups of 10 to 12 or in large groups of about 50. Even groups as small as 10 would be regarded as rather large by most group counselors and therapists. There are no data presented at all about actual participation in the groups, e.g., how many men spoke, what topics were discussed, whether there was any absenteeism, or how many men (or counselors) may have slept through meetings. Groups were to meet for "an hour or two" once each week, and the counseling program went on for, apparently, 2 years. However, no data are reported for the amount of time spent by individual inmates in counseling. Presumably some might have had 2 full years of counseling while others had only a few months at best.

In short, the study by Kassebaum et al., which has been cited by some as a definitive study of the value of group counseling with prison inmates, may justify only very narrow conclusions about counseling. It is not unimaginable that a counseling program involving professionally trained counselors operating from a clear theoretical position and with volunteer inmates might be put together. Thoughts of doing so should not necessarily be inhibited by the results of the program actually tested by Kassebaum and his associates.

Probation and Parole. When considering probation and parole as rehabilitative efforts, the issues of treatment strength are as important as for other interventions and the importance of that dimension seems to have been more clearly recognized by those testing probation and parole. For a good many years there has been an obviously persistent expectation that success of probation and parole should be enhanced by increasing the intensity of contact or supervision (i.e., the strength of treatment).

The primary strategy to increase intensity of contact or supervision has been to decrease average caseloads of probation and parole officers. Actually, rather widely differing levels of caseloads have been tested, particularly in the California Special Intensive Parole Units (SIPU) projects in which case loads varying from 90 cases down to 15 cases have been studied.¹⁷ Those caseloads span the limits of what would seem to be reasonable caseloads from the high range to the low, and from the narrow perspective of caseload size, one could probably not support the notion that treatment would be stronger with even smaller caseloads. However, the theoretical rationale for smaller caseloads is more intensive contact between parolee and officer, and there may not be a linear relationship between caseload size and amount of contact.^{18,19} Results of comparison between caseload sizes have fairly consistently shown little relationship between caseload size and recidivism except that small caseloads do tend to be associated with higher levels of technical violations that stem from the closer supervision.²⁰

Other persons working in corrections have thought that probation and parole supervision might be enhanced (by a stronger treatment) if probation and parole officers could be selected for some special characteristics making them more attractive to and effective with prisoners. For example, the California Special Intensive Parole Unit studied the concept of matching parole officers with offenders,²¹ but whatever matching was achieved had no effect on recidivism rates. The Ohio program attempted to use ex-convicts on the grounds that they would be more understanding, in the sense of empathy, and perhaps in the sense of not being easily conned.²² Berman went in the opposite direction and studied the use of volunteer lawyers.²³ It is not obvious that either type of person constitutes a strong parole supervision and contact treatment, but certainly the two groups do represent fairly extreme instances of the range of persons who could plausibly be considered

for the role of probation or parole supervisor. Nonetheless, neither type of supervision resulted in recidivism rates below those in the comparison groups.

What is not so clear, in part because it is so rarely discussed, is the strength of probation and parole supervision that has been achieved and tested when viewed from the standpoint of personal qualifications, training, and dedication of the supervisors. It requires little investigation to discover that probation and parole officers in most jurisdictions are not as a whole highly trained, that they work under conditions not conducive to high levels of dedication, that they are poorly paid, and so on. Whether changing case loads for such officers could be expected to have much effect on the way they work and on the effectiveness of their efforts is questionable. In any case it may be that the responsibilities of policing and surveillance are so integral to the concept of the probation and parole supervisor that the notion that such persons could be effective in the rehabilitative role is misguided in the first place.

The Generality of the Issue of Strength of Treatment

The writers do not wish to imply that by limiting the discussion of strength of treatment to the topics given above that its applicability is limited. In fact, the issue of strength of treatment could be raised with respect to any treatment proposed, since such factors as timing, length, training, or personnel, and the like, are factors that affect all treatments.

One additional factor involved in strength of treatment, albeit in a nonobvious way, is the treatment x subject interaction, i.e., the possibility that certain treatments are better suited for some subjects than for others. In view of the fact that at this point one could not with any confidence specify differential treatment for subjects such that little more than marginal improvements in obtained results would occur, the writers consider the case for treatment x subject interactions in offender rehabilitation to be still rather weak. Nonetheless, it seems likely that at some time in the future the capability for specification of differential treatments will be realized. Those like Warren,²⁴ Glaser,²⁵ and Palmer,²⁶ who are strong advocates of the notion that it is possible to match treatments to offenders, may be somewhat ahead of the field and its actual capacities, but it must be admitted that, in the long run, it will be necessary to assess strength of treatment in light of offender characteristics just as medication dosages are adjusted to body weight and other characteristics of sick persons.

Rehabilitation--in the sense of reduction of recidivism--is only one of several rationales for the diverse treatments that have

been tested with criminal offenders. The discussion of strength of treatment refers solely to the rehabilitative rationale. To the extent that other rationales are relevant or paramount, the issue of strength of treatment becomes moot or takes on a different coloring. For example, an alternative rationale for some offender dispositions is strictly economic, i.e., that offenders can be managed with equal public safety in a way that costs less money. That is the rationale often invoked for work release programs (e.g., Witte²⁷, Waldo and Chiricos²⁸), early release and parole (Havel and Sulka²⁹), and other treatments. With such a rationale the question of whether a treatment is as strong as possible is at best irrelevant and at worst counterproductive since a stronger treatment might cost more money and vitiate the economic rationale for alternative management.

Another rationale might be that prison intramural programs keep prisoners busy and out of mischief, e.g., in vocational training, education, recreation, and even counseling. If keeping prisoners busy is a rationale for treatment, then the question of strength of treatment comes down to the question of how busy the treatment keeps them. Presumably a strong treatment would be one with the potential for keeping many prisoners occupied and out of mischief for a good many hours every day. The outcome measure would be the number of hours of prisoner-time occupied per unit time, say one week.

Still another rationale for some treatments might be the humane rationale that, as human beings, prisoners do not deserve, on top of everything else, to be idle and bored (see, for example, Wexler³⁰). Under that rationale, strength of treatment would be assessed in terms of its interest value to prisoners. Active recreational programs might then be regarded as potentially stronger treatments than, for example, vocational training or counseling.

A Priori Assessment of Strength of Treatment

One can certainly assess the strength of a treatment in terms of its outcome. If effects were large, then the treatment was, by definition, a strong one; if effects were small or nonexistent, the treatment was weak. However, a weak treatment may have been weak for any of three reasons: (1) some treatments are inherently weak and will do rather little for any condition; (2) some treatments are potentially strong but are inappropriate for the condition for which they are used; and (3) some treatments are potentially strong but are given in only a weak form. In the latter case the problem may have resulted from an initial plan that incorporated a weak form of a treatment or from a plan that began with a strong form but went awry. To avoid abandoning a potentially strong treatment that has not been adequately tested, it is necessary to make some a priori

determination of treatment strength so that one can differentiate between poor outcomes attributable to the three reasons given above: reason 1 recommends abandoning the treatment, while reasons 2 and 3 recommend further exploration. Is it possible to make an a priori determination of the strength of a treatment, and if so, how?

Theoretical premises. These writers would suggest that the evaluation of strength of treatments begin with assessment of the theoretical premises on which the treatment is based. Greenberg has noted that the theoretical bases for many interventions "border on the preposterous,"³¹ a somewhat polemic but perhaps warranted description. At the very least it is evident that a large proportion of the theorizing, if even present, is implicit. It seems, in many cases, as if theory is lacking altogether, and interventions are being tried for no better reason than that they seem like good ideas at the time.

Glaser has asserted that the failure to employ adequate theory is a key deficiency in criminal justice evaluation research.³² He notes, for example, that it is pointless to test the effects of psychotherapeutic interventions on prisoners without differentiating among prisoners' probable response to psychotherapy. Glaser believes that when proper differentiations have been made, a distinct value for psychotherapy can be shown.

At the outset, then, in evaluating the potential strength of a treatment, one can determine whether the treatment is well-grounded in a theory which provides links between the type of intervention which is proposed, the types of offenders to which it is to be applied, and the anticipated outcomes. Were more treatments subjected to such scrutiny, it seems likely that many of them would never be tested, or at least not tested on such a large scale and so seriously.

A case in point is the study of counseling by Meyer et al.³³ Had those planning the intervention considered the realities faced by the girls at the vocational high school who were their subjects--poor family situations, physical abuse, bad housing, undesirable social conditions--one wonders whether it would have seemed at all credible that any kind of personal counseling, whether individual or group, could have made any difference. What was the presumed link between the activities of group counseling and response to a physically abusive parent? It is not surprising that the evaluators concluded that the kind of treatment tested was not likely to be effective and that more direct ways of intervening in the girls' real life circumstances needed to be developed.

A systematic and rigorous theoretical analysis prior to the Kassebaum et al.³⁴ counseling study might well have reached the conclusion that an hour or two a week of counseling of the type

provided was unlikely to have detectable effect on the response of offenders to parole over a period of 3 years and probably little if any effect on the outcomes studied. These writers do not believe that any such prior theoretical analyses were ever done and believe that they are infrequently done for other types of interventions. The closest thing to a theoretical hypothesis in the Kassebaum et al. study is the statement that counseling should weaken the commitment of prisoners to the criminal value system. There was no specific link proposed between the counseling and the weakening of that commitment and subsequent adjustment on parole. The remaining hypotheses stated were rather mundane expectations about effects of counseling that specified no mechanisms at all by which the effects were to be achieved.

The case may not necessarily be better for other interventions. Testimony to the lack of theoretical analysis is provided repeatedly by studies which conclude with the investigators being surprised that more intense supervision of probationers or parolees leads to higher levels of detected and reported offenses, especially technical violations. To the best of these writers' knowledge, not one investigator made such a prediction, even though it should have been obvious in advance that reducing case loads would in itself not change the handling of technical violations. Nor have the links between intensity of supervision and parole adjustment been well specified, if more intense supervision is to have its effects through greater surveillance to be produced and specify the types of offenses for which the greatest decrements would seem likely.

A study of prisoners released from Massachusetts forest camps revealed no lower levels of recidivism than for prisoners released from other types of facilities.³⁵ It is difficult to absorb such a finding. Why would one have expected any relationship between release from a forest camp and subsequent recidivism? The research report is totally nonrevealing of the origin of interest in the question. There was a difference in recidivism rates for offenders released from two different forest camps, and the report attributes these differences to the difference in visitation frequencies between the two facilities. That is a hypothesis possibly worth exploring, but as after-the-fact theorizing it is totally unsatisfactory.

As indicated previously, many studies appear to involve a good bit of implicit theorizing, but generally the theory is limited to rather general expectations that doing what seem like good things for offenders will eventually have some good effects. In the recently published report on the Unified Delinquency Intervention Services,³⁶ a variety of services provided to juvenile offenders are described and, while the services generally seem like good things to do, specific rationales are not always given for them (e.g., rural programs), nor is there specification of the links by which the ser-

vices might be expected to have effects on subsequent offense rates. It might be both humane and otherwise desirable to see to it that juvenile offenders who need dental work have it done, but it is quite unclear how having that work done might decrease subsequent delinquency.

Either the Theory or the Practice May Be Wrong. One of the great disadvantages of a weak or nonexistent theoretical position is that when, as so very often happens, expectations about the effect of a treatment are not met, there is no good basis for interpreting the failure of the treatment. If one has a theory (more or less) that psychotherapy will improve self-esteem, and that with improved self-esteem some types of offenders will experience reduced pressure to follow their peers in criminal activities, and that such reduced pressure will result in lower rates of recidivism upon release, then if the obtained results are not consistent with expectations, e.g., if recidivism is not reduced, one is left with a series of questions of interest and potential import.

One may conclude that the theory was wrong and abandon that line of thinking. But one might also wonder whether that type of psychotherapy was wrong, whether the increment in self-esteem was achieved. One's conclusion might be limited and not totally pessimistic. However, if one's theoretical perspective were no better thought out than that in some unspecified way psychotherapy ought to have an impact on recidivism, the failure to find an effect would be minimally informative and devoid of heuristic value.

Threshold Effects. Careful analysis in the planning stages of an intervention may well suggest the likelihood of threshold effects of treatment, i.e., that unless a treatment reached some given level of strength or intensity, no effects would be detectable and that above the threshold, effects should be rather immediately detectable. A compelling, if specific, illustration is provided by schemes which propose to give financial support to released offenders for a period of time while they look for work, the purpose of the subsidy being to make it at least possible for ex-offenders to avoid crime until they can find jobs.³⁷ In all probability there is something akin to a threshold effect in the amount of money necessary to make such a program work. If the amount of money necessary is too small, it can only supplement funds acquired in illegal activity. Above threshold the effects may be described by a fairly flat function to some level, since if parolees are really interested in going straight, having enough money to get by on will be enough, and more than that will not help them much. Probably, beyond a certain point, the money might become sufficient to make the parolee disinterested in looking for work, since he might have enough money for a time, even if only for 3 months or so, to enable him to indulge himself in comfortable idleness.³⁸

There may be threshold effects for other treatments also. A little bit of vocational training will probably be totally wasted, and if a trainee cannot be brought to some particular skill level, there may not be much point in beginning the training. From a rehabilitative standpoint there may be little value in educating prisoners who are high school dropouts. If one adheres to labeling theory, there may not be much advantage in a diversion program which leaves an offender only a little bit labeled. Even for some types of counseling there may be threshold effects if there are stages in the counseling process, e.g., building a relationship between client and therapist may have to precede a therapist's more direct attempts at client change.

To summarize, a strong treatment is likely to have a clear and acceptable theoretical rationale that specifies not only a plan for treatment but the mechanisms or processes expected to produce the desired change. Paul Samuelson is said once to have remarked that an economic theory that an economist cannot explain to his spouse is scarcely likely to be true. In these writers' view a theory about treatment that cannot be clearly stated and that is not persuasive to most readers is scarcely likely to be useful and productive in effecting change.

Expert Judgments. A second approach to a priori assessment of strength of treatment might be to have the treatment plan examined by experts in the field who would then rate the treatment on the likelihood of its producing important changes or who would estimate the amount of change likely from implementing the treatment. For example, one might have a group of experienced, professionally trained parole officers estimate the reduction in recidivism that might result from providing each parolee with a financial subsidy of \$60 a week for 13 weeks (the arrangements in the previously cited U.S. Department of Labor document of 1977). Or one might have counseling experts estimate the amount of change in criminal behavior that might be achieved by a particular counseling program. Alternatively, or additionally, one might have experts in corrections estimate effects likely to be produced by various programs. One might even have offenders estimate the strengths of different interventions. If the strength of an intervention is to be estimated by an expert, it is important that the intervention be described as specifically as possible.

In some respects the idea of having experts assess strengths of treatments on an a priori basis may seem a foolish enterprise quite unlikely to be revealing, but the senior author and one of his students, William Yeaton, have been investigating that possibility in various fields, not yet including corrections. Twelve experts on the design of interventions to control cigarette smoking estimated the percent of treated subjects who would cease smoking in response to 50 different experimental interventions reported in

the literature. The average correlation between judges' estimates and the actual percentage of those who ceased smoking was 0.47, and some of the judges achieved correlations up into the 0.70's. We believe that it is not beyond reason that carefully picked judges could make good estimates of the strengths of various treatments for criminal behavior in terms of the change likely to be produced. Even if such estimates were not totally accurate, they would almost certainly be better than no estimates at all, which is the current state of affairs.

Norms. For some treatments or aspects of some treatments, it might be possible to develop normative data that would be useful in assessing strengths of treatments. For example, if it were known that the typical counseling program in prisons involved 1 hour per week for 16 weeks, that the typical counselor had a college degree and one course in counseling, and that the typical counseling recipient was a volunteer in the last 6 months of his sentence, any proposed counseling program to be tested could be compared against that standard. A program which offered only eight sessions would appear on a priori grounds to be weak and probably not worth testing unless, in some manner, a special case for it could be made. Similarly, norms could be developed for length of work release, time in sentence at which work release begins, rate of pay, type of job, etc., and any work release programs could be compared to the norms to determine whether they seemed stronger as treatments than the typical program. Parole norms could be established for caseload size, frequency of encounters, level of training, and any other variables thought important.

Cost. Obviously it would be a mistake to suppose that there is any direct or even necessary relationship between the cost of an intervention and the probable outcome. Nonetheless, within types of treatments it might be useful to document costs so that treatments could be compared. It would be of at least passing interest to know that the cost of maintaining an ex-offender on parole in one jurisdiction was three times the cost elsewhere. Substantial differences in costs do not necessarily denote differences in treatment strength, but they may at least call attention to the issue and demand explanation that may be enlightening.

Parametric Study of Treatment Strength

In order to facilitate systematic comparison of treatments, there might be an advantage in establishing a standard set of parameters for comparison, along with guidelines for assessing treatments with respect to the parameters. Experts might at some point be used to establish weightings of the parameters so that they could be combined into a smaller set of indices or even into a single index

of treatment strength. It might never be possible to make good and legitimate comparisons across treatment types, but it might be quite possible to make comparisons within treatments, e.g., counseling might not be comparable to work release, but one counseling program could be compared to another.

Listing possible parameters for consideration is not difficult, but any list established would be arbitrary and would need to be revised in light of opinions of experts in the field. Despite the arbitrariness of any list, for what it is worth, the following appear, to these writers, to be some of the aspects of treatments worth considering.

Qualifications of staff. Until some standards are established for qualifications of staff, it may be difficult to make highly objective judgments. Even if standards existed, e.g., for qualifications of instructors in vocational training, it might be very difficult to obtain and present the information necessary to judge whether staff met the standards. Nonetheless, even now if sufficient information were presented, it would seem likely that qualified persons could rate staff on something like a 5-point scale ranging from "not at all qualified" to "highly qualified."

Intensity. This parameter refers to the amount of treatment per unit of time, say per week in the case of vocational, educational, or counseling programs. In the case of parole supervision, intensity might refer not only to number of hours of contact per month but to the overall level of supervision made possible by access to various sources of information such as reports from employers, from family members, and from personal observation. Again, at present, it might not be possible to do much more than rate some treatments on a simple scale ranging from low to high intensity treatment. Degree of probable involvement of staff and clients in the treatment process should also be assessed. Proponents of interventions might be asked to specify how they would assure commitments of staff and motivation of clients wherever either might be in doubt, e.g., from previous studies.

Length. The time span over which treatment is carried out is another important parameter, the standards for which would obviously differ from one treatment to another. Ordinarily treatments extended over time would be considered stronger than briefer treatments, but there could be a point of maximum benefit.

Intensity-times-Length. Totality, or at least optimality, of treatment may not be completely reflected in the separate factors of intensity and length. There is probably some optimal level of total treatment for some interventions.

Focus of treatment. Generally speaking, a treatment would be considered stronger if it is focused on one or a few problems or out-

comes than if it is diffuse. For example, 24 counseling sessions devoted to alcohol problems may be a stronger treatment than 24 counseling sessions devoted to a variety of problems. Six months of training on a milling machine may be a stronger treatment than six months of general machine shop work.

Clarity of treatment plan. A treatment that had a well-developed protocol to back it up would be considered stronger than one lacking such a protocol.

Differential assignment. A treatment plan which involved assessment of the suitability of different candidates and which assigned them according to suitability would be considered a stronger treatment.

Again, there is not now a very good basis for rigorous quantification of strength of treatment on the parameters listed, but even to begin would be a notable step. Empirical work might help to delineate the specific features of each parameter and the most useful ways of scaling them. Further work could produce weights or rankings of the parameters according to their importance for different types of treatment. Qualifications of staff might be of limited importance for some treatments because of near standardization of personnel or because of limited staff input; the importance of qualifications could be great for other treatments. If agreement on weights could be reached, the end result of the strength of treatment assessment could be a single index permitting useful and revealing comparisons across different evaluations, settings, and programs, not to say an index that could be related to treatment outcomes.

A Posteriori Evaluations of Strength of Treatment

Treatment strength must often be evaluated after the fact, e.g., when reading a research report on the evaluation of a treatment. The ideas sketched above for evaluating treatments a priori are, of course, applicable a posteriori also. However, researchers could themselves improve estimates of treatment strength by planning for collection of data at the end of treatment which would indicate its strength. A fairly strong tradition in laboratory social psychology is for the use of so-called manipulation checks³⁹ to determine whether the experimental manipulation was successful and had its intended effect. Similarly it might be possible to build into some treatment evaluations procedures of data collection which would indicate something about strength of treatment. For example, Kassebaum et al.⁴⁰ administered questionnaires to inmates after some period of counseling. Responses of inmates suggested that counseling may not have been an especially strong treatment since inmates tended to view their own motives for participating as being primarily based on receiving the approval of prison administrators rather than re-

ceiving therapy. Further, weakness of treatment is suggested by the inmates' expressed doubts about the value of counseling. Meyer et al.⁴¹ found that girls in the counseling groups at a vocational high school were more likely to report that they had had sufficient opportunity to talk over their problems with an adult than were girls in the control condition, not a very discriminating finding, but one that at least indicates that treatment probably had some detectable impact. One final example is the Dickover et al.⁴² study of vocational training. A posttreatment questionnaire given to released offenders indicated, among other things, that the training they received had been with equipment quite comparable to or better than equipment found in places of employment on the outside. Thus, in that particular study the treatment was strong in at least one technical aspect.

In instances in which treatment outcomes are expected to be mediated by some intervening processes, posttreatment measures might help to determine whether those processes ever occurred. If treatment is expected to improve self-esteem, which is in turn expected to improve resistance to peer pressures, a posttreatment self-esteem measure might indicate whether the strength of treatment was sufficient to produce the expected change in the mediating variable. Kassebaum et al.⁴³ thought that favorable outcomes from counseling would be mediated by changes in inmates' commitments to criminal value systems, but at least some questionnaire evidence indicated that change in commitment never was achieved. The UDIS study⁴⁴ makes very explicit its immediate and intermediate goals, and data developed and presented by the investigators tend to support the conclusion that most of the immediate and intermediate goals of the project were achieved. In the case study of the UDIS study, had there been a failure of the treatment to show an effect, or the effect wash out, the failure presumably could not be readily attributed to a very weak treatment.

Greater efforts should be required of investigators to establish strengths of treatments, and posttreatment questionnaires, if judiciously employed, could be very helpful in accomplishing that task.

The Case for Strong Treatments

We would like to take advantage of this opportunity to make the case for testing treatments in a strong form. In any test of the effects of a treatment there are several risks to be run. One risk, for example, is that results will occasionally appear positive simply by chance. That risk is quantifiable by adopting a given level of significance, for use in statistical tests. Another risk, one much more difficult to quantify, is that a result will be regarded as negligible by chance alone. That risk is difficult to quantify because there is no single convention to operate as a guide and be-

cause setting any level for the probability of being able to detect a real effect, if it exists, requires a decision about the size of an effect one wishes to be able to detect.

We are rarely in a position in applied research settings to specify in advance just what we would consider an effect worth detecting. Cohen⁴⁴ has proposed some rules of thumb for what would be considered small, medium, and large effects in research in the behavioral sciences, but his guidelines bear no relationship to needs required for decisionmaking in the real world.

However, apart from statistical considerations, in applied research settings a powerful factor determining the probabilities of Type II errors (see page 19) is likely to be the strength of the treatment tested. Once a treatment has been tested and found wanting, there is an impetus for abandoning it, particularly if it is at all expensive, and the impetus may be imparted without respect to specific characteristics of the treatment. Thus, if a family support program appears to have no effect, further efforts to explore the usefulness of such programs may cease without anyone ever inquiring whether the program was anywhere near optimal in its design and implementation. In order to minimize the probability that Type II errors will be made, these authors urge that, at least in initial tests of treatment, the strongest possible treatment be devised and tested. If even the strongest treatment has little effect, perhaps one may be justified in abandoning that line of effort. If the strongest treatment does have an effect, then an appropriate strategy may be to proceed to "decompose" the treatment to determine whether it might be effective in a weaker or less expensive form.

An argument that is sometimes raised against testing of strongest forms of treatments is that they are like the proverbial million-dollar cure for cancer, a solution simply not affordable no matter how effective. These authors' response to that argument is twofold. First, it is usually important to know whether in principle some goal can be accomplished. If we then choose not to try to accomplish it because of the expense, we at least know why we are not pursuing it and do not delude ourselves with the mistaken, however comforting, excuse that "nothing can be done." Knowing that a goal is attainable in principle can be very important in establishing priorities. The second response to the million-dollar cure argument is that if we knew a way, however expensive, of rehabilitating criminal offenders with a high degree of certainty, we would have a very useful guide to further research. When originally developed the artificial kidney was in the million-dollar cure category, but further research and technological development have decreased the price of dialysis to a point at which it is expensive but not insufferably so. If we knew what to do about criminal offender rehabilitation, even if it involved a very strong and expensive treatment, we would know something of great value.

Another argument in favor of testing treatments in their strongest form stems from the very substantial difficulties in the field of criminal justice of planning and successfully carrying out true experiments, the methodological achievement which provides the strongest basis for causal inferences linking treatments to outcomes. In general, the larger the effect one produces, the less likely one will entertain the possibility of variables other than the treatment accounting for the effect. When there are lapses in methodology, whether at the planning stage or in actually carrying out an experiment, those lapses admit, as more or less plausible rival hypotheses, a variety of explanations for the findings, e.g., that results may be attributed to maturation, to initial differences between noncomparable experimental and control groups, to extraneous and unanticipated events coincident with treatment, and so on (for an extended discussion see Campbell and Stanley⁴⁵). Most rival hypotheses are plausible about in proportion as the results are small or limited. When there is a really sizable effect associated with a treated group, it is implausible that it will have been caused by something other than the treatment. It is, unfortunately, the case in research on offender rehabilitation that the effects likely to be achieved will be of modest size at best and therefore vulnerable to a variety of alternative explanations when the methodology is deficient. Weak treatments exacerbate the case considerably.

Consider, for example, a study summarized by Greenberg⁴⁶ from an unpublished report by Carney and Bottome.⁴⁷ Inmates who had received at least 25 weeks of individual therapy were followed up after 4 years, and their recidivism rate was compared with that expected from base expectancy scores. The expected rate was 68 percent and the actual rate was only 53 percent, so the program was considered to be highly successful. However, the base expectancy score was rather crude and may not have given a good estimate of actual risk for the therapy subjects. It is easy to believe that the 15 percent difference in recidivism rates might have been accounted for by biases in selection of subjects for therapy. Suppose, however, that the recidivism rate for the therapy group had been only 35 percent; at that level it might be easier to believe that therapy had an effect than that by some unspecifiable process a group of subjects could have been identified for whom the recidivism rate was so much lower than expected. Our intuition will usually suggest to us whether it is plausible that unspecified factors could have been powerful enough to produce effects observed. How might a much lower recidivism rate have been produced? One can only answer this question hypothetically, but presumably, assuming therapy to be effective at all, if only expert therapists were involved, if therapy sessions were held with the right frequency and over a sufficiently long period of time, perhaps with postrelease followup sessions, more favorable outcomes might have been produced.

In those instances in which treatments are found to have positive effects, arguments from weak treatments might be thought espe-

cially persuasive since the results might be taken to imply that if stronger treatments were used, even more positive outcomes would be obtained. Probably this is often the case although the relationship between any assessment of treatment strength and outcome is not likely to be linear, and perhaps not even monotonic, i.e., some treatments may be stronger than optimal and actually produce less favorable outcomes than weaker ones. In psychotherapy, for example, individual therapy would usually be considered a stronger treatment than group therapy, and a probing, insight-oriented treatment would usually be considered a stronger treatment than a more rational problem-solving or supportive therapy.

There is evidence, however, that with some kinds of patients, especially those considered borderline in adjustment or those likely to be labeled schizophrenic, group and less intensive treatments have better outcomes.⁴⁸ Intensity of parole supervision probably has some optimum value beyond which less favorable outcomes are likely.

It is also sometimes true that very weak treatments can be suspect because they are unlikely to have produced the results attributed to them. A widely known study showing the extensive effects of brief psychotherapeutic intervention on utilization of health services⁴⁹ is suspect precisely because the results seem to indicate that a single interview could be shown to have had an effect still detectable after 8 years! These writers are not aware of any quite comparable case in offender rehabilitation, but there have been many treatments tested that were so obviously weak that had they had an effect, it would have been immediately suspect as an artifact of some sort.

Are Really Strong Treatments Realistic?

The question might be raised whether the strong treatments that are encouraged in this paper are at all realistic in the world of corrections and offender rehabilitation. Is it realistic to suppose that work release programs could be developed in which prisoners would be released for substantial periods of time into meaningful and good-paying jobs in which they might continue postrelease? Is it realistic to think of counseling programs with expert staff, individualized counseling programs, intensive treatment over considerable periods of time, and the like? Could a prison develop and operate a vocational training program with highly qualified instructors, high quality training materials, carefully selected occupational specialities, good placement programs, and all the characteristics that most people regard as desirable? Perhaps not; these writers are not in a position to say. However, in the absence of strong treatments, care should be taken in drawing pessimistic conclusions about treatments that supposedly do not work.

Implementing stronger treatments might become possible if strength of treatment were made more salient as an issue and if greater pressures were put on those responsible for developing programs to plan stronger treatments. Not all increments in strength of treatment necessarily depend on availability of greater resources. For example, to the extent that a clear and thoughtful treatment protocol results in a stronger treatment, all that is required is clearer thinking and more thoughtful people to develop those protocols.

In any case, it should be made clear whether particular treatments are studied because no better ones can be imagined, or because current administrative arrangements would not permit better ones, or because some segment of society is unwilling to make the resources available that would be required for better ones.

INTEGRITY OF TREATMENT AS AN ISSUE

Even if a seemingly strong and potentially effective treatment is planned, there is no guarantee that the treatment will actually be carried out. Many things can go wrong. The staff charged with carrying out the treatment may lack qualifications, staff may not comprehend the treatment plan, interest in the treatment may wane, carelessness about implementing features of the plan may be widespread, those to be treated may fail to attend treatment sessions, and on and on. There is, then, a serious problem in evaluating treatments: there may be a very serious discrepancy between the treatment as described in the plan, the treatment that was intended, and the treatment actually delivered. We refer to this discrepancy as a problem of treatment integrity. Any conclusions about the outcome of a treatment should be limited to the treatment actually delivered rather than to the treatment as intended. The construct validity problem stemming from weakness of treatments is that the treatment actually tested may not resemble what most people have in mind when they interpret findings, e.g., brief training occurs but findings are interpreted as reflecting on extended training.

The construct validity problem involving treatment integrity is that the treatment actually delivered may not be the same as the treatment about which conclusions are drawn. If the prescribed treatment is not actually delivered, then any conclusions about efficacy will be misleading. Although it is possible that a treatment delivered could be better than the prescribed treatment, it seems unlikely in practice. In most instances any alterations that take place will be for the worse.

Quay⁵⁰ has suggested that there are four elements to be considered when evaluating the integrity of a program: the conceptual

clarity of the basis for the intervention; the service actually delivered; the personnel involved in service delivery; and the appropriateness of the treatment for those treated. These writers prefer to allocate some of Quay's considerations to the issue of strength rather than integrity of treatment, but Quay's analysis has been of substantial value in simulating consideration of both problems. When Quay examined the report by Kassebaum et al.⁵¹ with respect to his list of issues, he found the study seriously deficient on all counts. The conceptual basis for the intervention, group counseling, was very weak, a point made previously in this paper. Also, as noted previously, the personnel were not necessarily well qualified, were not well trained, and were, themselves, doubtful of the value of the intervention, and were probably poorly motivated. Participation by the inmates was scarcely voluntary, and they did not believe that the treatment was effective even at the beginning. The groups were formed rather haphazardly, and there was clearly no attempt to match inmates to treatment methods.

More clearly relevant to what is here called integrity of treatment are the observations made by Quay concerning the monitoring of services delivered. As he notes, the project staff itself identified a series of problems that characterized the groups: superficiality and lack of emotional involvement, tendencies for talkative members to monopolize discussions, lack of confidence in leaders, tendencies to focus on stories that were not analyzed but merely stimulated competitive accounts from other inmates, and tendencies for staff members to permit periods of silence as long as entire sessions because of misunderstanding of nondirective counseling or their own inability to stimulate discussion. Kassebaum et al. themselves report that "there were some, but not many, groups...in which the conduct of the sessions approximated the goals of counseling set forth in the departmental training manual" (p. 123). As if the foregoing problems were not sufficient, the groups were unstable with respect to both leaders and members. The report by Kassebaum et al. quotes the group counseling coordinator as remarking that a group could be regarded as stable if there were less than one change in leader per month, and they note that no group had the same leader for an entire year. Inmate members of groups moved in and out of groups as their prison assignments required and some stopped attending counseling sessions altogether.

Given the problems that existed in treatment, these writers conclude along with Quay that the treatment was so lacking in integrity that no conclusion at all is justified concerning the impact of group counseling on recidivism. What is surprising is that Kassebaum et al. did not make clearer the limitations on their conclusions about the usefulness of group counseling. Furthermore, it is truly astonishing that Martinson⁵² described the treatment as one "which had been able to effect an exceptionally intensive and experimentally rigorous transformation of the institutional environment." Group counseling may or may not have any usefulness in prisons, but

it would be unfortunate, to say the least, if the Kassebaum et al. study were used as a justification to abandon study of group counseling under more felicitous circumstances than existed in the California institution they studied.

It is worth noting here that not all improvements in the California group counseling program would necessarily involve greater resources, since one very major change could be accomplished by making participation in counseling truly voluntary by assuring prisoners that no record of any kind would be made of their participation or nonparticipation in a counseling program.

Were the Kassebaum et al. study the only example of a treatment gone awry, the issue of treatment integrity could be easily dispensed with. Unfortunately that study is by no means an anomaly, although a case cannot so easily be made against many other treatments. As Quay noted, the Kassebaum et al. study is remarkable for the detailed information it gives about the program under investigation. As will be noted here later, most descriptions of treatments are so deficient that it is impossible to be very certain even what the treatment was, let alone how well it was delivered. Still there are enough other instances in the literature of failures of treatment integrity to make it evident that there is cause for serious concern. Without being exhaustive, we will present several examples which illustrate the problem.

One issue of importance has to do with the level at which one conceptualizes, and ultimately examines, treatment and its integrity. There have been a number of studies of caseload size in probation and parole supervision. At an obvious level the treatment consists of assignment of probationers or parolees to different case supervisors whose caseloads differ. At that level there is probably not much cause for concern about integrity of treatment since, if a report states that some supervisors had 15 cases and others had 50, it is likely that actual caseloads approximated those figures reasonably well. However, at a more meaningful level, the treatment presumably was to involve variation in intensity of supervision, probably as reflected in amount of contact between supervisor and the probationer or parolee. If, however, case supervisors with smaller caseloads used their freedup time to attend more meetings, to write more reports, or merely not to work as hard, then more intense supervision would not be achieved, and the integrity of the treatment would be jeopardized. Glaser⁵³ notes that there is reason to believe that reduction in caseload size was not accompanied by increased intensity of supervision. This inconsistency (see Greenberg⁵⁴) may be partially explained by the variable tendency of reducing caseloads to increase intensity of supervision. In order to protect integrity of treatment in parole supervision studies, the treatment should be described and implemented in terms of amount of contact or other supervisory activities rather than simply in caseload size per se.

The Community Treatment Project (CTP) provides an example of a program whose implementation did not match its description. The CTP was supposed to deliver a variety of services to delinquents placed on probation. However, as Lerman⁵⁵ has shown, it actually involved extensive social control and detention, so much so that youths on probation experienced a good many more days in postrelease detention than did those who had served their sentences and been released outright. Moreover, the CTP youths had been classified according to maturity level and were to have been treated accordingly. However, as Lerman has shown, treatment was administered largely independently of maturity level. Lerman pointed out a number of other ways in which the CTP departed from the treatment plan as described but which were not necessarily taken into account in interpreting the findings of the study. Lerman also makes similar points with respect to the California probation subsidy program.

Some problems with maintaining the integrity of the programs probably stem from difficulties in finding staff who are motivated toward, and capable of, adhering to the requirements of a treatment plan. Jesness et al.,⁵⁶ for example, reported the results of a contingency contracting study with delinquents and noted that many field officers were not very successful in implementing the treatment, even though they had been trained to do so. Some officers seem unable to write contracts at all, and ultimately contracts were written for only 269 of 1,248 identified problem behaviors. Moreover, only 104 of the 269 contracts written were judged adequate. Obviously, contingency contracting cannot be expected to be useful if it is not properly implemented, and perhaps it really cannot be implemented with personnel likely to be available in most places. It would be unwarranted, however, to conclude that contingency contracting cannot work, since capable personnel may be available in some places and since it may be possible to improve on the training or supervision of personnel to achieve greater adherence to the demands of the treatment.

The examples of problems with treatment integrity given here are certainly not the only instances that may be found (see Saari and Selo⁵⁷). One limitation, however, is that few programs and their implementation are described sufficiently well to make it possible to know what the level of treatment integrity might have been. This is a facet of evaluation that badly needs correcting.

A Priori Assessment of Integrity of Treatment

One can assess integrity of treatment only after it has been delivered. However, it should be possible to make some assessment of the likely integrity achievable at some point prior to delivery of treatment. Essentially that assessment involves examination of the nature of the treatment and the plans devised to assure its de-

livery according to protocol. The conceptual distinctness of strength and integrity of treatment is revealed by noting that it is perfectly possible to have a weak treatment delivered exactly according to plan; in fact, as will be suggested later, weaker treatments are quite likely to have higher integrity of delivery.

The Treatment Plan. The first element to be examined is the treatment plan itself. As Quay⁵⁸ notes, a conceptually sound and clear treatment plan is more likely to be delivered with integrity than is a vague or poorly described plan--or at least so it would seem. It is possible that the failure to delineate in any clear way the treatment which is to be delivered is more a failure of description, of communication, than of concept and protocol. Certainly there could be a well conceived and sharply defined treatment concealed by dense, imprecise, or careless verbiage. Nonetheless, when a treatment plan is not clearly spelled out and comprehensible, those responsible for providing approval and funds for implementing the treatment should be on guard. In these writers' opinion, as detailed in the last section of this paper, neither approval of nor funding for the test of a treatment program should be provided until a detailed treatment plan is available.

The conceptual basis for a treatment should make clear the assumptions about target behavior on which the treatment is predicated and the links by which the processes involved in treatment are expected to result in behavior change. For example, if as staged by Kassebaum et al.⁵⁹ the effects of group counseling on recidivism are thought to be mediated by weakening the commitment of inmates to a criminal value system, the mechanisms or processes in counseling that produce that weakening should be spelled out; the way in which weakening of commitment to the value system is to reduce criminality at rather remote times in the future should also be specified. If reduced parole officer case loads are thought to lead to more intensive surveillance and hence to reduced opportunities for criminal behavior, that chain of actions and consequences should be made explicit--for one reason, so as to enhance the prospect that the original action of reducing caseload size would actually lead to closer surveillance.

There are many elements which might be listed in an adequate treatment plan; just which elements should be listed would differ depending on the type of treatment. Still, it is possible to specify some that should almost always be present. These would include: (a) the methods used to select persons for treatment(s) and how any matching of persons to treatment occurs; (b) the frequency, length, and circumstances of any intervention with individuals or groups; (c) a description of the total services to be delivered, including elapsed time in treatment; (d) the identities, experiences, and training of the staff who are to deliver treatment, including vendors, if any. These four listed items would be merely a minimal set, but even they are not described in many reports (see last section of this paper).

Staff Commitment. For most treatment interventions, the commitment of the staff to the treatment program is vital. As investigators have found^{60,61} if staff are not committed to a treatment program, it is unlikely to be well implemented and delivered. In fact, it is difficult enough to deliver many treatments even with a completely committed staff. A half-hearted commitment, let alone outright resistance, is likely to be disastrous. Ideally the description of a proposed treatment program should include positive evidence of staff commitment. Such evidence might come in the form of questionnaire responses, interview data, or even unsolicited testimonials. In the absence of any positive evidence of staff commitment to a treatment, those evaluating a proposal are forced to rely on their own judgments about the probabilities of staff enthusiasm, but such judgments should not be shirked. Where there are reasons to suspect a low level of staff commitment, an investigator should be required to produce evidence of satisfactory commitment as a condition of project approval.

Supervisory Plan. Despite the best plans and intentions, programs can fail for want of adequate supervision. A satisfactory proposal for evaluation of a treatment should include a specific plan for supervision of delivery of treatment to insure that delivery conforms to plan. For example, the supervisory plan should provide for monitoring of training personnel, for checking on adherence to treatment protocols, for verification of staff activities, and the like. To some extent an adequate program of supervision may seem to cast doubt on the qualifications and dedication of staff, but with a well-motivated and committed staff, supervision can be seen as a valuable method of verification and accountability rather than as an intrusion. Without a good plan for supervision, a project is always open to a charge of inferior treatment, especially if the treatment proves to have minimal or equivocal effects.

Documentation of Service Delivery. Related to the plan for supervision of treatment staff is the plan by which it can be documented that treatment as described did, in fact, take place. Every research proposal that is approved should include a specific plan by which it can be documented that services were delivered according to plan as described in the proposal. The plan should include mechanisms for establishing that treatment sessions took place, that those to be treated attended, that the treatment protocols were adhered to, and that the total amount of treatment can be estimated with confidence. Proposals to evaluate treatments should include such things as forms on which dates and places of sessions could be recorded, along with the names of those in attendance, provisions for tape or video recording of sessions to be analyzed later for conformance to requirements of treatment, spot-checking of treatment sessions by supervisors, interviews or questionnaires with subjects of treatment, and so on. Reputable survey research firms regularly perform checks to determine that their interviewers perform according to instructions, that interviews actually take place, etc., and it does not

seem less important that such quality control mechanisms should be applied in evaluation of treatments.

Complexity and Difficulty of Treatment. Generally speaking, we can expect that the more complicated or difficult a treatment is to administer, the greater the likelihood that its integrity will suffer. Where it is possible on a priori ground to surmise that a treatment will be complex and difficult, concern for integrity of treatment should be sharpened, and demands for protection of integrity should be increased. Unfortunately it is also the case that the more complex the treatment, the more difficult it is to establish its integrity. When a treatment is so apparently simple as delivering a monetary subsidy to parolees for a period of time, it should not be difficult to document that the treatment was carried out as planned, although even with such simple treatment it is reasonable to ask how parolees will be kept in contact with staff, how there can be assurance that parolees receiving the subsidy are in fact unemployed, and so on. Where the treatment is more complex, as in vocational training, counseling, etc., there are many more possibilities for shortcoming, and anticipations of problems in maintaining treatment integrity will be greater. When programs are very complex, e.g., requiring the delivery of a wide variety of services by diverse agencies, as in the UDIS study,⁶² apprehensions become even stronger. Patton⁶³ provides an example of an intervention that was evaluated even though it never took place at all. Some treatments may also raise doubts about probable integrity because of the likelihood of resistance within the corrections community or even the community at large, as might occur with ambitious plans for community placements, or because of infeasibility growing out of excessive resources required to implement the program.

The fact that complex or otherwise difficult programs are suspect from the beginning may seem unfair--an indictment of treatments before they have had a chance to be tried. However, the history of treatment efforts in many fields justifies greater doubts for more difficult innovations. What is required in such cases is that investigators who propose to test the efficacy of unusually complex interventions propose correspondingly careful and thorough devices to insure that departures from the treatment plan are minimized and that when departures occur they are detected and assessed. The more complex the treatment, the more detailed should be the plan for protecting treatment integrity by appropriate treatment protocols, supervision, monitoring, and motivating staff.

A Posteriori Assessment of Treatment Integrity

Unfortunately, the very concept of treatment integrity has as yet so little currency that a priori assessment is rarely done.

Equally unfortunately, a posteriori assessment is rarely done and is often very difficult if not impossible. Worse yet, the information required for a posteriori assessment of treatment integrity is infrequently available. As will be pointed out in the last section of this paper, the information necessary to assess strength and integrity of treatment is rarely given in research reports and may not even be available at all in many cases. In these writers' view, the current situation demands stringent corrective measures.

When it is possible to assess treatment integrity after the fact, usually from a final report on a project, the factors to be considered are much the same as for a priori assessment. One wants to know whether there was a treatment protocol to be followed and, if so, how sensible and detailed it was. One wants to know what provisions were made for supervision of treatment delivery and one wants documentation of delivery of services along with indicators of their quality. Data should be provided on intensity and length of treatment, attendance at sessions, and many other relevant variables.

There remains the discouraging question of what to do when a treatment is inadequately documented and appears to have produced no significant results. The solution proposed here is expensive but based on the assumption that potentially useful treatments should not be abandoned on the basis of inadequate tests. The solution proposed by these writers is that if the strength and integrity of treatments cannot be documented, reports on their outcomes should be strictly ignored and their outcomes should not be cited in any way by responsible investigators or program developers.

NEED TO DOCUMENT TREATMENTS

The need for better, more complete documentation of treatments is, in the judgment of these writers, imperative. The current situation borders, in fact, on scientific disgrace. In the course of preparing this paper, 29 papers (listed following notes) involving a variety of treatments in the general categories of work release, vocational placement, probation and parole supervision, counseling, halfway houses, and miscellaneous other interventions were reviewed for adequacy of description of the treatment. Only six papers (Jesness, 1972; Collins, 1969; Acquilano, 1972; Metametrics, 1975; State of Minnesota, 1974; and Varki, 1977) met even minimum standards of adequacy. (We judged the reports by Collins, Jesness, and Metametrics to have quite good descriptions.) It is absolutely impossible to arrive at a meaningful interpretation of the findings of a treatment evaluation unless one knows what the treatment was. One would not, from most treatment evaluation reports, have more than the foggiest notion of what the treatment was.

One of the canons of science, and presumably the treatment evaluation reports examined aspire to the status of science, is replicability. One absolutely could not replicate most studies of rehabilitative efforts in the criminal justice field because one could not even begin to replicate the treatment. It avails little that a research design is specified clearly and that outcome measures are carefully described, if the treatment description is so sketchy that a reader cannot tell what was really done in the intervention.

Perhaps it might be objected that the treatments usually being tested in the literature on rehabilitation of criminal offenders are so standard and well understood that they need not be described in each research report since there is general agreement as to what the treatment involves. Is this true? Then what specifically is denoted by a treatment which is described only as assignment of prisoners to a forestry camp (Fitzpatrick, 1967)? What minimum happenings may be assumed to have occurred when a treatment is described in not much more detail than that in a planned parole process prisoners "represent themselves" in negotiating their parole program (Wisconsin Department of Corrections, 1977)? Or when offenders are said to be released during the day to work in State hospitals "giving direct care services" and to "attend weekly meetings and in-service training," neither of which is described in any way (Farrington, 1977)?

These writers would not agree that very much at all may be assumed about treatments based on a common understanding of what they are supposed to be. Nothing should be assumed that is not specified, but more often than not so little is specified that almost nothing about a treatment should be assumed.

The first need is to have a clear and reasonably complete description of the treatment to be tested. By the implicit standards which now prevail, research reports are so sketchy that it is difficult to get more than a vague, general idea of what a treatment may have been, let alone enough information about the treatment to imagine being able to replicate it. Even to approximate most treatments presented in the literature would be a chancy prospect at best.

Consider for the moment something like an ideal research report on the evaluation of a treatment. Imagine the structure of the report, and supposing the entire report to be of about 100 pages in length, how many pages would it take to describe the treatment? One's imagination can scarcely match reality, for it seems unbelievable that reports that require 40 to 50 pages to present findings of no treatment effect should require no more than 3, 4, or 5 pages to report what the treatment was. Yet this is commonly the case.

Without meaning to vilify some particular set of investigators who are only following custom, two or three examples might be cited. Wylie (1977), for instance, reports on a treatment involving multiple services in a community based probation program; the description of treatment takes up only about 2.5 pages of a 50-page report. A report on the job placement program for ex-offenders by the Maryland Department of Corrections (1973) devotes only the following sentence of a 20-page report to description of treatment: "Job openings are discerned through various sources; directories of job openings, manual of union positions, information from the Maryland Employment Association, seminars and direct contact with individual employers" (p. 2). A report of 50 pages on probation management alternatives (Davis, 1976) gives a one-page description of treatment and gives no rationale for either of the two alternatives to traditional probation, which is assumed to be the same for everyone.

Examples of better-documented treatments are available. In the report by Metametrics (1975) on a volunteer citizen counseling program for inmates, an appendix constituting half the report contains a rather well-documented description of treatment for each site, although the overall rationale for counseling is a bit vague. Jesness' (1972) volume on a comparison of effectiveness of group counseling with transactional analysis includes two chapters totaling about 50 pages devoted to detailed descriptions of each treatment and how each is specifically employed in the program in question. Collins (1969) devotes 18 pages to a detailed description of clients and their needs and the various services which were provided. Very specific details are not given, e.g., what clients were told about money management, but the general kinds of services available are spelled out, and the program could be reasonably well replicated. The task of describing treatments is not beyond that realm of the possible and reasonable.

Of greater direct pertinence to the themes of this paper than inability to replicate adequately described treatments is that those inadequate descriptions frustrate completely any efforts to assess either the strength or integrity of treatments. Of the 29 papers reviewed, only 8 even gave basic information about length of treatment, and that was often no more than a range of length of stay in a program. Again, however, it is not impossible to report more useful data. Varki (1977) reports that groups of about 10 girls each met for 1.5 hours twice a week for 60 sessions. Metametrics (1975) at least reports that groups met once a week for 1 hour for an average of 4 weeks. Jesness (1972) also reports that the transactional analysis groups he studied met twice a week in group session and one to three times a week in community meetings. There was also an educational program from 8 a.m. to 3 p.m. every day, and average length of stay in the program was 7.6 months. Even such details as these are quite infrequent in research reports. Rarely is information given about training of staff, other than an occasional description of staff as "trained." Again, these writers believe

it imprudent to the extreme to draw conclusions about the effectiveness of a treatment when one cannot make even crude judgments about the strength at which treatment was even planned.

We could find indications of attempts to assess integrity of treatment in any way for only 6 of the 29 reports (in one or two cases the issue of integrity may have been inapplicable). In the study of use of ex-offenders as parole officer aides reported by Scott (1975), field observers did record the number of parolees seen per day, percent of aide's time spent with parolees, and how well the aide got along with parolees and fellow staff members. Wylie (1977) suggests that the multiservice community program studies involved periodic "process assessment" with feedback that included assessment of staff attitudes and morale. Gabriel (1977) studied a residential home for juveniles in which group therapy sessions were said to be conducted by personnel not highly trained, and Gabriel does report on deviations from planned treatment in the home. Higgin's (undated) study of a youth advocacy corps involved observations of advocates, one per day, which revealed that of 25 advocates, 3 were involved in inappropriate services and several were not in contact with the delinquents assigned to them. Of those delinquents finally returned to the public school, 71 percent had been contacted at least once by an advocate, with the most frequent of six services delivered being encouragement to attend school, counseling, and coordination of academic program planning. (No further definition of services is given.) In the study by Davis (1976) of probation management alternatives, probationers were asked their opinions and perceptions of the probation process. In this study it was found, quite by accident, that one officer was fabricating contact reports. Finally, Jesness (1972) states that quality of transactional analysis sessions was evaluated by ratings by supervisory personnel and by participating subjects, and there is an extensive description of the rating system. The behavior modification program was found to be failing because many wards were not receiving the personalized behavior therapy implied in their contracts, but Jesness does not say how that fact was discovered.

Our analysis of the adequacy of reports, from the standpoint of judgments of strength and integrity of treatment, strongly suggests that the deficiencies in current literature are massive. The few positive instances cited are, with but rare exception, merely steps in the right direction and should not be regarded as indicative of totally satisfactory reporting. Yet the examples do, especially in the aggregate, illustrate that adequate reporting is at least possible.

In the view of these writers, both funding agencies and journal editors are being remiss in not insisting on better, more complete reporting of the details of treatment, and both should begin to refuse to accept reports which are obviously deficient in those details, just as they would be rejected if the reports of methodology were inadequate. Still, there is indication of a more funda-

mental failure of training of investigators and of socializing them to the demands of applied science. More attention needs to be paid in training programs to issues of strength and integrity of treatment. The concerns expressed in this paper are not idiosyncratic to these authors. Boruch and Gomez⁶⁴ have provided a clear rationale for and estimate of the degradation of treatments that is almost bound to occur when they are transported from the laboratory to the field or from one site to another, and their article is worth consideration by every professional involved in service delivery and by every evaluator of professional services. In another very revealing development, Klerman,⁶⁵ Director of The Alcohol and Drug Abuse Mental Health Administration (ADAMHA), has indicated that in the future, ADAMHA will insist on preparation and submission of a detailed treatment manual for every therapy evaluated with National Institute of Mental Health (NIMH) funds so that the treatment may be replicated by others with high fidelity.

Should the criminal justice system lag behind in efforts to achieve the quality of program description that will have genuine scientific and social worth?

NOTES

1 The authors wish to acknowledge the original stimulus provided by Herbert Quay (see note 8) for the development of their own thinking as reflected in this paper. In addition they wish to thank Don Gottfredson, Glen Kercher, William Yeaton, and Stephen West for very helpful comments on earlier drafts. The issue of integrity of treatment has not been quite as neglected as strength. A recent Government Accounting Office (1978) publication on evaluation at least makes the point that integrity should be assessed but Patton (1978, noted infra.) deals with the topic at some length. M. Gottfredson (1979, noted infra.) shows how failure to take treatment integrity into account can make interventions appear worthless.

2 Thomas D. Cook and Donald T. Campbell, "The Design and Conduct of Quasi-experiments and True Experiments in Field Settings," in M. D. Dunette and J. P. Campbell, ed., Handbook of Industrial and Organizational Research (Chicago: Rand McNally, 1975), pp. 223-325.

3 Christian R. Klimt et al., "A Study of the Effects of Hypoglycemic Agents on Vascular Complications in Patients with Adult-onset Diabetes: I. Design, Methods, and Base-line Results, UGDP," Diabetes 19, supplement 2:747-783.

4 Ibid.

5 Douglas Lipton, Robert Martinson, and Judith Wilks, The Effectiveness of Correctional Treatment: A Survey of Evaluation Studies (New York: Praeger, 1975).

6 R. Martinson, "What Works? Questions and Answers about Prison Reform," Public Interest 10 (Spring 1974):22-54.

7 Gene Kassebaum, David A. Ward, and Daniel M. Wilner, Prison Treatment and Parole Survival (New York: John Wiley & Son, 1971).

8 Herbert C. Quay, "The 3 Faces of Evaluation: What Can Be Expected to Work," Criminal Justice and Behavior 4 (1977):341-354.

9 Gordon P. Waldo and Theodore G. Chiricos, "Work Release and Recidivism: An Empirical Evaluation of a Social Policy," Evaluation Quarterly 1, (1977): 87-108.

10 Ann Witte, "Work Release in No. Carolina-A Program that Works?" Law & Contemporary Problems 41, 1 (1977):230-251.

11 Waldo and Chiricos, "Work Release," 87-108.

12 Norman Holt, "Problems and Prospects of Vocational Training in a Prison Setting," in Leon G. Leiber (director) Crime and Employment Issues (Washington, D.C.: American University Law School, 1978), pp. 97-110.

13. William C. Parker, "Offender Training and Employment: The Potential of Prison Industries," in Leon G. Leiber (director) Crime and Employment Issues (Washington, D.C.: Employment and Crime Project, Institute for Advanced Studies in Justice, American University Law School, 1978), pp. 111-118.

14 Richard M. Dickover, Verner E. Maynard, and James A. Painter, A Study of Vocational Training in the California Department of Corrections, Report No. 40 (State of California: Research Division, Dept. of Corrections, n.d.)

15 Henry J. Meyer, Edgar F. Borgatta, and Wyatt C. Jones, Girls at Vocational High (New York: Russell Sage Foundation, 1965).

16 Kassebaum et al., Prison Treatment.

17 Lipton, et al., Effectiveness of Correctional Treatment.

18 Daniel Glaser, The Effectiveness of a Prison and Parole System, abridged ed., (Indianapolis: Bobbs-Merrill, 1969).

- 19 Roger Hood and Richard Sparks, Key Issues in Criminology (New York: McGraw-Hill, 1970).
- 20 David F. Greenberg, "The Correctional Effects of Corrections: A Survey of Evaluations," in D. Greenberg, ed., Correction and Punishment (Beverly Hills, Calif: Sage, 1977), pp. 111-148.
- 21 Joan Havel, SIPU--Phase 4: The Parole Outcome Study (Research Report No. 3) (Sacramento, Calif: California Department of Corrections, 1962).
- 22 Joseph E. Scott, Ex-offenders as Parole Officers (Massachusetts: D.C. Heath & Co., 1975).
- 23 John J. Berman, "An Experiment in Parole Supervision," Evaluation Quarterly 2 (1978):71-90.
- 24 Marguerite Q. Warren, "Recent Findings in the Community Treatment Project," in Correction in the Community: Alternatives to Incarceration, Monograph No. 4, (Sacramento, Calif: California Board of Corrections, 1964).
- 25 Daniel Glaser, "Achieving Better Questions: A Half Century's Progress in Correctional Research," Federal Probation, vol. 39 (1975).
- 26 Ted Palmer, "Martinson Revisited," Journal of Research in Crime and Delinquency 12, 2 (1975):133-152.
- 27 Witte, Work Release, 230-251.
- 28 Waldo and Chiricos, "Work Release," 87-108.
- 29 Joan Havel and Elaine Sulka, Special Intensive Parole Unit, Research Report 3, (Sacramento, Calif: California Department of Corrections, 1962).
- 30 David B. Wexler, "Token and Taboo: Behavior Modification, Token Economies and the Law," California Law Review 61, 1 (1973):81-109.
- 31 Greenberg, "Correctional Effects of Corrections," pp. 111-148.
- 32 Daniel Glaser, "Concern with Theory in Correctional Evaluation Research," Crime and Delinquency 23 (1977):173-179.
- 33 Meyer et al, Vocational High.
- 34 Kassebaum, et al., Prison Treatment.
- 35 John J. Fitzpatrick, An Analysis of Recidivism among Inmates Released from the Forestry Camps (Massachusetts Dept. of Corrections, 1967).
- 36 Charles A. Murray, D. Thompson, and C. B. Israel, UDIS: Deinstitutionalizing the Chronic Juvenile Offender (Washington, D.C.: American Institutes for Research, 1978).
- 37 U.S. Department of Labor, Unlocking the Second Gate: The Role of Financial Assistance in Reducing Recidivism among Ex-prisoners, R & D Monograph No. 45 (1977):43.
- 38 Peter Rossi, Richard A. Berk, and Kenneth J. Lenihan, Money, Work and Crime: A Report of an Experiment on Reducing Recidivism through Limited Financial Aid to Released Felons (Amherst, Mass: Social and Demographic Research Institute, University of Massachusetts, 1978).
- 39 Elliot Aronson and J. Merrill Carlsmith, "Experimentation in Social Psychology," in Gardner Lindzey and Elliot Aronson, ed., Handbook of Social Psychology vol. 2, 2nd ed., (Reading, Mass: Addison-Wesley, 1968), pp. 1-79.
- 40 Kassebaum et al., Prison Treatment.
- 41 Meyer et al., Vocational High.
- 42 Dickover et al., Vocational Training.
- 43 Kassebaum et al., Prison Treatment.
- 44 Jacob Cohn, Statistical Power Analysis for the Behavioral Sciences, revised ed., (New York: Academic Press, 1977).
- 45 Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Design for Research (Chicago: Rand McNally, 1966).

- 46 Greenberg, "Correctional Effects of Corrections," pp. 111-148.
- 47 Francis J. Carney and Estelle D. Bottome, "An Evaluation of a Mental Health Program in a Maximum Security Correctional Institution." (Unpublished manuscript, Massachusetts Department of Corrections, 1967).
- 48 Harry Gottesfeld, Alternatives to Psychiatric Hospitalization (New York: Gardner Press, 1977).
- 49 Nicholas A. Cummings and William T. Follete, "Brief Psychotherapy and Medical Utilization: An Eight Year Follow-up," in H. Dorken & Associates, The Professional Psychologist Today: New Developments in Law, Health Insurance, and Health Practice (San Francisco: Jossey-Bass, 1976).
- 50 Quay, 3 Faces of Evaluation, 341-354.
- 51 Kassebaum et al., Prison Treatment.
- 52 Martinson "What Works?" pp. 22-54.
- 53 Glaser, Effectiveness of Prison and Parole System.
- 54 Greenberg, "Correctional Effects of Corrections," pp. 111-148.
- 55 Paul Lerman, Community Treatment and Social Control: A Critical Analysis of Juvenile Correctional Policy (Chicago: University of Chicago Press, 1975).
- 56 Carl Jesness et al., Cooperative Behavior Demonstration Project, (Sacramento, Calif: California Youth Authority, 1975).
- 57 Rosemary C. Sarri and Elaine Selo, "Evaluation Process and Outcome in Juvenile Corrections: Musings on a Grim Tale," in Evaluation of Behavioral Programs in Community Residential and School Settings, P.O. Davidson, F. W. Clark, and L. A. Homerlyak, ed., (Champaign, Ill.: Research Press, 1974) pp. 253-302.
- 58 Quay, 3 Faces of Evaluation, 341-354.
- 59 Kassebaum et al., Prison Treatment.
- 60 Ibid.
- 61 Murray et al., Chronic Juvenile Offender.
- 62 Ibid.
- 63 Michael Q. Patton, Utilization-focused Evaluation (Beverly Hills, Calif: Sage, 1978).
- 64 Robert F. Boruch and Hernando Gomez, "Sensitivity, Bias, and Theory in Impact Evaluations," Professional Psychology 8 (1977):411-434.
- 65 Gerald Klerman, "Psychotherapy Research and Public Health Policy," (paper presented at the Society for Psychotherapy Research, 1978).

REPORTS REVIEWED

- John N. Aquilano, "Monroe County Probation Program: A Follow-up Report." Probation and Parole, Summer 1972, 55-62.
- J. A. Collins, Chicago Federal Offenders Rehabilitation Project Final Report. Springfield, Ill., 1969.
- R. D. Davis, Findings of the Probation Management Alternatives Project. LEAA grant, June 1976.
- F. Farrington and S. Wittenberg, The Concord Achievement Rehabilitation Volunteer Experience: An Evaluation. Massachusetts Department of Correction, August, 1977.
- J. J. Fitzpatrick, An Analysis of Recidivism Among Inmates Released From the Forestry Camps. Massachusetts Department of Correction, March 1967.

R. A. Gabriel, Intensive Evaluation of the Dover Odessey House, Inc., The Dover Project. New Hampshire Governor's Commission, March 1977.

Ronald Lee Hamberg, Family House: Program Evaluation. State of Washington, n.d.

J. P. Heuser, Project PICTURE: Evaluation Report No. 2, State Planning Agency of Oregon, February 1977.

P. S. Higgins, Minnesota Youth Advocacy Corps. State of Minnesota, n.d.

J. J. Hug, A First-Year Examination of Educational and Employment Temporary Absence Programs. Province of Ontario, n.d.

F. Hussey, Conestoga Cottage Project: Final Report. Pennsylvania State University, July 1975.

M. Iutcovich and J. Iutcovich, Evaluation of the Intensive Treatment Unit at Gannondale School for Girls. Northwest Institute of Research, October 1977.

W. O. Jenkins et al., A Postprison Follow-up Evaluative Study of the Alabama Work Release Program. Rehabilitation Research Foundation, March 1975.

C. F. Jesness et al., The Youth Center Research Project. California Youth Authority, July 1972.

Maryland Department of Public Safety and Correctional Services, Job Placement Program: Final Report. April 1973.

MetaMetrics Inc. Evaluation of the Offender Aid and Restoration Program. November 1975.

Mutual Agreement Program Practical Outcome Evaluation. Wisconsin Division of Corrections, July 1977.

M. Polonoskik, The Community Adjustment of Male Training School Recidivists: The Total Experience, III. Province of Ontario, October 1976.

L. H. Ricker and F. C. Walker, "Effectiveness of a Therapeutic Camping Program for Delinquent Adolescents." JSAS Catalog of Selected Documents in Psychology, 1976, 6(1), 43.

J. E. Scott, Ex-Offenders as Parole Officers, Massachusetts: D. C. Heath and Co., 1975.

G. A. Shady and S. G. Kuc, Preparing the Hard-core Disadvantaged For Employment: Social Skills Orientation Course--An Evaluation. Canadian Journal of Criminology and Corrections, 1977, 19(3), 303-309.

Minnesota Governor's Commission on Crime Prevention and Control, PI House, Inc. March 1974.

D. F. Stroman, An Evaluation of Corrections Educational Coordinator Program. Juniata College, September 1973.

C. P. Varki, Sleighton School: Project Change through Group Processes, An Evaluation Report. Sleighton School Media, January 1977.

D. Vaughn et al., "Shock Parole: A Preliminary Evaluation." International Journal of Criminology and Penology, 1976, 4, 271-284.

P. S. Venezia and R. O. Steggerda, Residential Corrections: Alternative to Incarceration. Des Moines Office of Community Development and the Iowa Crime Commission, July 1973.

W. L. Ventis, Volunteer Youth Research: An Evaluation of a Treatment and Research Program with Juvenile Probationers. NILEJC grant, n.d.

G. P. Waldo and T. G. Chiricos, "Work Release and Recidivism." Evaluation Quarterly, 1(1), February 1977, 87-108.

M. R. Wylie, Project START: Evaluation of First Grant Period. Wayne State University, March 1977.

Probation, Parole, and Correctional Programs

A Review of Their Evaluation

Lawrence A. Bennett *is Director of the Office of Program Evaluation at the National Institute of Law Enforcement and Criminal Justice, formerly of Southern Illinois University.*

INTRODUCTION

From the title one might suspect that this was an exploration along the lines of Lipton, Martinson, and Wilkes.¹ However, let the reader be assured that no such indepth, analytic study has been accomplished in this brief overview. Some opinions will be developed regarding the adequacy of research and evaluation in probation, parole, and institutional programs. No attempt at strict delineation of boundaries will be attempted--some programs such as halfway houses can be easily classified as diversion programs rather than probation and parole programs.

After outlining some of the major policy issues, a selection will be made among those programs that have suggestive findings. From these findings a framework will be drawn outlining the current state of knowledge in the field. An attempt will be made to distinguish among the levels of our knowledge--what we know, what we think we might know, and what we think is knowable, if we only had some evidence. From this set of general observations, policy implications will be derived. Areas where further research and evaluation are most urgently needed will be specified.

Areas Not Explored

While an attempt will be made to avoid comingling diversion projects into the discussion of parole, probation, and institutional programs, it seems inevitable that some diversion efforts must be considered in order to round out the picture. There are a number of areas of importance that will not be addressed. For example, the important matter of manpower and training is seen as outside the purview of this study.

Along similar lines, techniques or methods of classification will be avoided, for the most part. In probation and parole, however, classification as to levels of supervision must be dealt with. Another area deliberately skirted will be that of management and specific organizational structures.

Copyright (c) 1979 by Lawrence A. Bennett

SOME BROAD POLICY ISSUES

Some of the major social policy issues in this area, as well as some of the underlying assumptions, tend to remain largely implicit in most studies.

A basic question that has only recently been examined is whether parole should exist. The issue has been beclouded because of assumed relationships. For example, parole as a control or helping function, or both, has been seen as a part of granting release from institutional confinement. Thus it gets entangled in the controversy over determinate or indeterminate sentencing. As a part of this confounding of purposes, parole is viewed as an extension of the medical model of treatment of offenders. It is claimed, for example, that the very existence of parole rests on the assumption that offenders are deficient in certain societal coping skills and need some kind of help to get them back on track. Thus "rehabilitation" might be viewed working toward changed attitudes. Even the newfound "integration" model has similar assumptions in that the agency workers are needed to help the offender in opening the doors to sources of assistance and support, leading the individual toward social acceptance.

With determinate sentencing, however, we find parole being viewed from a different perspective. If we drop the former basic assumption that offenders are sick and assume instead that they need punishment (or "to become acquainted with the consequences of their behavior"), then the imposition of parole can no longer be a helping function but rather should be seen as a continuation of the "just deserts" associated with the penalty for the offense.

In addition, parole can be seen as an effort to insure at least some continuing protection for society. While most experienced observers would question whether a parole officer with a case load of 100, or even a caseload of 30, could be in sufficient contact with a parolee to seriously modify behavior, many strongly support the assumption that protection is there and is of considerable value. It is clear that we do not as yet know to what extent the simple existence of the parole arrangement acts as a suppressor of illegal behavior.

But our problem has not been resolved by the admission of insufficient knowledge. One decision is suggested by Manson,² that the whole concept be abandoned. Others, of course, steadfastly argue both the necessity and the value of the parole function.

Perhaps we need to shift the focus of the assessment from program effectiveness to a concern for the social impact of the existence of parole, thus helping to arrive at a tentative deci-

sion. From such a modified vantage point, probation for example, might be deemed as a valuable contributor to social values, not in terms of the usual measures of effectiveness or efficiency but rather as a mechanism that allows decisionmakers an option to make certain kinds of rulings that may be viewed as socially beneficial. An approach of this kind is seen as quite negative by some. Lerman, in his reanalysis of the evaluation of the Community Treatment Program, vehemently wrote:

...concluded that the Community Treatment Program had demonstrated that it could change the discretionary decision making behavior of adults--but had been unable to demonstrate that it had any appreciable impact on youth behavior.³

Despite negative views of this type, it can be argued⁴ that the decisionmaking process is a legitimate area of examination (perhaps even an essential part of any systematic study) and that such effort can result in greater gains toward societal goals than the detailed study of specific projects or programs.

If we accept arguendo this perspective, we can return to the question of continued existence of parole and answer a rather strong "yes." The value of parole when the indeterminate approach to setting length of institutionalization is used would appear to be unquestioned. Decisionmakers can feel more comfortable about placing individuals in a more open setting if parole is there to provide at least the semblance of behavior control as well as the possibility of aid and assistance should problems develop.

Well and good, but what about those jurisdictions with a "just deserts" determinate sentencing approach, which seem to be growing in number? Here also, it can be argued that the existence of parole can serve to reduce the length of prison stay, for few can deny the punitive aspects of supervision. Punishment may be seen to include both institution and parole time. Such views are not universally held, however. Some see only the incarceration period as the punishment. Parole, or supervision under some other euphemism, is provided to aid in the transition from prison life to the relative freedom of outside society.

Thus, accepting these kinds of assumptions, we can proceed to study the programs of parole in terms of efficiency and effectiveness.⁵ Apparently a similar position can be taken on probation.

Probation Subsidy Programs

If parole and probation should exist, how should these services be delivered?

There seems to be little question about parole being a State-level function, but a great furor can easily arise at the mere mention of State-operated probation programs. Those jurisdictions that have adopted statewide probation operations apparently find that approach desirable, but there seems to be a dearth of policy-analysis studies examining the alternatives. Given that a large number of jurisdictions provide probation services through local governmental structures, however, alternative methods of support such as full or partial State subsidy vs. local taxes will be explored.

Several efforts have been made to improve probation delivery systems by State augmentation of local resources. Programs have been initiated in California, Washington, Nevada, Minnesota, and Missouri,⁶ but no comprehensive evaluation has been attempted in most jurisdictions. However, the large-scale program of California has been subjected to considerable evaluation.⁷

Feeney et al.⁸ list seven major conclusions resulting from findings of their study:

- Subsidy did reduce significantly the commitments to State-level institutions;
- Subsidy saved the State a great deal of money in direct costs;
- Subsidy saved the State a much larger amount in terms of delayed construction of State-level institutions to provide bed space;
- The subsidy program did not increase county costs beyond that for which they were reimbursed by the program;
- The program changed only slightly the nature of the individuals remaining in secure facilities at both the State and county level;
- It did not appear to have affected welfare costs although there was some suggestion that some savings may be involved; and
- It may have affected the use of State-level diagnostic service, but the influence was difficult to unravel.

In all of this, there is little indication that the local special probation programs, designed to rehabilitate the more difficult cases, tended to be any more effective than regular, good quality probation supervision. While in a few cases⁹ the

special efforts could be shown significantly superior, in most situations the seemingly better outcomes could be attributed to differences in the kinds of clients.

Thus, we see a broad social policy that results in dollar savings to the taxpayers, does not appear to be any less effective than the former policy of direct local support, and keeps offenders in the community. This allows for the possibility of reduced welfare costs, improved employment opportunities, and conditions more conducive to meeting some of the personal, psychological needs of the individuals involved.¹⁰

The endorsement of a policy such as this does not rest on any belief that individuals will be "cured" of their criminal tendencies, nor on the idea that recidivism will be reduced, nor that crime in the community will be markedly reduced. Rather, the thrust is toward programs, policies, or procedures that increase the efficiency of the system as long as the measures of effectiveness do not drop below acceptable limits.

Deinstitutionalization

While the substitution of community-based alternatives for State-level incarceration is at the heart of all diversion programs, the systematic attempt to eliminate State-operated correctional facilities as an alternative seems to fall in a somewhat different category.

The work of Jerome Miller in virtually eliminating juvenile institutions in Massachusetts is well-known. However, of importance nearly equal to the action itself is the valuable research that accompanied it. The initial analytic work by Rutherford¹¹ was followed by the detailed work of Ohlin and his associates.¹²

Critics of the effort are quick to point out that institutionalization was not eliminated but rather that juveniles were shifted from one kind of lockup to another. Detailed study¹³ confirms that while some must be removed from the community, far fewer are controlled in this manner than was the case under earlier procedures. Other conclusions are that participants under the modified system do not perform less well in the community, and in some cases do much better, in a program which, without increases in overall total costs, maintains the majority of juvenile offenders in less restrictive settings.

Along similar lines the Unified Delinquency Intervention Services (UDIS) program of Illinois, designed to divert juveniles from State-level incarceration by providing individually determined services, was evaluated¹⁴ and was found to be very effective, in terms of the type of analysis applied, in reducing

the level of reported delinquent activity after intervention as compared with before apprehension. However, when this "suppression effect," as the authors characterize it, is compared to the use of institutionalization as an intervention technique, the two approaches were found to be nearly equal in results. Further analysis suggested that costs, when both direct and indirect effects were considered, were also quite similar. The high cost of intervention is not so startling as one might think when it is realized that the clients served are special problem cases requiring quite costly services.

Given the equal effectiveness of two alternatives, it seems quite clear that the choice should be the lesser penetration into the criminal justice apparatus. Thus, retention in settings of lowered restrictions avoids the negative effects of institutional living and the potential effects of the labeling process.

Punishment, Control, or Reintegration?

A continuing struggle centers on the basic goals of the correctional apparatus. While the pendulum seems to be swinging toward punishment and "just deserts," there are jurisdictions that continue to support the idea that their first duty is to protect society. Still others set goals that involve transition of the offender from some restricted state, such as incarceration, to a rightful place in ongoing society.

The point being raised here is not one evaluating the merits of these various positions but of the necessity of insuring that these differences of orientation are considered when attempting to analyze or evaluate probation or parole programs. The objectives deriving from each position are, or should be, quite different.

The protection of society through control, for example, will require that parole or probation officers maintain a level of supervision that will help them learn of any maladjustment that might lead to a return to wrongdoing. Such knowledge can form the basis for corrective action, whether it be referral to a mental health clinic or participating in the arrest of the client.

The kind of service to be provided under a philosophy that emphasizes punishment is quite unclear. The reintegration model, of course, would place less emphasis on surveillance, devoting most efforts toward guiding the individuals to those sources that could help fulfill unmet needs.

WHAT DO WE KNOW ?

Institutional Correctional Programs

As the material in the previous section suggests, there is an accumulation of evidence that a fair proportion of those individuals incarcerated do not appear to require the controls presently imposed. Thus the correctional apparatus has an option for modification of services or the nature of sanctions to be applied.

In the area of specific correctional programs and projects, it would appear that since the publication of the report of Lipton, Martinson, and Wilkes¹⁵ which stressed the limitations of findings supporting the effectiveness of institutional programs, there have been few attempts to examine specific projects. It is as though all key decisionmakers unconsciously accepted the "nothing works" philosophy derived from the review but continued institutional programs on faith. However, there are several projects that seem to be of value, pointing the direction for further development.

Specific Treatment Approaches

Since the report by Martinson¹⁶ there has been a tendency to reject any thought that any program might be effective in dealing with modification of attitudes or behavior, no matter what systematic treatment approach might be applied. While there should be considerable skepticism in approaching this area because of the many antitherapeutic forces within the correctional setting, care should be taken to ensure adequate evaluation of all possibilities.

On the side of finding no difference is the work of Kassebaum, Ward, and Wilner¹⁷ evaluating group counseling. The study was very soundly designed with random assignment to conditions, careful training of the group counselors, and a systematic analysis of outcome. Conducted at California Men's Colony--East Facility, four quads separated the comparison groups. Findings revealed neither differences in measured attitudes nor differences in parole performance that could be related to participation, type of program, length of exposure, voluntariness or stability of group leadership. Followup measures were recorded over a 36-month period.

Along more positive lines, the study of the comparative effectiveness of behavior modification, a token reward system, and transactional analysis, based on a system of self-adjustment development formulated by Eric Berne, by Jessness and Derisi,¹⁸ and Jessness¹⁹ suggests that treatment can have a significant impact. Again we see an outstanding example of a carefully designed program based on sound and well thought-out theoretical grounds, incorporating a solid approach to evaluation. Two institutions were chosen

with each to carry out one of the treatment modes. The staff of each was thoroughly trained in the technique and the entire institutional operation focused on the treatment approach assigned. The youthful offenders were randomly directed to each of the programs. Programs were carefully monitored to insure constant and "consistent application of the techniques and to assess within program impact." Parole followup data were provided at 3, 6, 12, 15, and 24 months after release.

The findings failed to reveal any major differences in outcome between the two institutions (meaning between the two treatment modalities) nor were there differences between treatments by subgroups within institutions. (One purpose of the study was to determine if the different treatment approaches affected different types of youthful offenders--as defined by a typology measuring interpersonal maturity levels--in a differential manner). However, there was strong suggestive evidence that recidivism was reduced for both institutions below comparable cases released from other programs. Since there was no randomly assigned control group the comparison was with baseline rates at these same institutions prior to the program, corrected for age at release. Statistically significant differences were found in each case (42.6 percent vs. 31.4 percent violation rate for Close, and 44.9 percent vs. 35.1 percent violation rate for Horton). Since there were no significant differences between the two experimental institutions, their results were combined for comparison with the 12-month outcome from two other institutions releasing youthful offenders of a similar age during the same time period. Again the difference in favor of the experimental programs was statistically significant (violation rate for comparison institutions was 42.7 percent compared to 31.1 percent for the experimental institutions). While the treatment methodology cannot be discounted, given these fairly strong indicators, one cannot avoid the speculation that perhaps the variable of total institutional commitment to a treatment approach, any treatment approach, may be a powerful influence in terms of impact.

Considering more traditional treatment, there are two studies suggestive of positive value. First the work of Carney²⁰ examined not only the effect of psychotherapy on recidivism but also looked at the contribution of volunteering for such a program. The evaluation involved 115 inmates who participated in psychotherapy and 138 who did not. Several approaches were used in analyzing the available data. First, a derived expected recidivism rate was developed against which to compare actual outcome for the therapy group. Second, applying appropriate statistical controls, therapy participants were compared with nonparticipants. Participation was found to be associated with a significantly lower rate of recidivism, defined in this study as return to any prison or jail within 4 years after release. Further, it was found that the act of volunteering accounted for a fair proportion of the observed outcome but did not eliminate the difference due to

psychotherapy. Complex relationships emerged in the search for what kinds of inmates profited most, but those with shorter criminal records did significantly better than others. Additional indices of the benefits of the treatment program were also reported--longer time to violation and reduced severity of offenses committed.

Along similar lines the work of Jew and Clannon²¹ found evidence of the positive effects of psychotherapy. Based on the parole outcome of 257 diagnosed as having personality and character disorders, this study found that those who had treatment did significantly better than a comparison group at 1 year after release. The differences disappeared after 4 years, however. The comparison group was carefully constructed to be comparable to the treated group on all variables known to be associated with recidivism. They were unable to correct, however, for the voluntarism factor.

Similarly, Jew, Kim, and Mattocks²² look at a much larger number of participants as evaluated against a similarly constructed comparison group and found that overall, the treated group (1) had fewer parolees with major problems, (2) had fewer returned to prison, and (3) had more who remained free of arrest or difficulty on parole. They found that older homicide cases and sexual offenders with short prior records seemed to profit to a greater extent than others. Both studies noted the need for supportive aftercare psychotherapy in the community to sustain treatment goals.

A treatment approach that has gained considerable attention is that of the Asklepion Community. Using a confrontive type of group interaction popularized by the Synanon program for drug abusers within a transactional analysis theoretical framework, it has been used to assist inmates to become paraprofessionals and professionals in the treatment field. Unfortunately, the evaluation process, although constantly ongoing, has not been brought to a satisfactory conclusion. The work of Paddock and Scott²³ can only be seen as a progress report. In that evaluation, reports from supervising parole officers indicated that participants had better employment-related adjustment than a comparison group, as well as lower unemployment rates. As for recidivism, the claim was made that participants were adjusting as favorably, if not more favorably, than nonparticipants. However, small numbers were involved, no real data were presented, and the length of postinstitutional exposure was not specified.

Guided Group Interaction (GGI) was a popular approach that enjoyed a wide following in the late 1960's. It is still used in a wide variety of settings but no evaluation data of any consequence has been presented. The work of Silverman²⁴ describes the problems in implementing such a program and the difficulties of evaluation but no data nor statistical analyses were reported.

However, GGI may still have a place, according to the findings of Hirschorn and Burck.²⁵ Using a post-test-only design, a study was conducted to determine if inmates could be as effective as staff in leading admission groups designed to reduce situational anxiety, to promote a positive attitude toward the program, and to develop an acceptance of future psychotherapy. Inmate counselors were selected on the basis of clinical judgment of having mature attitudes and were given 14 hours of human relations training using the Carkhuff²⁶ model. Inmates at admission were randomly assigned to inmate-led, staff-led, or control groups. It was found that there were no significant differences in terms of promoting favorable attitudes toward the program or in promoting a more positive attitude toward future psychotherapy. However, the inmate-led groups showed a significantly greater reduction in situational anxiety and a significantly greater understanding of institutional rules and regulations. It was concluded that group counseling, particularly peer-led counseling, could be effective in terms of admission processing.

The importance of stable family support may be assumed to be great, judging from the value of family visiting reported by Holt and Miller.²⁷ Generally speaking, however, the attention paid to this aspect of treatment by most correctional systems is minimal. One study, that of Filks,²⁸ may illustrate why. Evaluating a family counseling program operating over a 3-year period he found that the basic objective of increasing marriage stability and reducing recidivism was far too ambitious for the size of the effort mounted. Because of poor supervision, high turnover among counselors, and a host of other problems, few couples were able to complete the 15-week cycle of counseling. These inefficiencies increased costs beyond that which would prevail if the counseling were purchased on the open market on a fee-for-service basis. In general it was concluded that the funds could be expended to better effect in a variety of alternate ways.

The Value and Meaning of Outside Contacts

As noted above, contacts between those on the inside with those on the outside should have value, whether we approach the issue from the point of view of rehabilitation (changed attitudes) or from the concept of reintegration (introduction to the path toward social acceptance). A few studies have attempted to evaluate programs that purport to arrange for visits from interested outsiders, a program designed for those who have no outside social contacts or whose family and friends are too far away.

The work of Lewis et al.²⁹ represents one such study. Community sponsors served as members of an inmates support team, attempting to influence the individual toward a realistic release plan. Tentative conclusions were that while having a spon-

sor did not influence the institutional behavior of the resident, there was some evidence that residents with sponsors were less likely to encounter problems with the law following release. Unfortunately, no data or analysis were presented to substantiate the conclusion.

A similar program, Norfolk (Mass.) Fellowship, was evaluated by Cannon.³⁰ Investigating both postrelease behavior and the relationship between length and type of involvement and recidivism, he found a 9.4 percent lower recidivism rate for those receiving visits than for a comparison group. The comparison base could be judged less than desirable, but probably adequate. He concluded that the program was effective in reducing recidivism for regular participants, resulting in considerable monetary savings to the justice system.

While also utilizing a far from satisfactory research approach, Lawyer³¹ evaluated the Man-2-Man Job Therapy program in California. He used a nonrandom comparison group, a return-to-prison outcome criterion, and an uneven followup period ranging from 24 to 51 months. The evaluation effort was noteworthy in that a serious attempt was made to examine and discard alternate explanations of the observed findings. The comparison group was constructed by matching a number of variables and differed significantly only on two--a history of drinking and occupation prior to commitment. It seems doubtful, all other things being equated, that these two variables could account for the observed differences. The findings were that while the Man-2-Man group had a failure rate similar to the total population of releases during the same period, they were significantly superior to the comparison group (33.3 percent failure rate vs. 52.2 percent). When the voluntary aspect was considered and those who had sponsors were compared to those who had volunteered but could not be matched with sponsors, it was found the differences were even greater. However, while the number of visits was significantly and positively related to parole outcome, the receiving of letters made a similar impact.

Vocational Training

Here again we see very few studies. Vocational training programs represent a costly program effort in most institutions offering them. On the other side, it can hardly be denied that following release the individual must possess the capability to earn his own livelihood if the avoidance of illegal behavior is to be expected. One study that came to the forefront was from the Texas Department of Corrections.³² Findings suggested no impact in terms of recidivism for those participating. Over 40 percent (40.3) were in jobs related to the area of their training. This is a somewhat higher percentage than found in an earlier California study³³ where about 35 percent found jobs related to their

training. In both studies the factor most responsible for trainees being employed in the field of training was previous employment in that field.

Although the evaluation process had a number of serious limitations, the most apparent one being the lack of specification as to the period of the followup, there was one intriguing finding. Some of those going through the training program also were involved in a group-counseling effort entitled "Reality Adjustment Program." Those participating had both a higher percentage in jobs related to training and a lower rate of recidivism. Because of the limitations of the evaluation process, such a finding can be, of course, no more than suggestive.

While earlier evaluations had suggested that vocational rehabilitation training programs had little impact on either employment or recidivism,³⁴ a broad based evaluation of MDTA projects³⁵ suggests that soundly financed training can make a difference. While the outcome criterion "back in prison" used in this evaluation might be questioned in view of the short-term followup (3 and 6 months) it is at least clearly specified. Analyzing 25 projects it was found that trainees had a significantly lower recidivism rate (4.7 percent difference) than control group members. The employment situation was unclear, but since it was found that recidivism and employment success were related, the evaluator reasoned that the trained group must have had a better employment record because of its lower recidivism rate. The study concluded that a comprehensive inmate training program, properly designed and implemented, can have a significant impact on postrelease success.

Educational Interventions

There is a continuing, nagging, common-sense belief that improved education attainment will enhance the societal roles of inmates after release. Certainly it is obvious that most inmates entering prison show an academic deficiency usually averaging about two grade levels. The temptation is to relate this lack to lowered skills in coping with the environment, leading to maladjustment and crime. However, it is just as likely that the social maladjustment of the individual brought about both his failure to profit from classroom exposure and his failure to avoid illegal behavior.

At any rate the concern remains and should be examined from time to time to see if new insights can be generated. In this regard, a study by the California Youth Authority³⁶ was conducted to determine if large gains in reading skill or participation in remedial education programs would lower recidivism. Dealing with a population of over 1,000, it was found that (1) reading skill gain was not related to recidivism and (2) participation

in a remedial education program was not related to recidivism. The program was effective in dramatically raising reading level scores, so the intervention was achieved.

A side finding, however, is of interest. Reading ability was found to be related to length of satisfactory parole adjustment--those with low reading ability were more likely to recidivate within the first 3 months, while those with higher reading ability were more likely to survive beyond that period. Age and race were examined and appear to have only a negligible effect on the relationship found.

When college level programs are examined, we see mixed results. A District of Columbia study of the Lorton Prison Project,³⁷ while failing to present any analysis of findings, does present sufficient information to allow for data analysis. The statistics suggest that those participating in the college program did adjust better than those who did not participate, but the difference fell short of statistical significance.³⁸ This finding is in contrast with a California Youth Authority study³⁹ in which it was found that parole outcome was significantly better for participants than for comparable populations. In addition, self-esteem showed a significant increase. This, in itself,⁴⁰ may represent a valuable gain in view of the finding by Bennett⁴⁰ of a significant and positive relationship between measured self-esteem and parole outcome. The problem of placing heavy credence on these findings is the artificial nature of the construction of the comparison group. One of the major problems confronted by studies of this type is the variable of volunteerism, since the volunteer is likely to be more highly motivated and potentially different in some significant way from any nonvolunteer counterpart selected, no matter how well matched on a number of variables. Blumstein and his associates⁴¹ attempted to overcome the problem by drawing the comparison group from the same population, but from a time period during which the program was not available.

Inmate Participation in Programing

With the shift away from indeterminate sentencing toward flat-time or determinate sentencing, there will be a need to reexamine the voluntary aspects of program participation. It is obvious that when determination of release is believed by the inmate to be related to involvement in self-development programs, many will sign up for the record with little real interest in benefiting from the knowledge to be gained. With the removal of this belief, program participation could conceivably become more meaningful. However, for those systems continuing with the indeterminate mode, a system to increase communication and agreements about program participation would likely be helpful.

Such a program has been tried. Mutual Agreement Programming (MAP) conducted it on a demonstration basis in three states, Arizona, California, and Wisconsin. Attempts at evaluation by Rosenfeld, Abt Associates, and Susmilch⁴² suggest that the benefits may be minimal. Findings reported indicated that participants did not perform significantly better than nonparticipants either in nonrecidivism or employment. In the case of the Wisconsin application, one of the objectives was to reduce the length of incarceration. While participants did spend an average of 2.8 months less institutional time than nonparticipants, more detailed analysis revealed that this difference could largely be accounted for on the basis of the characteristics of the group volunteering for the program. The conclusions were that the program had value, "...in reducing parole board arbitrariness, reducing inmate anxiety, encouraging more realistic postrelease planning, and improving coordination of training programs with release."

Some Miscellaneous Concerns

A couple of other studies come to light that do not fit into any neat category but seem worthy of consideration.

An item of considerable importance that was poorly researched has to do with diagnostic services rendered by correctional institutions to assist judges by supplying information upon which to base sentencing. A study of a project of this type, developed in Utah, addressed the wrong issues, comparing those referred with the prison population rather than examining the decisionmaking process and the potential for diversion from prison.

The work of Dickover and Durkee⁴³ provides a more accurate and valuable portrayal of the diagnostic process, illustrating how feedback can be of considerable value in shaping behavior toward preestablished guidelines, and according to Bennett,⁴⁴ divert large numbers from State-level incarceration and improve consistency of decisionmaking.

Transition Programs

Having covered institutional projects and related efforts, we now turn our attention to probation and parole services. The plan in this section is to avoid as much as possible the ground that has been covered so thoroughly and soundly by others. The works of Adams,⁴⁵ Adams and Vetter,⁴⁶ and Banks et al.⁴⁷ will be frequently cited to provide information spanning a wide number and variety of projects. The aim will be to supplement these reports in an attempt to develop a complete picture.

Supervision Intensity and Organization

Efforts are constantly being made to reduce probation caseloads in the belief that such procedures will result in more adequate service to the clients, leading to reduced recidivism. A review of a number of studies⁴⁸ leads to the conclusion that no significant differences could be determined in recidivism rates among adult offenders in caseloads of different sizes. However, according to the review of Lipton, Martinson, and Wilkes⁴⁹ there did seem to be strong evidence that intensive probation supervision (caseload size 15 to 20) did reduce recidivism for both males and females under the age of 18.

Among those studies reviewed was the San Francisco Project (Adams et al.,⁵⁰ Robinson et al.⁵¹) but the equivalent carried out in the State of Washington was not included. The Washington State study by Carter and Dightman⁵² fell far short of standards in terms of rigorous evaluation. The aim of the project was similar to that of the San Francisco study, namely to determine if parolees could be classified in such a way that some portion of the total caseload could be responded to on an "as needed" basis, allowing for much larger caseloads in this category. Outcome comparison basis was faulty, the length of followup varied, and there was no comparison group. The resulting low violation rate of about 15 percent may well be the result of selection or some other unspecified variable.

Two broad-based evaluative efforts (Jordan and Saska⁵³ and Saska⁵⁴) suggest that there is the possibility of profit in the area of intensive supervision. While most of the projects failed to demonstrate adequate evaluations, two projects provide considerable evidence of crime reduction, educational gains, employment, and improvement in self-concept. Using project outcome against baseline data, Project New Pride (Denver) records a reduction of some 40 percentage points (22.5 percent vs. 66.7 percent) in recidivism. However, varied lengths of exposure to risk plus selection factors becloud the issue. Similar problems detract from the findings of the Providence Center (St. Louis) program. A point well made, however, emerges from the review with the conclusion that intensive supervision can be useful when efforts expended are related to the differential needs of clients--case count alone provides little understanding of the problem. This concept is echoed by the study of Davis et al.,⁵⁵ where it was found that extensive intervention among bad-risk cases is associated with a reduced reconviction rate.

In a careful study by the California Youth Authority⁵⁶ the supervision variable was clearly validated. It was found that reducing caseloads from 72:1 to 50:1 achieved the following:

- increased parole agent/parolee contact by 25 percent on the average--this was supplemented by paraprofessional contact;
- increased the availability of adequate out-of-home placements;
- reduced revocations;
- those revoked stayed on the streets an average of 2 months longer;
- no change in types of violations;
- increased slightly the number of favorable discharges from parole;
- differential caseloads seemed to work in large metro areas but not too well elsewhere; and
- close coordination with institutional staff for those close to institutions only.

However, special diagnosis of needs followed by individualized treatment often fails to affect outcome. While Nath⁵⁷ claims 75 percent success for a program of this type in Florida, a closer look at the data reveals an overall 53.4 percent failure rate, the bulk of the failures occurring during program participation. Since the evaluation's design, reporting, or both are so poor, it would be unfair to judge the intervention strategy on the basis of these findings.

Lewis et al.,⁵⁸ in evaluating the Pennsylvania probation and parole special caseloads, found that not only was there no relationship between type of supervision and outcome, no significant differences were found in level of supervision or types of services provided. They concluded, as did the previous two studies, that caseload size had little meaning unless controlled for the varying needs of the clients involved.

Despite the confusing and sometimes conflicting findings reported there seems to be a thread of evidence suggesting that special efforts can be effective. Turner⁵⁹ reports, for example, a reduction in recidivism from 53 percent to 26 percent related to early probation officer contact. Overall recidivism of 38 percent compared favorably with previous general rates of 53 percent. While the length of followup is indefinite, outcome limited to rearrest, and the comparability of the comparison base not investigated, measures were at least relatively consistent and gains of this magnitude are significant. In another study in Pennsylvania⁶⁰ objective reporting noted a wide range of outcomes among units--from 6 to 89 percent--but failed to specify any ex-

planation of these differences. In Pittsburgh the probation revocation rate was significantly lower for the intensive probation unit as compared to the general supervision program. However, when all intensive units were compared with all general supervision units, no significant differences were observed. While it is scientifically improper to attribute causality within the framework of an ex post facto analysis, at this stage of the development of our understanding of criminal justice processes it might be well to view many of our efforts as exploratory and at least examine what is happening to create such strange differences, hoping that the design of hypotheses for future testing can be somewhat more reality-based.

Along somewhat more positive lines, Perlman⁶¹ reported on an intensive probation program for youthful first offenders and found that "re-arrests and incidents of violation were very low for those in the program." Since no comparison group was identified it is, of course, difficult to tell if the results are related to the program or the selection. However, it was noted that since 27 percent of the sample had a prior record and 30 percent were over 25, the group was not quite the uninitiated population originally intended. Reported outcomes do indeed look quite favorable: with a followup period of 27 to 36 months, 82 percent had no arrests after successful completion of the program, 5 percent had been arrested on felony charges, but only 1 percent convicted. Inprogram failures were estimated to be around 6 percent with only 1 percent arrested on felony charges during the program. It is unfortunate that selection variation could not be ruled out in what appeared to be strong positive results.

Along similar lines in terms of programs with potential but without adequate evaluation to support the case, is the San Francisco misdemeanor parole program.⁶² The program was able to reduce the county jail population but because of inadequate followup data the goal of improving parole eligibility standards went unmet. Again, when one considers the many thousands serving sentences in local jails who could safely be returned to the community under supervision, it is lamentable that there are not greater efforts being made toward the accumulation of sound data in this area.

Sweet,⁶³ in studying a diversion project in Oakland County, Michigan, found that while the objective of reducing recidivism was not achieved, the program did demonstrate that offenders with prior felony convictions now imprisoned are acceptable candidates for intensive supervision in the community.

Decentralization of decisionmaking vs. centralization and standardization has been heatedly argued for some time. The attempts to assess objectively the effectiveness of either approach have been few, which is understandable considering the difficulties of evaluation involved. A study of the decentralization effort

in Pennsylvania⁶⁴ revealed that recidivism dropped from 39.9 percent to 19.9 percent. However, the evaluators were careful to point out that the effects of the separate variables could not be ascertained:

Decentralization policies, changes in regulations governing parole and a change in treatment philosophy has resulted in a lower return to prison rate. The impact of each of the above on the return to prison rate is not separable nor identifiable. (p. 5-2)

Failure in implementation has been a chronic problem in federally funded projects but one that is inadequately dealt with in the literature probably because no final report is ever produced. The work of Taylor and Masters⁶⁵ is refreshing in that specification of the independent variable--reduced caseload size--was never applied because the intrusion of outside forces caused caseloads to remain above 70, making intensive probation impossible. While the group identified as "treated" did better than the comparison base on rates of revocation and absconding, they had a higher new conviction rate. The resulting overall failure rates were quite comparable--19 percent for the treated vs. 17 percent for the comparison group.

Several studies have attempted to unravel some of the correlates of successful adjustment associated with probation or parole. Morel et al.,⁶⁶ for example, found that while presentence reports were of little value in predicting behavior on probation, there were a number of factors strongly associated, one of the most prominent being prior criminal history. Of greater importance was the finding of significant differences in post-probation adjustment related to probation adjustment. Whether their subsequent adjustment could be attributed to the probation experience is problematical--it seems more likely that those who have the potential to manage life in society can handle probation restrictions in a satisfactory manner. In any case it raises the possibility of identifying those who can be removed from probation supervision at an early date without major threat to society (in this regard see Bennett and Ziegler⁶⁷).

Browne and Markman⁶⁸ compared the perceptions of parole failures and successes. Both saw conflicts with staff and inmates as harmful; both saw counseling, vocational, and academic programs as helpful. Successful parolees tended to identify assistance as coming from events outside the prison and reported few harmful events occurring to them on parole. Those failing, on the other hand, felt that success must be related to job training and placement. Neither group felt any great pressure because of parole status. The evaluators concluded that it was an identified need for inmates to maintain meaningful relationship with those in the community. It will be noted that this finding is quite congruent with those reported on earlier in relation to family and other

kinds of visits (Holt & Miller,⁶⁹ Lewis et al.,⁷⁰ Cannon,⁷¹ and Lawyer⁷²).

In a somewhat similar approach the work of Jenkins et al.⁷³ dealt with early parole behaviors and attitudes. They developed specialized scales to assess environmental stress, maladaptive behavior, and level of productive activity. It was found that institutional vocational training had some impact immediately following release in terms of time employed and earnings, but had no demonstrable effect on recidivism. They felt, however, that employment per se seemed to serve as a prevention of recidivism. Further both environmental deprivation and maladaptive behavior were strongly related to subsequent recidivism.

Volunteers and Ex-Felons as Probation or Parole Officers

As noted in Banks et al.⁷⁴ one way that caseloads can be reduced is through the use of auxiliary staff--volunteers and paraprofessionals. Such an approach sometimes has considerable appeal in that it can often be accomplished at considerably less cost than a general reduction in caseload size. Evaluations of such efforts are extremely difficult, but an excellent product in the area is the review of the program in Lincoln, Neb., by Ku et al.⁷⁵

A detailed description of the intervention is presented, involving screening, selection, and training of volunteers. Outcome evaluation suggests dramatic reductions in recidivism for those participating in the program, in contrast to a control group experiencing the regular probation program. To insure comparability and to avoid subject selection factor, high risk participants were compared with high risk offenders in the control group. It was found that participants had 45.7 percent fewer offenses, with 55 percent of the group involved in new offending in contrast to 70 percent from regular probation supervision.

Considerable educational improvement was found in a study of 20 subjects under the supervision of volunteers evaluated by Lonergan⁷⁶ as well as a slight improvement in adjustment. The followup period was described as "several months." Attitude change was assessed by the use of the Buss Durkee Scale, a questionnaire used to measure aggressive tendencies, and significant reductions were found in 3 out of 8 areas--assaultive tendencies, suspiciousness, and guilt. Unfortunately, no comparison group data were cited, so it is difficult to determine if these drops were related to the program or to maturation effects.

Indigenous paraprofessionals are often sought to supplement the efforts of probation and parole officers. Part of the think-

ing has to do with a reduction in the sociocultural distance between counselor and client. A study of one such effort, by Beless et al.⁷⁷ reviewed the use of paraprofessionals in a Federal Probation District and found that such workers do no better or no worse than regular staff. The findings upon which this conclusion was based were derived from an examination of status as of a certain date, leaving open to conjecture the length of the follow-up period. Also rearrest is used as outcome, the inadequacy of which measure is amply demonstrated when it is noted that for those for whom dispositions are known, 57 percent of the cases were dismissed.

The use of ex-offenders as aides to parole or probation officers is believed to be even better than the use of indigenous workers in "bridging the gap" between the counselor and the person being supervised. The parole officer aide program of Ohio is one such program and has had several evaluations (Scott and Bennett,⁷⁸ Scott,⁷⁹ and Allen and Priestino⁸⁰). Despite the number of assessments, the quality of evaluation that might be desired is simply not there. The attitudes of the aides were found to be quite similar to those of regular parole officers. Inmates and parolees like being supervised by ex-offenders, and supervisors felt that the aides tended to extend themselves to assist parolees. Of even greater importance, parolees supervised by ex-offender aides had significantly fewer failures than did the clients of regular parole officers. The basis for outcome assessment, however, is a strange one--failures based against "of all those supervised during a year." Further, the 4 percent difference, while statistically significant, would, it is estimated, make a difference of 95 cases staying out of prison from the sample--190 out of the entire parole caseload. A detailed costbenefit analysis would be required to determine if differences of that size really make an economic difference.

The study by Blew et al.⁸¹ represents another evaluation of the Ohio program. While most of the report is devoted to description and guidance as to how others might apply the program, recidivism data are presented. Again the 4 percent figure emerges but this time in terms of differences in new felony convictions. Technical parole violations were about the same for those supervised by the ex-offender aide and for those under regular supervision (2 percent). Rates of absconding were also very similar (10 percent under aides, 9 percent under regular parole officers).

Along somewhat different lines the work of Connett et al.⁸² outlines how ex-offender self-help groups can be effective in providing transition programs on a contract basis. It was found that while recidivism did not vary significantly between those going through such programs and a comparison group, the services offered were provided at a considerably lower cost than similar services provided through official channels.

In a number of areas there seem to be conflicting results out of which suggestive findings point toward practices that might well have positive impact. However, either imprecise application of the treatment intervention or inadequate evaluation has obscured the potential for the program to make its full contribution.

Work Furlough

While work furlough might as easily be considered either as a correctional institution program or as a part of community-based correction, a brief overview is included here on the basis that such programs serve a transition function and, as such, can be treated conceptually as the initial phase of the parole experience.

The work of Moran et al.⁸³ illustrates one of the many shortcomings of evaluation in the field. Reporting a 35 percent failure rate for a token economy community residential program, no attempt was made to provide a basis for comparison. It may well be that the assumption employed suggested that the majority would fail in that the program was aimed at high risk offenders, "...who had a very guarded prognosis for adapting to the community."

In evaluating the Alabama work release program, Jenkins et al.⁸⁴ found that participation in work release produces a large and highly significant reduction in post-prison law encounters for males; female participants did not benefit to any significant degree.

Bass,⁸⁵ evaluating the California work furlough program, found quite different results. For those male felons completing the program the parole failure rate was quite similar to all those released during the same period, but when inprogram failures were combined with parole failures for the group, program participation could be seen as a detriment. Because inprogram failures were returned to prison, the end result was a slight increase in time served for participants.

Also the feeling was gained from interviews that many participants found themselves placed in lower status jobs that did not provide them any great opportunity to save in preparation for entering parole. This tendency of work furlough programs to perpetuate the low status of ex-offenders was noted also by Rudoff.⁸⁶

In evaluating a program in New York City, Stanton⁸⁷ reported only a 4 percent inprogram failure rate, quite low in comparison to most programs. The evaluation carefully described the characteristics of the sample, suggesting a high potential for failure. However the postprogram record suggests that 68 percent had no arrests nor did they abscond during the followup period. As is

often the case, no comparison group was employed so it is difficult to assess the extent to which these findings have meaning.

Seiter et al.⁸⁸ evaluated eight halfway houses operated by the State of Ohio and found that on an offense severity scale as well as in terms of recidivism, the houses participants did significantly better. While not discussed as a part of the impact evaluation, it was noted that the program had a 17 percent inprogram failure rate, possibly a level that would invalidate the postprogram success rate.

In an assessment of the state of the art,⁸⁹ it was pointed out that the value of a work furlough program could not be evaluated until program rationale, assumptions, and goals are identified.

Special Service Projects

As noted earlier, academic problems and delinquency seem to go together. It is not surprising then that efforts to undo delinquent tendencies often involve remedial or corrective educational efforts. The work of the State of Washington⁹⁰ seemed a laudable effort in this area. Juvenile offenders had to volunteer for the project and be accepted by a counselor and a teacher who assisted in the reintegration into the world of academic achievement. The program was quite successful in improving academic levels (average achievement was equivalent to 9 months accomplishment in the 3.6-month program) but made only minimal impact on "social adjustment." The inprogram failures accounted for 20 percent of the sample but only 16 percent could be counted as recidivism. This figure was seen as quite low considering that the sample was drawn from a high-risk group. However, no comparison group data were provided.

Group counseling as a part of the parole function has been suggested as a possible program. McCord⁹¹ reports on a project in Philadelphia involving the application of Guided Group Interaction techniques to parolees and probationers. In the case of parolees the program had no demonstrable impact. For female juvenile offenders there was a significant drop in recidivism for participants as compared to a carefully matched comparison group. Results were equivocal for male juvenile probationers, for while there were no significant differences overall, new convictions for participants were for less serious crimes. Some individual caseloads showed reduced recidivism, suggesting that conscientious application of the technique might improve on regular parole. A factor to note here is that the results obtained, as minimal as they are, came from a very light application of the independent variable, for attendance was often low.

Project New Pride of Denver⁹² has already been mentioned but follows a similar line in that remedial education plays a major role among the services offered to juvenile offenders. Adding vocational and individual counseling plus cultural enrichment services it was found that fewer clients from New Pride were arrested than from a randomly assigned control group (27 percent vs. 32 percent). Considerable cost savings are also reported.

A very similar program, Project START, was evaluated by Lewis and Lichtman.⁹³ While project participants experienced fewer arrests than nonparticipants, the difference was not statistically significant. Differences that approached significance were noted in terms of program impact in the case of those probationers who had a prior record (28 percent recidivism for those participating vs. 45 percent for others).

Assistance in job placement should enhance the probability of satisfactory parole adjustment and a program along this line was tested in Texas. Evaluated by Killinger and Archer⁹⁴ it was reported that "the project contributed to substantially reducing the recidivism rates of parolees assigned..." Closer examination raises questions as to the adequacy of the control group despite random assignment. Figure I, for example, suggests a noncomparability among the groups.

FIGURE I

PERCENTAGE OF SUBJECTS IN EVALUATION GROUP
WHOSE INTELLIGENCE QUOTIENTS WERE 85 OR BELOW

Group	Percent
Experimental	55.55
Control	72.16
Extra	83.26

Source: Killinger and Archer.⁹⁴

Baca⁹⁵ reports on an evaluation of the use of group homes for juveniles in Denver, noting nearly a 20 percent lower rate of recidivism for participants as compared with a control group. Minnesota, in reporting on its group home program for juveniles⁹⁶ found that only one-fifth completed the program successfully while 33 percent became involved in an additional offense. In the absence of an adequate comparison base it is difficult to know if this success rate is high or low. It might be quite high, given that the sample is drawn from a high risk category.

Since ability to earn a livelihood seems essential to societal adjustment, rehabilitative efforts, as previously noted, often make use of vocational training. Most such programs are placed in institutional settings but in recent years have been incorporated into parole and probation programs. A study in this area that has been subjected to a sound evaluation⁹⁷ is that of Monroe County, New York. Findings suggested that the program offered reduced the unemployment rate, aided employment satisfaction and reduced recidivism. The evaluation design involved a specified followup period (6 months), employed a control group, and defined failure as revocation, new convictions, or both.

In attempting to improve employment, Project HIRE (Helping Industry Recruit Ex-offenders) was designed to assist in placement. The evaluation conducted⁹⁸ was better than most and revealed that the program did effect both recidivism and employment. A matched comparison group was used and carefully reported.

Different approaches were taken in two evaluations of a special Monroe County probation program designed to provide academic upgrading, vocational assessment, and job placement. From Acquilano⁹⁹ the view is presented that the program was a very strong one with 88 percent of successful completions placed in jobs or training with an 80 percent job retention rate. The percent of recidivism was reported to be very low--less than 2 percent. No comparison group data were presented. This same program was evaluated by Chitren¹⁰⁰ from a costbenefit point of view. He constructed a comparison group and found that recidivism was reduced neither by participation in the program nor by increased wages. However, the fact that skills learned in the program carried over in improved jobs suggests that benefits would exceed costs over a 3-year span.

The problem of inprogram failures is clearly reflected in the evaluation of the Pennsylvania Community Treatment Services program.¹⁰¹ While the report states, "Approximately 31 percent of parolees eventually return to prison, Community Treatment Services has reduced this return rate by one-third and a potential exists for further reductions" (p. 2-3 and 2-4), if inprogram

failures are viewed as a part of total failure pattern, then participants have a failure rate of 24.48 as compared to parole failures of 23.78. Again, careful conceptualization of the process has to be determined before outcome measures can take on meaning for policy decisions.

Drug problems tend to accelerate parole or probation failures and thus some effort must be made to provide for individuals so afflicted. Two programs, Family House in the State of Washington and the Narcotics Education League's Residential Treatment program in California, have been evaluated. In the case of the first of these Hamburg¹⁰² reports that those who stay for at least 11 months refrain from undesirable behavior as well as do individuals released from prison, while those who had been involved for more than 11 months participated in no adverse incidents. Langer,¹⁰³ in evaluating the California program, reports that the project met its modest goals (having 20 percent of those clients completing the 90-day program free from further criminal justice involvement for at least 6 months). But a close examination reveals that the total failures (combining inprogram and postprogram failures) was fairly high (56.9 percent failures by 6 months following release). In the absence of a comparison group, however, not much meaning can be attached to this finding.

To assist in the transition process for youthful offenders, Minnesota developed a youth advocacy program¹⁰⁴ which involved selecting a member of a high school faculty who would visit a youth during incarceration, encourage school attendance, arrange a program of study, and generally act as a social buffer. On every index of adjustment, participants did better--attendance, grades, avoidance of new offenses, and nonreturn to the institution. However, none of these differences were significant. When all indices were combined into a single indicator the difference between the groups was significant.¹⁰⁵ Return to the institution for the controls was twice that of the experimentals (24 percent vs. 12 percent).

In an attempt to provide a comprehensive system for meeting the needs of both probationers and parolees, the Oregon Impact Program developed a series of interrelated programs that provided extensive coverage. The evaluation by Baker et al.¹⁰⁶ assessed the achievement of objectives--most were met or exceeded--and impact. Despite the fairly adequate delivery of required services, no significant differences in recidivism in relation to intensity of service were found. Similarly the utilization of subsistence support appeared unrelated to recidivism.

Financial Support and Employment

The matter of subsistence, mentioned briefly above, is often seen as a valuable assist in the transition from life in prison

to adjustment on parole. Even a cursory overview will reveal that a considerable amount of funds is required to establish oneself in a community--first and last months' rent, cleaning deposit, utility deposit, funds for transportation, and enough reserve to provide support until the first paycheck--which may be as long as a month away. In this section findings from a few studies dealing with this problem will be examined.

First a study in Connecticut providing a stipend of up to \$470 was evaluated by Feeley.¹⁰⁷ This excellent evaluation included two control groups--one receiving the standard \$20 gate money and another receiving \$50. Outcome was assessed in terms of parole violation, rearrest, parole officer assessment, and employment. While the experimental group consistently performed better than the comparison group, the differences were not statistically significant. Even in this very sound study flaws can be noted. The comparison was not with randomly assigned control groups and no analysis was supplied to reveal to what extent the comparison group might differ from the experimental group. Further it would have been helpful if an attempt had been made to examine whether, within the total sample, certain subgroups did respond differently to the support provided.

On the other hand the Direct Financial Assistance program in California appeared to have a somewhat more positive impact. Using a randomly assigned control group, evaluators (Reinarman and Miller,¹⁰⁸ Miller and Waldorf,¹⁰⁹ and Reinarman and Miller¹¹⁰) found that participants did considerably better at 6 months (80 percent satisfactory parole vs. 71 percent) and at 1 year (47 percent arrest-free vs. 40 percent). While these differences failed to reach statistical significance, there were a number of subgroups for whom the support allowance of \$80 per week for up to 12 weeks, if needed, make a significant difference. The subgroups that seemed to profit most included some groups who have a high potential for failure--low Base Expectancy (success potential), narcotic offenders, property criminals, and multiple termers were among those for whom a 10 percent or more differential outcome was required.

In the now famous study by Lenihan,¹¹¹ it was found that financial assistance was of considerable importance in reducing recidivism for property offenders but not for those involved in crimes against persons.

To summarize, it would appear that monetary assistance is not needed by many who are released and of little benefit to others. However, financial aid can have strong positive effects on certain kinds of offenders. It would seem that we have once again fallen into the trap of attempting to find a single solution to deal with the problems of all offenders despite the fact that we quickly reject any assumption that implies that all offenders are identical.

THE ADEQUACY OF EVALUATION

The assessment of the adequacy of the evaluations that form a part of a wide range of funded projects must be judged as "poor." Many programs do not go beyond a description of the activities performed, while others interject a subjective opinion that the services rendered were "valuable" or that the project is making a positive contribution, without supporting evidence.

The next level of evaluation tries to determine in a somewhat more objective way how people associated with the program feel about it. Thus, staff and clients are interviewed, on an as-available basis, and their views summarized.

For those few that attempt data collection and data analysis the level of quality and adequacy of treatment fall far short of any acceptable standard. Most of the reports examined concerned projects started in the 1970's with the evaluation component conducted around 1975. While an excuse might be generated for the evaluations of the 60's that the state of the art had not been sufficiently developed to allow for a broad general application, no such excuse would seem acceptable at this late date. Yes, there are still numerous difficult issues to be dealt with and the present level of sophistication of design and analysis does not encompass all evaluation situations, but the reported evaluation efforts often do not appear to represent even a reasonable effort.

Fortunately for the field and for society there are a number of studies that incorporate outstanding evaluation efforts with highly innovative approaches to very difficult problems. It is from these findings that we can be relatively comfortable about suggesting the possibility of application of programs, projects, and approaches on a somewhat more general basis.

But with regard to the widespread inadequacies of evaluations we cannot simply say that things will get better in the future, basing our belief on wishful thinking. It is strongly felt that the specific deficiencies have to be depicted in a straightforward manner so that at least some few project managers, some few evaluators, can learn from the past errors of their colleagues. Thus an attempt will be made to briefly highlight the major shortcomings seen in many of the evaluation efforts reviewed. The concerns raised here have not been newly discovered but receive the emphasis of repetition because of their extreme importance. Banks et al.¹¹² raise many of the same issues in their evaluation of intensive special probation projects. Matthews¹¹³ suggested that many of the same problems have been occurring for some time.

Lack of a Comparison Basis

At the risk of being accused of name dropping, the sayings of Richard A. McGee come to mind when this subject is introduced. When confronted with the casual inquiry, "How's your wife?" he would often respond, "compared to what?--to how she was yesterday?--to how she was last year?--to some movie star?--to Granny Goose?" All of which illustrates that we are in a comparative mode when we attempt to evaluate.

Many evaluation studies fail to specify any sort of a comparison group at all.

Some evaluators attempted to develop a randomly assigned control group, but in the maelstrom of operating agencies such procedures often break down. Even so, the effort is applauded, for if it could be brought to successful fruition and proved to be a nonbiased sample, the generality of the findings could be greatly enhanced.

The more typical approach has been to select a somewhat similar group or sample as a comparison base. It is often erroneously referred to as a "control" group, lending further confusion to attempts at interpretation of findings. Caution must be applied in dealing with the findings and results of such studies.

Subject as His Own Control

Various schemes have been attempted to make use of the approach that considers each subject as his or her own control. The level of functioning prior to the introduction of an intervention strategy is compared to the level of functioning after the intervention. Such procedures, if well done, are considerably better than a simple rough comparison among groups that are presumed to be equivalent, inasmuch as the matter of difference before the intervention has been handled. An emerging problem, however, has to do with how long before and after one observes or records behavior--the selection of different time periods can markedly affect the resulting findings.

Another approach rests on quite a different assumption. Expected behavior, against which one might compare the postintervention observed behavior, is determined by a prediction based on an extrapolation of the pretreatment behavior. Such an approach, while only occasionally applied, enhances the possibility of the findings being positive. There are a number of dangers involved in this approach, the most obvious of which is the lack of precision with which predictions can be made.

The best procedure available in the area of the individual as the control is that of the base expectancy approach suggested

CONTINUED

1 OF 5

by Mannheim and Wilkins¹¹⁴ and Gottfredson and Beverly.¹¹⁵ In this procedure variables known to be associated with subsequent adjustment are tested against several large samples to determine the relative weights to be given each variable. Usually five to nine variables provide the most economic prediction, typically accounting for about 20 to 25 percent of the variance. The resulting scale is then applied to the sample being studied. The extent to which deviations from expected outcome are observed is the extent to which treatment impact may be assumed. This is a powerful technique that can be applied when comparison groups are difficult to construct, also controlling for intake variation between groups when group comparisons are attempted.

The Length of Followup

There seems to be a strong tendency to form a firm attachment to some date and develop "as of" outcome measures. The finding may read something like, "As of April 17, only 10 parolees (17.4 percent) had been returned to prison." What usually isn't known is how many had been released only one month previously.

There are times, of course, when these kinds of data are all that are available. It approaches acceptability when similar measures are taken at specific times over an extended period. For example, if status is recorded "as of" the end of each year, rough comparisons can be made between years.

A variation of the "as of" approach occurs most often in studies dealing with probation. That is to report outcome in terms of nature of terminations, whether by successful completion of the probation obligation, by revocation and placement in prison, or by revocation and continuation on probation, during a specified period. Thus, a change attributed to some program intervention might read something like, "The percentage of those successfully discharged increased from 74.3 percent in 1975 to 81.2 percent in 1976." As can be seen, the procedure has a number of flaws that make interpretation almost impossible. We do not know how long people have been under supervision prior to termination. Thus, some of the group may have been on probation 3 years or more while others for only a matter of a few months before termination occurs. While the standards recommended by the National Advisory Commission on Criminal Justice Standards and Goals of specified periods of followup of 6 months, 1 year, 2 years, and 3 years may be excessive, certainly there needs to be a designated and determined period of time of followup if we are even to approach any sort of comparability among the findings from different projects and studies.

The Application of Statistics

While the application of statistical analysis is often lacking, the absence may be less devastating than the misapplication of statistics. It is unfortunate that either of these conditions exists, but the problem is far overshadowed by the serious deficiencies in evaluation design. Several points must be emphasized if evaluation is ever to achieve a role in decisionmaking anywhere near its potential.

First, statistical analysis should be applied to determine to what extent the observed findings might be the result of chance variation. Fortunately, the use of tests of statistical significance has become much more common in recent years, but there are still reports that do not provide this kind of an analysis nor even seem to feel the need of it.

Along this line, however, one must also guard against overreliance on statistical analysis, for when dealing with a large number of subjects or observations, statistically significant differences can be found that have little practical meaning. For example, two groups may be significantly different on a scale measuring psychological adjustment. Closer inspection may reveal, however, that the difference in mean scores is 2.4 on a scale with a range of 6-100. Thus the meaning of the difference is of little consequence.

The other area of statistical inadequacy is related to level of measurement. Statistical techniques rest on a variety of assumptions and are designed for application to certain kinds of situations and certain kinds of measurements. While reality constraints sometimes force modification of techniques that violate assumptions, such deviations should be explained. The application of parametric statistics to analyses that involve only nominal and ordinal data should be carefully avoided. Nonparametric techniques are available for almost any known applications.

Outcome Measures Need Refinement

As noted by Banks et al.,¹¹⁶ time factors seriously hamper full and adequate evaluation of projects. Because of the short time frame of most projects, handy indices are sought rather than sound ones. For example, arrest data are often chosen with the rationalization that they represent the most objective data available. Further, with a brief followup of, say, 6 months, there is insufficient time for adjudication in most jurisdictions so convictions cannot be considered. However, arrest data are a very poor indicator of behavior although they are a very good indicator of level of police activity. Thus, the evaluation design must specify the objective of the evaluation. For example, in one study the percent arrested was slightly higher for the group des-

ignated as treated compared to the nontreated comparison base, but when convictions were considered, the treated group was significantly lower.

The whole matter of recidivism as a measure of outcome is fraught with difficulties but is a matter that must be faced squarely. While recidivism can range from arrest, conviction, absconcion, return to confinement, etc., and most projects do a fair job of specifying which of these outcomes will be used, the fact of the matter is that probably all of these outcomes should be applied. In addition there should also be behavioral indices such as residential stability and employment, although only one is addressed (absconding). A few studies have attempted to classify outcomes in terms of the behavior involved, rather than official sanctions. While procedures of this type approximate reality to a greater extent than adjudication and disposition they nevertheless include the danger of some loss of objectivity. In summary, it would appear that a comprehensive outcome evaluation would account for all clients over a specified period of observation in terms of official actions, dispositions, and behavior. While the results often tend to be confusing because they are complex, a more realistic picture emerges, for over-simplification of complex phenomenon can be dangerous.

Too often, arrests are counted only within the immediate community of the project, reconfinement counted only when the individual is returned to the institution from which released. In this age of high levels of mobility, offending behavior can occur almost anywhere. And as our technology increases, records are available almost everywhere. There needs to be greater effort made to insure that all members of a release cohort or a post-treatment group are accounted for whether in the jurisdiction of immediate concern or elsewhere.

The Matter of Inprogram Failures

Far too little attention is paid to that portion of a sample that fails during the process of going through the program or the treatment process being evaluated. Depending on the objectives of the program, inprogram failures may be more important than the successes of those satisfactorily completing the program. For example, where inprogram failures are reported, the percentage may range from 5 to 30 for most programs, but isolated instances have been reported as high as 75 percent. If these failures are added to posttreatment failures, quite a different picture is portrayed than if only those completing the program are considered. This becomes especially crucial in situations where the intervention is designed to ease the transition between institutional life and parole. If the combined inprogram and postprogram rate exceeds the parole failure rate for the same period, it would not appear that the program is meeting its objectives. The admonition

then is to make sure that inprogram failures are recorded and reported and that they are taken into consideration when evaluating outcomes and assessing program achievement in terms of objectives.

The matter, while often overlooked, has not been totally ignored. Cavior and Cohen,¹¹⁷ for example, stress the need to carefully account for inprogram outcome measures as well as the post-program effects.

General Problems in Evaluation

Why is it that evaluation tends to be revealed in such a poor light? There are several interrelated aspects that appear to account for a great deal of the difficulty. First of all, evaluation is not usually an integral part of project planning. It is usually an "add on," sometimes in the thought process and often in actuality. Evaluation then enters the picture after much of the action has occurred and must manage the measurement process on a makeshift basis. Second, the objectives developed often defy measurement. As can be seen, if evaluation is built into initial planning, the aims of the project can be constructed in a form encouraging objective evaluation. Third, data systems are usually in a developmental state. Again early planning for evaluation would insure the systematic recording and reporting of data of concern to the management of the project or program and of value for evaluation. A number of projects reviewed apparently failed to maintain even an adequate system of recording events essential for good bookkeeping practices, falling far short of evaluation needs.

Hopeful Signs for the Future

Fortunately there have been a number of studies completed that have applied sound principles of evaluation. From these studies we can use the findings to build toward a solid body of knowledge.

Also in the plus column is the emergence of a number of replications of earlier studies. This occurrence is a mark of maturity in the field in that single study results may not represent the level of stability of findings that might be desired.

The work of Martinson¹¹⁸ led him to the conclusion:

This is not to say that we found no instances of success or partial success; it is only to say that these instances have been isolated, producing no clear pattern to indicate the efficacy of any particular method of treatment.

The outlook seems somewhat more hopeful at this time. There now appear to be a few findings that tend to support one another. Other studies offer suggestive evidence of direction. While some of these latter reports are admittedly imperfect it seems we are at a point where we can begin the process of the formulation of a tentative body of knowledge.

A danger here is the tendency to be overcritical. In the best evaluations conducted flaws can be found, especially if the process is subjected to intensely critical evaluation. While we await a comprehensive knowledge base, action will occur--people will be making decisions; they will be planning new programs. The concern that must be addressed is whether evaluation results will form a part of that decisionmaking process.

An additional problem is the difficulty in dealing statistically with multiple outcomes. This difficulty has been particularly emphasized by Williams.¹¹⁹ One attempt at a solution for this dilemma is proposed by Keller and Carlson¹²⁰ wherein a single index is suggested. Unfortunately, their proposal does not progress very far beyond the Wolfgang-Sellin Index of Offense Severity,¹²¹ and fails to consider other adjustment factors such as employment level, contribution to family support, and community action.

CONCLUSIONS

From this cursory review there seem to be some converging streams of evidence, while not yet completely validated, that strongly suggest the beginnings of a knowledge base upon which policy consideration can be based.

What Do We Know?

In the Institutions

- We know that a great many individuals now incarcerated need not be held in the level of control now imposed.¹²²
- We know that almost no programs work for all clients and only a very few work for some clients.¹²³
- It seems that well-planned treatment programs, totally applied, work to some extent.¹²⁴
- Visits and other family contacts have a positive relationship with postinstitutional adjustment.¹²⁵

- Contacts of a supportive nature from people outside have a positive relationship with after-prison adjustment.¹²⁶
- Vocational training carefully planned and coordinated with job counseling and placement efforts is related to lower rates of recidivism.¹²⁷
- While education in and of itself does not seem to be related to favorable parole outcome, participation in college level classes seems related to positive adjustment after release.¹²⁸
- Mutual agreement programming efforts do not seem to achieve the desired results.¹²⁹

During Transition

- Diagnostic services provided by correctional institutions to the court can be of value to the court and result in reduced commitment to State-level institutions.¹³⁰

In the Parole and Probation Setting

- Small caseloads are associated with lower violation rates in the case of juveniles.¹³¹
- A fair proportion (25 to 30 percent) of those under supervision need only minimal supervision for a short period.¹³²
- Carefully planned and executed intensive supervision programs can be effective in reducing failure rates with selected kinds of offenders.¹³³
- Some special types of well-executed service delivery programs are associated with improved adjustment.¹³⁴
- Volunteers can augment probation and parole services without a decrement in effectiveness.¹³⁵
- Ex-offenders can make a valuable contribution as aides to parole or probation officers and function at a level of impact comparable to that of regular officers.¹³⁶

What Do We Think We Know?

A number of studies have provided positive results but, because of inadequate evaluation designs, their findings must be viewed as very tentative and suggestive only. Other studies reveal evidence that falls short of statistical significance, but for a variety of reasons it is felt that positive results could be obtained if the program implementation were improved or if certain obstacles to evaluation were removed. Because of

the distinct possibility of subjective bias, the tentative findings listed here must be viewed with considerable caution if used for the planning of policy.

The following items represent, then, areas of potential value:

- Work furlough programs may be a valuable adjunct to institutional programs for some individuals if objectives are carefully delineated and implementation closely monitored.¹³⁷
- Financial support equivalent to unemployment insurance might be of value with certain kinds of offenders during the transition period following incarceration.¹³⁸
- Psychotherapy seems to have value for a carefully selected group of offenders.¹³⁹
- The positive impact of psychotherapy can be greatly enhanced if supportive continuity of treatment can be provided.¹⁴⁰
- Special vocational training, job counseling and remedial education may have a beneficial effect on some individuals when provided as a part of field services.¹⁴¹
- Ex-offenders can be effective in providing contract services training, counseling, job referral and placement, etc., to augment regular parole or probation programs.¹⁴²
- Decentralized decisionmaking in field services increases the probability of satisfactory adjustment on the part of clients.¹⁴³
- Guided Group Interaction and other counseling techniques have potential for improving client adjustment.¹⁴⁴
- It may be possible to identify a fair proportion (15-20 percent) of those under supervision who will have a low recidivism rate whether supervised or not and for whom direct release would be the most economical procedure.¹⁴⁵
- Drug treatment programs can help a few addicts some.¹⁴⁶

What Do We Need to Know?

Before too many far-reaching policies are developed and placed into operation there are a number of areas where the potential for sound knowledge is within the realm of possibility considering the state of the art. A number of areas suggest themselves as essential before further drastic action is contemplated.

- We need to know if the presence or absence of parole makes a difference in subsequent adjustment. Only one study tends to attempt to deal with this issue¹⁴⁷ but it was limited to juveniles. We need to know what aspect of parole makes a difference--surveillance, support, service, threat, or social contact. With jurisdictions contemplating the elimination of parole, it would appear that the time is ripe for a complex evaluation design with random assignment to conditions that would lead to evidence of considerable reliability.
- Similar studies are needed in probation.
- The combination of brief periods of incarceration (3 to 6 months) in combination with probation is often applied. Studies should be designed to determine if such an approach is associated with subsequent positive adjustment or if it is contradictory to the intent of probation.
- It would be helpful if we could develop classification criteria that would assist decisionmakers in selecting those offenders for whom neither incarceration nor supervision are required for the protection of society. Carefully planned studies could supply this information.¹⁴⁸
- Since financial assistance seems to be of value only to certain groups of individuals being released into the community, further exploration appears essential to insure maximum impact from the application of limited resources.
- More detailed analyses are suggested to determine how and why participation in college programs in the institution improves postrelease adjustment.
- Further study of parole from county jails should be undertaken. If a clear and acceptable system can be developed, the potential exists for a sizable reduction in the population of such facilities.¹⁴⁹

Potential Results

If a system applied all of the reasonably solid and tentative findings in developing a comprehensive set of policies for operation, there is little assurance that great reductions in recidivism would be achieved. It is also unlikely the crime rate would be markedly reduced. There would be some reduction in the failure rate but, more importantly, the participants would be dealt with in a more efficient manner and at a lower cost. Further, a larger proportion of offenders would be safely functioning in the community, offering the potential for contribution to family support, for the payment of taxes, and reintegrating into society with fewer incapacitating side effects.

NOTES

1 Douglas Lipton, Robert Martinson, and Judith Wilkes, The Effectiveness of Correctional Treatment (New York: Praeger, 1975).

2 John R. Manson, "Do We Need Parole?" Corrections Magazine, (May 1977).

3 Paul Lerman, Community Treatment and Social Control--A Critical Analysis of Juvenile Correctional Policy (Chicago: University of Chicago Press, 1975), p. 5.

4 Lawrence A. Bennett, "Should We Change the Offender or the System?" Crime and Delinquency 19 (1973):332-342.

5 Lawrence A. Bennett, "Evaluation, Feedback, and Policy" (Paper presented at the National Conference on Criminal Justice Evaluation, Washington, D.C., February 1977).

6 Lawrence A. Bennett, "Recent Trends in Juvenile Corrections Research and Evaluation" (Paper presented at the annual meeting of the American Correctional Association, Portland, Oreg., August 1978).

7 Floyd Feeney, Travis Hirschi, and Edwin M. Lemert, Evaluation of the California Probation Subsidy Program, Volume 6, Summary (Davis, Calif: Center on Administration of Criminal Justice, 1975); Dennis A. Johns, Philip D. White, and Sheldon A. Berkowitz, California's Probation Subsidy Program--A Progress Report to the Legislature--Report Number 2 (Sacramento, Calif.: California Youth Authority, 1975); Gail S. Monkman, Cost-Benefit Analysis--Three Applications to Corrections...Probation Subsidy, Diversion, Employment (Washington D.C.: American Bar Association, 1974).

8 Feeney, Hirschi, and Lemert, California Probation Subsidy Program.

9 Dennis A. Johns, Philip D. White, and Sheldon A. Berkowitz, California's Probation Subsidy Program--A Progress Report to the Legislature--Report Number 2 (Sacramento, Calif.: California Youth Authority, 1975).

10 D. Lollar, "Institutionalization: A Viable Alternative in the Treatment of Delinquent Adolescents?" Journal of Research in Crime and Delinquency 10 (1973):195-202.

11 Andrew Rutherford, "The Dissolution of the Training Schools in Massachusetts," in Calvert R. Dodge, ed., A Nation without Prisons: Alternatives to Incarceration (Lexington, Mass.: Lexington Books, 1975).

12 Lloyd E. Ohlin, A. D. Miller, and R. B. Coates, Juvenile Correctional Reform in Massachusetts (Cambridge, Mass.: Harvard University, 1977).

13 Ibid.

14 C. A. Murray, D. Thomson, and C. B. Israel, UDIS: Deinstitutionalization of the Chronic Juvenile Offender (Chicago: Illinois Law Enforcement Commission, 1978).

15 Lipton, Martinson, and Wilkes, Correctional Treatment.

16 Robert Martinson, "What Works? Questions and Answers about Prison Reform," in John A. Gardiner and Michael A. Mulkey, ed., Crime and Criminal Justice (Lexington, Mass.: D. C. Heath & Co).

17 Gene G. Kassebaum, David A. Ward, and Donald M. Wilner, Prison Treatment and Parole Survival--An Empirical Assessment (New York: John Wiley & Sons, 1971).

18 Carl F. Jessness and W. J. Derisi, Youth Center Research Project (Rockville, Md.: National Institute of Mental Health, 1972).

19 Carl F. Jessness, Comparative Effectiveness of Behavior Modification and Transactional Analysis Programs for Delinquents (Sacramento, Calif.: American Justice Institute, 1974).

20 Francis J. Carney, "Evaluation of Psychotherapy in a Maximum Security Prison," Seminars in Psychiatry 3 (August 1971):363-375.

21 Charles Jew and T. Laurence Clannon, "Effectiveness of Group Psychotherapy in a Correctional Institution," American Journal of Psychiatry 129 (November 1972): 602-605.

22 Charles C. Jew, Luke I. Kim, and Arthur L. Mattocks, Effectiveness of Group Psychotherapy with Character Disordered Prisoners, Research Report 56 (Sacramento, Calif.: California Department of Corrections, 1975.)

23 Arthur L. Paddock and Ronald L. Scott, Evaluation of the Increment Effectiveness of the Asklepion Community on Post-incarceration Adjustment of Incarcerated Inmates (Rockville, Md.: NCJRS Microfiche Program, n.d.).

24 M. Silverman, Guided Group Interaction as an Instrument of Correctional Treatment--Evaluation and Standards (Washington, D.C.: NILECJ, 1973).

25 Stephen I. Hirschorn and Harmon D. Burck, "Utilizing Inmates as Group Leaders in the Admissions Phase of Incarceration," Offender Rehabilitation 2 (1977):45-52.

26 Robert R. Carkhuff, Helping and Human Relations (New York: Holt, Rinehart & Winston, 1969).

27 Norman Holt and Don Miller, Explorations in Inmate-Family Relationships Research Report 46 (Sacramento, California Department of Corrections, January, 1972).

28 Alfred I. Fiks, Family Counseling for Inmates and Their Wives--Evaluation Report (District of Columbia: Office of Criminal Justice Plans and Analysis, 1973).

29 Morgan V. Lewis et al., Community Sponsors and Support Teams in Corrections--An Experiment and Its Evaluation (College Park, Pa.: Pennsylvania State University, 1974).

30 Tom Cannon, Evaluation of the Norfolk Fellowship (Boston: Massachusetts Department of Corrections, 1975).

31 Alfred Gordon Lawyer, An Evaluation of the Man-2-Man Program (Hayward, Calif.: Man-2-Man Job Therapy, 1978).

32 Charles M. Whitson et al., Vocational Follow-up Project--Final Report (Windham School District, Texas Department of Corrections, 1974).

33 Robert M. Dickover, Verner E. Maynard, and James A. Painter, A Study of Vocational Training in the California Department of Corrections, Research Report 4 (Sacramento, Calif.: California Department of Corrections, December 1970).

34 Merlyn Matthews, "Methodology and Findings of the Federal Offenders Rehabilitations Program," in Stanley J. Cohn and William B. McMahon, ed., Proceedings of the Third National Symposium on Law Enforcement Science and Technology (Chicago: ITT Research Institute, 1970), pp. 103-107.

35 John McDonnell, Evaluation of the Training Provided in Correctional Institutions under the Manpower Development and Training Act--Section 251--Final Report--Impact of Training Program on Trainees (Cambridge, Mass.: Abt. Associates, 1971).

36 California Youth Authority, Reading and Recidivism (Sacramento, Calif.: California Council on Criminal Justice, 1974).

37 District of Columbia Department of Corrections, Lorton Prison College Project--Third Year Final Report (Washington, D.C. 1973).

38 $\chi^2=1.07$, $df=1$.

39 Keith S. Griffiths, Joachim P. Seckel and Betty J. Raab, Assessment of Junior College Program for Youthful Offenders in an Institution, Research Report 65, (Sacramento, Calif.: California Youth Authority, June, 1973).

40 Bennett, "Self-Esteem and Parole Outcome," Criminology 12 (1974):346-360.

41 Al Blumstein and B. Pahl, Evaluation of a College-level Education Program in a Prison--Methodology and Data Base--First Progress Report (Pittsburgh, Pa.: Carnegie Mellon University, 1971).

42 Anne H. Rosenfeld, Mutual Agreement Programming Program Outcomes in the Initial Demonstration State--An Evaluative Summary of Research (College Park, Md.: American Correctional Association, 1975); Abt Associates, Wisconsin Mutual Agreement Program--Exemplary Project Validation Report (Washington, D.C.: National Institute of Law Enforcement and Criminal Justice, 1975); C. E. Susmilch, Impact of Mutual Agreement Program on the Amount of Time Served in Wisconsin Correctional Institutions--Special Report (Madison, Wis.: Wisconsin Council on Criminal Justice, 1977).

43 Robert M. Dickover and Kay A. Durkee, Guidance in Sentencing: The Pre-Sentence Diagnostic Observation Program, Research Report 53 (Sacramento, Calif.: California Department of Corrections, September 1974).

44 Bennett, "Evaluation, Feedback, and Policy."

45 Stuart Adams, "Some Findings from Correctional Caseload Research," Federal Probation 31, 4 (1967):48-57.

46 Reed Adams and Harold J. Vetter, "Effectiveness of Probation Caseload Sizes: A Review of the Empirical Literature," Criminology 8 (1971) 4:333-343.

47 J. Banks et al., Summary Phase I Evaluation of Intensive Special Probation Projects, NILECJ, LEAA (Washington, D.C.: Government Printing Office, 1977).

48 Ibid.

49 Lipton, Martinson, and Wilkes, Correctional Treatment.

50 William P. Adams, Paul M. Chandler, and M. G. Neithercutt, "San Francisco Project--A Critique," Federal Probation 35 (1971) 4:45-53.

51 James Robinson et al., The San Francisco Project: A Study of Federal Probation: Final Report, Research Report 14 (Berkeley, Calif.: University of California School of Criminology, April 1969).

52 Robert M. Carter and Cameron R. Dightman, Washington Description and Evaluation of the Minimum Service Caseloads in the Division of Probation and Parole, Research Report 5 (Olympia, Wash.: Washington Department of Institutions, 1969).

53 Frank R. Jordan and Joseph H. Sasfy, Review of Selected Issues and Research Findings Related to Probation and Parole--National Impact Program Evaluation (McLean, Va.: Mitre Corporation, 1974).

54 Joseph H. Sasfy, Assumptions Research in Probation and Parole--Initial Description of Client, Worker, and Project Variables--National Impact Program Evaluation (McLean, Va.: Mitre Corporation, 1975).

55 Martin Davis et al., Social Work in the Environment (London: Her Majesty's Stationery Office, 1974).

56 California Youth Authority Increased Parole Effectiveness Program--Final Report (Sacramento, Calif.: California Council on Criminal Justice, 1974).

58 Sumil B. Nath, Evaluation--Multiphasic Diagnostic and Treatment Program (Florida) (Tallahasee, Fla.: State of Florida, Parole and Probation Commission, 1975).

59 Morgan V. Lewis, Barbara J. Clark, and Jacob J. Kaufman, Evaluation of the Specialized Units Project of the Pennsylvania Board of Probation and Parole (University Park, Pa.: Pennsylvania State University, 1974).

60 Stanley Turner, Philadelphia--Court of Common Pleas--Evaluation of Community Related Institutional Program (CRIP) (Harrisburg, Pa.: Pennsylvania Governor's Justice Commission, 1974).

61 Pennsylvania Board of Probation and Parole, Development of Specialized Units of the Pennsylvania Board of Probation and Parole--Final Report (Harrisburg, Pa.: Pennsylvania Board of Probation and Parole, 1976).

62 Perlman, Deferred Prosecution and Criminal Justice--A Case Study of the Genesee County (MI) Citizens Probation Authority (Lansing, Mich.: Michigan Office of Criminal Justice Programs, 1972).

63 Abt Associates, Exemplary Project Evaluation Report--Improvement of County Parole Program (San Francisco) (Cambridge, Mass.: Abt Associates, Inc., 1976).

64 Ronald P. Sweet, Recidivist Felons in the Community--Final Evaluation Report of the Community Treatment of Recidivist Felony Offenders (San Francisco: National Council on Crime and Delinquency, 1975).

65 Lewis, Clark, and Kaufman, Evaluation of Specialized Units Project.

66 Philip L. Taylor and Kenneth W. Masters, Community-based Juvenile Probation--the Allegheny Count Program--Final Evaluation Report (Pittsburgh, Pa.: Allegheny County Court of Common Pleas, 1977).

67 Stephanie Morel et al., Probation and its Effect on Recidivism--An Evaluative Research Study of Probation in Nassau County (NY) (Garden City, N.Y.: Nassau County Probation Department, 1972).

68 Lawrence A. Bennett and Max Ziegler, "Early Discharge: A Suggested Approach to Increased Efficiency in Parole," Federal Probation 39 (1975):27-30.

69 B. S. Brown and E. M. Markham, Evaluation of Institutional and Community Experiences by Successful and Unsuccessful Parolees (Washington, D.C.: District of Columbia Department of Corrections, 1969).

70 Holt and Miller, Inmate-Family Relationships.

71 Lewis, Clark, and Kaufman, Evaluation of Specialized Units Project.

72 Cannon, Norfolk Fellowship.

73 Lawyer, Man-2-Man Program.

74 W. O. Jenkins et al., A Longitudinal Follow-up Investigation of the Post-release Behavior of Paroled and Released Offenders (Elmore, Ala.: Experimental Manpower Laboratory for Corrections, Rehabilitation Research Foundation, 1974).

75 Banks et al., Special Probation Projects.

76 Richard Ku, Richard Moore, and Keith Griffiths, Volunteer Probation Counselor Program--An Exemplary Project (Washington, D.C.: NILECJ, 1975).

77 J. Brian Lonergan, "Impact of the Volunteers in Probation (VIP) Program on Probationers," Journal of Volunteers with Delinquents 1(1972): 22-32.

78 Donald W. Beless, Edward R. Rest, and William S. Pilcher, Probation Officer Case Aide Project (POCA)--Final Report--Phase I (Chicago: University of Chicago Center for Studies in Criminal Justice, 1973).

79 Joseph E. Scott and P. A. Bennett, Ex-offenders as Parole Officers--An Evaluation of the Parole Officer Aide Program in Ohio (Columbus, Ohio: Ohio State University Press, 1973).

80 Joseph E. Scott, Ex-offenders as Parole Officers--An Evaluation of the Parole Aide Program in Ohio (Lexington, Mass.: D. C. Heath & Co., 1975).

81 Harry E. Allen and Ramon R. Priestino, Parole Officer Aid Program in Ohio--An Exemplary Project (Columbus, Ohio: Ohio State University, Program for the Study of Crime and Delinquency, 1975).

82 Carol H. Blew, Daniel McGillis, and Gerald Bryant, Only Ex-offenders Need Apply--Exemplary Project (Cambridge, Mass.: Abt Associates Inc., 1977).

83 Archie V. Connett et al., Utilizing Ex-offender Resources in Rehabilitation (La Jolla, Calif.: Western Behavioral Sciences, 1975).

84 E. L. Moran, W. A. Kars, and D. C. Munz, "In-program Evaluation of Community Correctional Agency for High-risk Offenders," Corrective and Social Psycial Psychiatry and Journal of Behavior Technology, Methods and Therapy 23, 2 (1977):48-52.

85 Jenkins et al., Follow-up Investigation.

86 Richard A. Bass, Analysis of the California Department of Correction Work Furlough Program, Research Report 57 (Sacramento, Calif.: California Department of Corrections, Research Division, 1975).

87 Al Rudoff, Work Furlough and the County Jail (Springfield, Ill.: Charles C. Thomas, 1975).

88 John M. Stanton and Donald R. Hoad, Parole Resources Center Program--The New York City YMCA Centers (Albany, N.Y.: New York Department of Corrections, 1974).

89 Richard P. Seiter, Joan R. Petersilia, and Harry E. Allen, Evaluation of of Adult Halfway Houses in Ohio (Columbus, Ohio: Ohio State University Program for the Study of Crime and Delinquency, 1974).

90 University of Alabama, Furlough Programs for Inmates--Final Report--National Evaluation Program--Phase I Project (Washington, D.C.: U.S. Department of Justice, Law Enforcement Assistance Administration, 1976).

91 State of Washington, Summary Evaluation of Juvenile Parole Learning Centers --1971-72, Department of Social and Health Services--Office of Juvenile Rehabilitation.

92 Joan McCord, Philadelphia Family Court, Correctional Group Counseling Evaluation Report (Philadelphia: Family Court, 1973).

93 Carol H. Blew, Daniel McGillis, and Gerald Bryant, Denver--Project New Pride--Exemplary Project (Cambridge, Mass.: Abt Associates, Inc., 1977).

94 Steven A. Lewis and Cary M. Lichtman, Project START: Evaluation of First Grant Period (Detroit: Wayne State University, Center for Urban Studies, 1977).

95 George A. Killinger and Glen A. Archer, Employment Assistance and Support Support for the Ex-offender (Project E.A.S.E.)--An Evaluation (Huntsville, Texas, Sam Houston State University, Institute of Contemporary Corrections and Behavioral Sciences, 1974).

96 L. L. Baca, Community Group Homes, Inc.--Final Report (Denver: Community Group Homes, Inc., 1975).

97 Minnesota Department of Corrections, Report on Juveniles in Group Homes (St. Paul, Minn.: Minnesota Department of Corrections, 1972).

98 National Council on Crime and Delinquency Effect of Vocational Upgrading upon Probationer Recidivism, One-year Evaluation of the Singer/Gräflex Monroe County, New York Pilot Probation Project (Hackensack, N.J.: 1972).

- 98 Russel Stricker et al., Evaluation on the Effectiveness of Hire, Inc., (Minneapolis, Minn.: Minnesota Correctional Services, 1974).
- 99 John N. Acquilano, "Monroe County (NY) Probation Program--Follow-up Report," Probation and Parole No. 4 (Summer 1972): 55-62.
- 100 Vince Chitren and Regis J. Reynolds, Cost/benefit Analysis of the Monroe County (NY) Pilot Program for Vocational Upgrading of Probationers (Rochester, N.Y.: University of Rochester, 1973).
- 101 Informatics Inc., Pennsylvania Community Treatment Services--An Evaluation and Proposed Evaluation Information System--Final Report (Rockville, Md.: Informatics, Inc., 1972).
- 102 R. L. Hamberg, Family House Program Evaluation (Olympia, Wash.: Washington Law and Justice Planning Office, 1973).
- 103 Jerry H. Langer, Evaluation Report of the Narcotics Education League's Residential Treatment Program for Chicano Heroin Addicts (Oakland, Calif.: Alameda Regional Criminal Justice Planning Board, 1975).
- 104 Paul Higgins, Minnesota Youth Advocacy Corps--An Evaluation (St. Paul, Minn.: Minnesota Governor's Commission on Crime Prevention and Control, 1974).
- 105 p=.03
- 106 Duane Baker et al., Oregon--Corrections Impact Program--Client Resources and Services Project--Evaluation Report (Sacramento, Calif.: American Justice Institute, 1976).
- 107 Malcolm M. Feeley, Effects of Increased Gate Money--Parolee Reintegration Project--Final Report (Hartford, Conn.: Connecticut Department of Corrections, 1974).
- 108 Craig Reinerman and Donald Miller, Direct Financial Assistance to Parolees Project--Research Evaluation (San Francisco: Scientific Analysis Corporation, 1973).
- 109 Donald Miller and Dan Waldorf, Direct Financial Assistance to Parolees Project--Research Evaluation (Sacramento, Calif.: California Council on Criminal Justice, 1973).
- 110 Craig Reinerman and Donald Miller, Direct Financial Assistance to Parolees: A Promising Alternative in Correctional Programming, Research Report 55 (Sacramento, Calif.: California Department of Corrections, Research Division, 1975).
- 111 Kenneth J. Lenihan, Financial Assistance in Reducing Recidivism (Washington, D.C.: U. S. Department of Labor, 1977).
- 112 J. Banks et al., Special Probation Projects.
- 113 Matthews, Rehabilitations Program.
- 114 Hermann Mannheim and Leslie T. Wilkins, Prediction Methods in Relation To Borstal Training (London: Her Majesty's Stationery Office, 1952).
- 115 Don M. Gottfredson and Robert F. Beverly, "Development and Operational Use of Predictive Methods in Correctional Work," in Edwin Golffield, ed., Proceedings of the Social Statistics Section of the American Statistical Association (Washington, D.C.: American Statistical Association, 1962), pp. 54-61.
- 116 Banks et al., Special Probation Projects.
- 117 Helene E. Cavior and Stanley Cohen, "Evaluation Research--Perspectives from a Corrections Setting," Criminal Justice and Behavior 2 (1975):237-251.
- 118 Martinson, Crime and Criminal Justice, p. 179.
- 119 Andrew T. Williams, Critical Evaluation of Research into Output Measures for Juvenile Correctional Programs (Stanford, Calif.: Stanford University, 1972).
- 120 Sandra L. Kellar and Kenneth A. Carlson, "Development of an Index to Evaluate Programs in a Correctional Setting," Canadian Journal of Criminology and Corrections 19 (July 1977):273-277.

- 121 Thorsten Sellin and Marvin E. Wolfgang, The Measurement of Delinquency, (New York: John Wiley & Sons, Inc., 1964).
- 122 Feeney, Hirschi, and Lemert, California Probation Subsidy Program; Johns, White, and Berkowitz, California's Probation Subsidy Program; Monkman, Cost-Benefit Analysis; Ohlin, Miller, and Coates, Juvenile Correctional Reform; Rutherford, A Nation without Prisons; Nath, Diagnostic and Treatment Program; Sweet, Recidivist Felons.
- 123 Lipton, Martinson, and Wilkes Correctional Treatment; John McDonnell, Training in Correctional Institutions.
- 124 Jessness and Derisi, Research Project.
- 125 Holt and Miller, Explorations; Lewis, Clark, and Kaufman, Specialized Units Project.
- 126 Tom Cannon, Norfolk Fellowship; Lewis, Clark, and Kaufman, Specialized Units Project; Lawyer, Man-2-Man Program.
- 127 McDonnell, Training in Correctional Institutions.
- 128 Griffiths, Seckel, and Raab, Assessment of Junior College Program.
- 129 Abt Associates, Wisconsin Mutual Agreement Program.
- 130 Dickover and Durkee, Guidance in Sentencing.
- 131 Lipton, Martinson, and Wilkes, Correctional Treatment.
- 132 Abt Associates, Wisconsin Mutual Agreement Program; Carter and Dightman, Minimum Service Caseloads; Robinson et al. San Francisco Project.
- 133 California Youth Authority, Parole Effectiveness Program; Ellis Perlman, Deferred Prosecution.
- 134 Stanley Turner, Philadelphia; Blew, Carlson, and Chernoff, Only Ex-Offenders Need Apply.
- 135 Ku, Moore, and Griffiths, Volunteer Probation.
- 136 Scott and Bennett, Ex-offenders; Scott, Ex-offenders; Allen and Priestino, Parole Officer Aid Program.
- 137 Seiter, Pertersilia, and Allen, Halfway Houses; University of Alabama, Furlough Programs.
- 138 Lenihan, Financial Assistance; Reinerman and Miller, Financial Assistance.
- 139 Carney, Evaluation of Psychotherapy; Jew and Clannon, Group Psychotherapy.
- 140 Jew, Kim, and Mattocks, Group Psychotherapy.
- 141 Acquilano, "Monroe County (NY) Probation Program"; Stricker, et al., Effectiveness of Hire, Inc.
- 142 Connett et al., Utilizing Ex-offender Resources.
- 143 Lewis, Clark, and Kaufman, Specialized Units Project; Pennsylvania Board of Probation and Parole, Specialized Units.
- 144 McCord, Philadelphia Family Court.
- 145 Bennett and Ziegler, "Early Discharge."
- 146 Hamberg, Program Evaluation.
- 147 Joseph H. Hudson, An Experimental Study of the Differential Effects of Parole Supervision for a Group of Adolescent Boys and Girls (Washington D.C.: National Institute of Law Enforcement and Criminal Justice, March 1973).
- 148 Bennett and Ziegler, "Early Discharge."
- 149 Abt Associates, Evaluation Report.

What Is Citizen Crime Prevention?

Robert K. Yin is a visiting associate professor at the Department of Urban Studies, Massachusetts Institute of Technology, and has a consulting practice in public policy research.

INTRODUCTION

Policymakers have traditionally seen community efforts as only a supplementary resource for preventing crime. In contrast, this paper makes the following proposition: Community efforts, far from being a supplementary resource, may actually be the essence of successful crime prevention activities. To the extent that this is true, research on citizen crime prevention must begin with adequate operational definitions of these activities. Otherwise, few conclusions can be drawn concerning the most desirable community conditions or the most effective interventions. Thus, this paper is based on a second proposition: Most research has failed to describe properly the actual activities involved in citizen prevention. The paper concludes by reviewing the ways that research can be improved in this regard.

In comparison to these propositions, many feel that the criminal justice system typically provides the major alternatives for crime prevention. For instance, crime vulnerability is thought to vary primarily with the amount of patrol by paid police officers, and crime pressure¹ is thought to be relieved by dealing with offenders--through incarceration, rehabilitation, deterrence, or other sentencing policies.² As part of this same viewpoint, private citizens are thought to play only a limited role in crime prevention, even though they may engage in any number of private activities (e.g., keeping private vehicles locked when not in use, using door bolts, or installing new alarm systems), in various group activities (e.g., watching a neighbor's property), or in assisting in law enforcement (e.g., reporting all crimes, assisting police investigations, or serving as witnesses in the prosecution of offenders).

The suggestion of this paper, however, is that it is the criminal justice system that serves as the supplemental resource, and that effective crime prevention must primarily be based on the informal social controls imposed by residents. Thus, the maintenance of public safety is seen as an integral function of the community, and effective crime prevention only occurs when this function is performed properly. This paper deals with one aspect of this claim--the role of neighborhood residents in reducing vulnerability to

Copyright (c) 1979 Robert K. Yin

crime. There is, however, an equally compelling literature that notes the importance of community affairs in reducing crime pressure through the appropriate socialization of youths and potential offenders, as well as in rehabilitating criminals.

THE NEIGHBORHOOD AS THE SOURCE OF PUBLIC SAFETY

Jane Jacobs was one of the first to call attention to the crime preventive role of neighborhoods. Her vivid description of the uses of sidewalks begins with the dominant function of providing public safety:³

The first thing to understand is that the public peace--the sidewalk and the street peace--of cities is not kept primarily by the police, necessary as police are. It is kept primarily by an intricate, almost unconscious network of voluntary controls and standards among the people themselves, and enforced by the people themselves....No amount of police can enforce civilization where the normal, casual enforcement of it has broken down. [Emphasis added.]

Others have also asserted the importance of safety in the neighborhood environment. As the sociologist Gerald Suttles puts it:⁴

The quest for a good community is, among other things, a quest for a neighborhood where one does not fear standing at arm's length from his neighbor, where one can divine the intent of someone heading down the sidewalk, or where one can share expressions of affect by the way adjacent residences dress up for mutual impression management.

Suttles argues that all neighborhoods must perform the function of residential defense against outsiders. This is because residential proximity creates its own dangers and difficulties of social control. People who are close at hand are impossible to avoid, and people thereby want to live in neighborhoods where they feel safe and are at a distance from those whom they cannot trust. "The defended neighborhood," according to Suttles, "segregates people to avoid danger, insults, and the impairment to status claims."⁵ Occasions arise when some neighborhoods are not capable of defending themselves, and thus compensatory actions--including neighborhood revitalization--must be initiated. The main point, however, is that the role of the neighborhood in defending itself against outsiders is an instrumental function that goes beyond the mere romanticization of the traditional neighborhood as a sentimental place. As a result, with the rising crime rates during the late 1960's, "law and order issues [became] basic to the defended neighborhood and, indeed, basic to the development of a community itself."⁶

The importance of public safety in neighborhood life has been noted in ways other than the words of keen journalists and academic observers, however. Residents implicitly expect their neighborhoods to provide protection and are affronted when such protection fails. It is thus not surprising that one of the roots of this nation's major war on crime, beginning with the Lyndon Johnson administration and the Great Society legislation, was an unbelievable event that occurred in a residential neighborhood: the murder of Kitty Genovese in full earshot of 38 people.⁷ The Genovese case influenced the adoption of an anticrime plank in the Republican platform of 1964.⁸ Following Goldwater's defeat, President Johnson then usurped the Republican position by appointing the President's Crime Commission (Katzenbach Commission) in 1965, and the work of the Commission led directly to the Omnibus Crime Bill (1968) and the nationalization of the war on crime. In the ensuing years, it has similarly been no surprise that the dominant public target of crime control programs has been crime in the street, i.e., robbery and predatory crime in the neighborhood, and not white-collar crime or organized crime--both of which may incur greater costs.⁹

The potential importance of citizen activities in preventing crime is based in part on the negative results of other courses of action. Thus the popular notion that more police can lead to less crime has not been corroborated by most research. Wilson, for instance, reviewed the key experiments on the effects of police preventive patrol and concluded that, with the possible exception of saturation patrol (a condition that may be socially undesirable and fiscally impossible for long-term application), there was no clear evidence that increases in preventive patrol led to a reduction in crime.¹⁰ Wilson's conclusion was based on the results of such field experiments as the Kansas City Preventive Patrol experiment,¹¹ in which police deployment patterns were systematically varied over time between comparable areas of a city.¹² Another review, involving no experimentation, reached similar conclusions. In this review, Levine analyzed robbery and murder rates as a function of different levels of police manpower in 26 cities and found no relationship between the number of police and crime reduction.¹³

An additional line of evidence comes from cross-cultural comparisons. Although such comparisons are always somewhat tenuous, one investigator has suggested that the existence of neighborhood crime prevention associations may be a major reason for Japan's low crime rate.¹⁴ In comparison to the United States, which is similar to Japan in its degree of industrialization and state of national development, Japan has a very low crime rate and yet has fewer police, an equivalent tradition of violence, and more lenient sentencing policies. The major difference between the two countries thus appears to be the role of neighborhood associations in Japan, which are present in nearly every neighborhood and which promote public safety and crime prevention practices.

At the same time, positive evidence on the effects of citizen crime prevention activities is admittedly still scarce. An important observation is simply that these activities have increased and continue to draw significant support from both residents and policymakers alike.¹⁵ The most common theme in these activities has been to promote improved surveillance, or the use of "eyes and ears."¹⁶ This effect may be promoted through a variety of community activities. For instance, Yin et al. found that resident patrols existed in numerous neighborhoods across the country and were predominantly resident-initiated, resident controlled activities aimed at crime prevention.¹⁷ Such contemporary patrols differed from their older counterparts¹⁸ in that riot-suppression activities and police antagonisms were minimal. The latter-day patrols ranged from residents actually conducting mobile patrols in their own cars to the hiring of private security guards to protect building entrances and premises. Similarly, Bickman et al. found a variety of citizen crime-reporting programs in different cities, including "whistle stop" activities, radio watch projects, and educational efforts aimed at improving residents' knowledge of crime prevention practices.¹⁹

The basic element of citizen crime prevention appears to be social interaction, in which residents retain an everyday familiarity with each other. One mechanism derives implicitly from normal or routine neighborhood behavior: The more that such behavior follows certain patterns and rituals, the easier it is to detect the presence of a stranger. One of the most important neighborhood functions is therefore the establishment of norms and a particular culture of normality. At the interpersonal level, as described by Erving Goffman:

. . . normal appearances mean that it is safe and sound to continue on with the activity at hand with only peripheral attention given to checking up on the stability of the environment.²⁰

Goffman also describes the use of alarms, all-clear signals, the perceptually relevant world of the individual (the Umwelt), disguises, and zones of vulnerability, all of which are important in protecting oneself from possible crimes.

A second important mechanism for neighborhood defense is to stimulate as much casual and public surveillance as possible. This can be facilitated by an appropriate architectural design, as uncovered for instance by Oscar Newman's research.²¹ However, people often forget that "defensible space," to use Newman's term, is not purely a product of an architect's plans. Defensible space requires not only the appropriate physical design (for buildings as well as for public areas around buildings) but it also requires surveillance by residents. In Newman's words, "fundamental to this study is the proposition that through the manipulation of building and

spatial configurations, one can create areas for which people will adopt concern [emphasis added],²² a theme whose importance has been reaffirmed in Newman's later work.²³

Third, citizen crime prevention activities need not be carried out in complete independence from the criminal justice system. In fact, the success of most activities depends on positive interaction with local law enforcement agencies. Reppetto, for instance, conducted a study of residential crime, including interviews of offenders involved in burglary cases.²⁴ He concluded with the suggestion that the most fruitful course of future action would be the development of a crime prevention approach that would:²⁵

...blend the deterrent effects of the criminal justice system with citizens' anticrime efforts.... It is possible, for example, that the "rapid response" techniques of the police could become a more meaningful deterrent to residential crime if environmental characteristics could be modified to maximize surveillance possibilities and encourage a sense of territorial concern among residents; citizens would take a few more precautions aimed at "slowing down" the prospective burglar so that his suspicious activities might attract the attention of neighbors; and observing neighbors might feel a "social commitment" sufficient to prompt them to summon the police.

In summary, citizen crime prevention may be seen as an integral neighborhood function, going far beyond individual actions such as installing new target-hardening devices. The social interactions upon which crime prevention is based may also lead to the further cohesiveness of a neighborhood,²⁶ in which residents also help each other in dealing with other everyday functions, such as child care, shopping, schooling, and emergency assistance. In contrast, neighborhood interactions and cohesion might even diminish as a result of other approaches to crime prevention, e.g., increasing preventive patrol by police.

In this context, a formal criminal justice system can thus be interpreted as a compensatory system, being needed when such informal mechanisms of social control as might operate in a neighborhood setting fail to work properly. At the neighborhood level, disruption of these mechanisms could result, for instance, from:

- high rates of residential turnover;
- the displacement by automation of such traditional neighborhood artisans as doormen, elevator operators, and janitors;
- the declining presence of small store owners, corner druggists, and sidewalk concessions;

- an imbalanced demographic pattern, in which, for example, children far outnumber adults; or
- changing cultural styles in which, for instance, both adult members of a household have full-time jobs outside of the neighborhood.

However, a compensatory system, no matter how effective, cannot become the primary system. Instead, ways must be found to facilitate the traditional role of neighborhood and citizen crime prevention efforts. For this reason, policymakers and researchers alike have attempted to identify and test the major alternatives.

IDENTIFICATION OF CITIZEN CRIME PREVENTION ACTIVITIES

The Definitional Problem

New citizen crime prevention efforts can be based on two sources of information. First, improved practices may emerge as a result of experiences in specific neighborhoods. Such knowledge may be defined as non-R&D-based knowledge, and citizens in one community may directly inform those in another community of their experiences. Second, improved practices may emerge as a direct result of new research experiments or other evaluation studies. Such knowledge may be defined as R&D-based knowledge, and research investigators may be responsible for disseminating the results to potentially interested citizens.²⁷ In either case, the communication of new knowledge depends on one essential condition: The practice must be operationally defined before it can be tried elsewhere. For a full evaluation, additional information is also essential--e.g., an assessment of the outcomes of the efforts and an understanding of causal relationships (what produces what and why). However, if a citizen crime prevention effort has not been adequately described in the first place, there is no way of knowing what happened, much less allowing for any subsequent evaluation of the experience.

For example, consider the problem of describing block organizations. Such organizations are frequently credited with having initiated significant crime prevention activities.²⁸ A block organization may hold meetings, disseminate literature to residents, or conduct specific public safety campaigns--such as getting residents to improve their cautionary behavior or to carry whistles or other devices capable of emitting signals for help. In order to duplicate these experiences, the specific activities must be documented--e.g., how many people were involved, in what way, and using what specific materials and procedures. But there is an even more serious problem--that of defining the basic block organization. Residents in one neighborhood may want to contact such organizations in other neighborhoods to exchange information, or a research investigator may want to design a study to determine what generalizations can be made about these orga-

nizations. In either case, the question remains: How are existing block organizations to be identified? Because organizations vary in their formality and do not "announce" their existence in any consistent manner, residents may have to rely on word-of-mouth techniques that may be highly inefficient. The research investigator, however, faces a more serious problem: The study must begin with a functional or operational definition²⁹ of a block organization; otherwise there will be no assurance that the proper sample has been studied. Naive investigators will settle this problem in an entirely unsatisfactory way--they will simply include all organizations that call themselves "block organizations."

Figure I lists 19 different kinds of citizen crime prevention activities. The figure gives only the briefest description of each activity, but all require some degree of formal organizational effort. The figure includes a variety of voluntary efforts that may be sponsored by the police or by a community organization or by both. The purpose of the figure is to show the varied forms that citizen crime prevention can take, and to suggest that, just as with the problem of defining "block organizations," each of these activities must be described in considerable detail before any transfer of the ideas to new sites or before any multisite assessment can occur.

Typical Treatment of the Definitional Problem

Although an operational definition of a citizen crime prevention activity is essential for understanding what occurred in a particular neighborhood (or for extending the experiences to other neighborhoods), existing research guidelines and many previous studies have ignored this problem. This has occurred in spite of the fact that, over the last decade, researchers have become much more sophisticated about how different kinds of programs--especially policy interventions--should be evaluated.

Evaluation Studies. For instance, there now exist any number of handbooks in evaluation research that cover the basic methodological issues involved in assessing public programs.³⁰ These efforts have tended to focus on several important issues such as:

- the establishment of outcome measures by which to determine the degree of program success or failure;
- the development of research designs that permit causal inferences and therefore relate specific actions to specific outcomes; and
- an appreciation for the politics of evaluation, which include the relationship between the evaluation team and the staff operating the program as well as the broader political setting within which an evaluation is conducted.

FIGURE I

DIFFERENT TYPES OF CITIZEN CRIME PREVENTION ACTIVITIES

Program	Description
Block Security Programs	Attempt to improve residents' awareness and education concerning public safety, and may also assign surveillance or assistance responsibilities to specific individuals such as block mothers. These programs are very diverse and may include actual patrol activities.
Blockwatchers and Neighborhood Watch Programs	Attempt to improve citizen reporting of crimes and suspicious events in the neighborhood. Residents are sensitized to signs of criminal activity and are given specific names or numbers to call in case of emergencies.
Citizens Alert Programs	Attempt to improve citizens' education about public safety and ability to report crimes. Typical emphasis is on disseminating information about physical security for the home and giving citizens a specific name or number to call in case of emergencies.
Citizen Volunteer Programs	Attempt to reduce crime pressure in the community at large. Typical activities are aimed at: improving educational and employment opportunities for youths, providing recreation services, and providing counseling and rehabilitation services. Many programs cover courts and corrections activities and not just police-related activities.
Community Radio Watch	Attempts to improve citizen reporting of crime and other emergencies. Business, firms or private citizens with vehicles having two-way radios are encouraged to report incidents to their dispatchers.
Community Service Officers	Attempt to improve communications between police and community. A neighborhood youth is trained by the police to perform community services. The services do not necessarily involve serious crime prevention responsibilities.
Escort Services	Attempt to reduce vulnerability to crime. Residents escort children to school or elderly persons when collecting and cashing checks. Escort service follows specific routine and time of operation.
Home Security Program	Attempt to reduce vulnerability to crime. Citizens are encouraged to install new and more effective security devices to protect their homes.
Improved Alarm Systems	Attempt to provide improved reporting of crimes in progress. In some cases, citizens carry special devices that transmit digital signals indicating need for help. In other cases, special alarms are installed in adjoining homes so that neighbors can help each other out.

Program	Description
Police-Community Councils	Attempt to increase informal contact between citizens and the police. Resident groups may be organized as an advisory body, meeting at intervals with the local police. The meetings increase communication in both directions, with residents having an opportunity to voice complaints and the police having an opportunity to explain regulations and other departmental changes.
Police-Community Relations Programs	Attempt to increase informal contact between citizens (mostly youths) and the police. Typical activities are: officers speak at school functions, officers and youths go on outings, and youths ride in police cars to get exposed to the officer's world.
Police Reserve Units	Attempt to reduce vulnerability to crime. Volunteers undergo extensive training and devote specific periods to police work. Volunteers may be considered part of an auxiliary police group, and the activities of such groups are fully controlled by the local police authority.
Property Identification	Attempt to reduce vulnerability to crime. Citizens are encouraged to mark property and to report registration numbers to police so that property may be more easily traced if stolen.
Resident Patrols	Attempt to reduce vulnerability to crime. Residents actively monitor specific physical locations, either by patrolling or by observing for a given period of time. Locations may involve either streets and outdoor areas or hallways and corridors within housing projects.
Secret Witness Programs	Attempt to increase information from citizens about criminal activities. A community organization or newspaper may offer monetary rewards for information from anonymous callers regarding a particular crime or specific lawbreakers, including pushers or fences.
Security Guards	Attempt to reduce vulnerability to crime. Citizens hire guards to patrol specific areas, usually a residential block or a housing project. Guards may operate in isolation or may be part of a professional security service.
Street Lighting Programs	Attempt to reduce vulnerability to crime. Brighter street lights are installed in public areas.
Vigilante Groups	Attempt to enforce laws and administer own justice. Volunteers exist outside of the regular public safety network and carry out both surveillance and apprehension activities.
Youth Patrols	Attempt to reduce tensions in times of neighborhood disorders and to reduce crime vulnerability. Volunteers circulate within the community, especially during times of civil unrest, and attempt to reduce hostilities among residents and between residents and police.

SOURCE: Robert K. Yin et al., Patrolling the Neighborhood Beat: Residents and Residential Security, R-1912-DOJ, (Santa Monica, Calif.: The Rand Corporation, 1976)

The first two topics are indeed essential for conducting rigorous studies of citizen crime prevention activities.³¹

Nevertheless, these guidelines have given only passing attention, if any, to the problem of defining the citizen crime prevention activity in the first place. One handbook, directed specifically toward the evaluation of crime control programs, typifies this gap.³² The handbook provides detailed guidelines for developing adequate outcome measures, with appropriate comments on the strengths and weaknesses of using reported crime data and victimization surveys. The handbook also provides a summary of the basic issues in research design,³³ e.g., the establishment of appropriate control and experimental groups and the use of longitudinal comparisons. Finally, the handbook describes three illustrative examples in which crime reduction effects may be studied: programs calling for increased police patrol, for improved burglary investigation, or for increased street lighting. No single chapter of the handbook, however, addresses the problem of defining the crime prevention activity. Furthermore, the three illustrative examples all overlook the fact that the crime control activities discussed can be conducted in a variety of ways. Thus, police patrol can be increased through any number of specific resource allocation procedures (and the detailed definition of any particular way is a difficult task); improved burglary investigation involves a complex set of organizational changes within police departments that are also difficult to define; even efforts to increase street lighting can involve a variety of procedures that are difficult to define operationally (e.g., the specific types of bulbs used, the number of lights installed per street mile, the height of the lights, and the placement of the light fixtures in relation to the width of the sidewalk).

Without such definitions, it is impossible to determine what activity occurred. Hence, even if the activity produced desirable outcomes, little can be learned or replicated. For one thing, it would not be known whether such outcomes were related to the entire activity or to a specific subset of those activities. This flaw permeates most research studies on citizen crime prevention activities. For instance, a thorough review of 41 street lighting projects--an intervention that may be considered a citizen crime prevention activity because it facilitates citizen surveillance in an area--found that none of the projects adequately described the system characteristics of the street lighting efforts.³⁴ In fact, the basic measure of the degree of "treatment"--i.e., the amount of light produced in the newly installed systems--could not even be ascertained because most municipalities did not know how to make systematic light measurements, and few municipalities even owned standard light meters in working condition. In contrast, Figure II gives a list of the various system characteristics that need to be assessed in operationally defining a street lighting program.

Similarly, there have been many studies of other types of citizen crime prevention activities in which an appropriate evaluation was conducted, but in which the basic activity was not defined. These have included public education programs in which residents are urged to keep their homes and automobiles locked to reduce crime,³⁵ efforts at organizing informal surveillance by residents or police,³⁶ and security survey programs, in which homes and other establishments are surveyed for potential weaknesses in preventing breaking and entering by potential burglars.³⁷ In every case, what need to be described are the actual procedures and content of the crime prevention activity. For instance, in a public education effort, a minimal description would cover the number of residents who participated, the degree of their involvement, the training methods of the "educators," and the curriculum used (or actual content of the information transmitted). This approach, of course has now been given increasing attention in the form of process evaluations.³⁸

Other Studies. The need for adequate definition also exists in studies not involving the evaluation of a specific intervention. Such studies may include assessments of neighborhood conditions presumed to be associated with effective or ineffective crime prevention. A typical situation would be the study of environmental design characteristics related to crime incidence, where assessments might be made of the physical characteristics of known crime sites. Some characteristics, such as the presence of a nonworking adult at home during the day, can be defined and assessed readily.³⁹ Other characteristics, however, such as the condition of the target or surrounding buildings, can only be assessed through ratings by human observers. In such cases, assessment is more difficult and the rating scheme itself must be replicated before any conclusions can be drawn.⁴⁰ Because it is well-known that building conditions are not easily rated with consistence, this particular characteristic is, in fact, not easy to define operationally.

One illustrative study that appeared to cope well with these definitional problems involved the examination of the physical characteristics of commercial stores that had been crime targets.⁴¹ The types of features that were assessed included: proximity to a major transportation route, the amount of traffic on adjacent streets, the residential or commercial type of land-use on adjacent blocks, and the nature of commercial activities immediately surrounding the target site. Normally, readers might consider these descriptions as sufficient operational definitions. (In many studies, this is the only level of description available). However, this is not the case. Each of the above four features was actually operationalized, for the purposes of the study, in the following manner:

Proximity to major transportation route: This was defined as the presence, within a 2-block radius of the target

site, of a transportation route that dominantly provided access to traffic of a nonlocal origin (e.g., interstate carriers, crosstown commuters, etc.).

Amount of traffic on adjacent streets: This was defined as the result of the traffic counts, obtained from the traffic department for a 2-year period, on streets immediately adjacent to the target site.

Land-use on adjacent blocks: This was defined on the basis of observers' ratings concerning the degree to which block faces, within a 2-block radius of the target site, contained residential units, commercial units, or vacant sites.

Commercial activities immediately surrounding target site: This was defined as commercial activities that provided a line of vision, within a block's distance from the target site, of the entrance to the target site.

This illustration, along with the ideal list of system specifications for street lighting programs in Figure II, is intended to suggest that the crime prevention activity or features deemed relevant must actually be measured in order to be acceptably defined in an operational sense. Thus, what has been lacking in most studies of citizen crime prevention activities is the empirical measurement of the activities. Only with such measurements can the needs for generalizing, making further assessments, or replicating various experiences be satisfied.

In this respect, physical features in the environment that are important to crime prevention may appear, at first glance, easier to assess than human or social characteristics. Studies of street patterns⁴² or other "defensible space" features would thus seem to have an advantage over studies of citizen activities such as public education programs. However, some features may involve subtle effects that are nearly impossible to measure. For instance, research on street lighting suggests that an important factor is the degree to which street lights produce shadows (or, technically, a high-contrast environment), in which potential offenders may hide. The shadow effects are hard to describe or measure, but may be so important that brightly lit streets are less safe than moderately lit ones.⁴³ Furthermore, it should be remembered that most changes in the physical environment--including increases in street lighting--are only important to the extent that specific types of human behavior are altered--e.g., increasing the actual or apparent surveillance behavior of residents, or increasing the sense of territoriality of residents. To this extent, the ultimate verification of the effects of physical features relies on the assessment of human behaviors. In this regard, previous studies⁴⁴ have failed to develop the necessary operational criteria for such behaviors.

Figure II

Purpose	Categories	Measures [Remarks]
<ul style="list-style-type: none"> • Project Plan 	<ul style="list-style-type: none"> • Performance Specifications • System Design • Target Area 	<ul style="list-style-type: none"> • Technical specifications, including average horizontal illumination, illumination uniformity, roadway/walkway luminance, glare, etc. for <i>vehicular roadways</i> and <i>pedestrian walkways</i>. [Compare with IES performance specifications--note that the IES specifications are expected to be revised in 1977.] • Management specifications: project budget and schedule. • Number and location of street lights. [Determine these measures for both the old and the new system.] • For each street light: light source type (i.e., high-pressure sodium, mercury vapor, etc.), wattage and initial lumen output; luminaire light distribution patterns; glare characteristics (i.e., full-, semi- or non-cutoff), and photometric data (supplied by manufacturers); pole mounting height, spacing and configuration, and bracket overhang; wiring type (i.e., overhead, underground; series, parallel). [Determine these measures for both the old and the new system.] • Selection criteria (e.g., high-crime, traffic safety, other program links, natural boundaries, political factors, technological factors, etc.) and decision-making process. • Target area boundaries and area in terms of number of street miles or number of blocks. • Land use (i.e., residential, commercial, industrial, etc.). [Note day/night land use <i>differences</i>.] • Environmental conditions, including classification and condition of streets and alleys; structural conditions of buildings; opportunities for concealment and surveillance; and distribution of targets. [Measures relevant to the proper design and effective use of the built environment are being developed and tested as part of the LEAA-funded Crime Prevention Through Environmental Design Program.] • Social indicators, including demographic and socioeconomic variables and trends.

SOURCE: James M. Tien et al., Street Lighting Projects. (Washington: Department of Justice, July 1977).

The Complexity of Citizen Crime Prevention Activities

There is a further complication to the definitional problem, not fully covered by the mere stipulation that citizen crime prevention activities be adequately measured. This is that most citizen crime prevention activities are complex affairs, not limited to simple actions. This creates almost insurmountable difficulties for any measurement effort. The problem may be likened to similar difficulties in controlling and assessing other types of interventions, such as changes in police preventive patrol tactics. In the well-known Kansas City experiment, one problem noted after the experiment was that:⁴⁵

... patrol cars for reactive beats--i.e., beats that received nominal preventive patrol--may have inadvertently produced a significant preventive effect by their longer and more visible drives to intervene in accidents. The desired treatment--variation in preventive patrol but consistency in response to actual calls for assistance--may be difficult to implement.

For citizen crime prevention activities, two illustrative examples may be described, one involving defensible space and the other involving a property-marking program known in many cities as Operation-Ident.

Environmental Design and Crime Prevention. No single citizen crime prevention activity has received more attention over the past decade than the "defensible space" notions developed by Oscar Newman.⁴⁶ Newman identified numerous features of the physical environment that appeared to facilitate residents' natural surveillance and sense of territoriality over the immediate environment--e.g., the use of single rather than double-loaded corridors in public housing units, or the use of glass fronts to increase the visibility of lobbies and entrances to multiple-dwelling buildings. On the basis of this and other research,⁴⁷ it was therefore hypothesized that alterations in environmental design would reduce a site's vulnerability to predatory crime and property crime. As a result, both the U.S. Department of Housing and Urban Development and the U.S. Law Enforcement Assistance Administration have supported extensive handbooks and demonstrations aimed at applying specific design guidelines to the construction of new sites or the modification of existing ones.⁴⁸

The largest demonstration activity has been a four-city effort, in which specific alterations were made in commercial, residential, and school sites.⁴⁹ This activity has become known by the initials CPTED (Crime Prevention Through Environmental Design), and it is presumed that many other local attempts at making changes in the built environment have occurred. The major CPTED demonstration has been implemented and evaluated by the Westinghouse Corporation. The eval-

uation results provide a basic illustration of the complexities encountered in mounting and assessing real-life citizen crime prevention activities.

Although the defensible space thesis and the initial intervention design called for a narrow set of changes in the built environment--e.g., the redesign of school grounds and locker rooms, the modification of recreational facilities for residential use, or the closing of residential streets for recreational use--the actual implementation process involved many citizen activities in addition to these microenvironmental changes.

For instance, at each site, public education programs told residents of the implications of the changes as well as of other crime prevention practices. Citizen involvement was encouraged and necessary in helping decide what changes were appropriate in the first place. Because there is reason to believe that this sort of citizen involvement also facilitates crime prevention, the demonstration project cannot be faulted for having instigated these broader activities. However, the addition of these activities to the narrower set of microenvironmental concerns effectively precluded any precise definition, much less measurement, of the basic intervention that occurred.

At one of the four sites, CPTED activity was augmented by other crime prevention projects. Thus, the narrower microenvironmental changes were not only augmented by broader citizen involvement, but also by independent projects such as: implementation of a major redevelopment plan for street improvement; installation of special bus shelters by the local transit company; a separate street lighting project (also funded by LEAA) in the same area; and other block-watch, cleanup, and public awareness programs. Again, such a situation is common in real-life interventions. If a residential or commercial area is suffering from a high crime rate, responsive public officials and citizens are likely to be engaged in a whole host of attempts to reduce crime, and the effects of any specific intervention are going to be difficult to extract. In addition, because these various attempts may be administered independently, the whole array at a given site may not even be known.

This experience in attempting to assess CPTED interventions thus illustrates two complexities that are likely to plague the assessment of any citizen crime prevention activity. First, crime prevention, as this paper has argued earlier, is much a matter of enhancing community involvement in the broadest possible sense; narrow approaches to the crime problem in an area are likely to be ineffective, but broad approaches, even given a particular core activity such as CPTED redesigns, are difficult to define and measure. Second, the problems of high-crime areas will be urgent enough that multiple attempts will be made at dealing with the problem. Residents may initiate their own patrols or hire private guards; private industry, such as commer-

cial stores or transit companies, may take their own actions; city government may initiate new programs; and Federal agencies may also provide assistance. In this context, there is little opportunity for precisely identifying any specific crime prevention effort.

Property-Marking Programs to Reduce Thefts. Another popular citizen crime prevention activity initiated in recent years involves the permanent marking of individual property so it can be easily identified, should it be later stolen. This activity has become known as Operation-Ident, and according to a review conducted several years ago, there were over 100 such projects being conducted in cities across the country.⁵⁰ The procedure is as follows:

- A resident borrows marking equipment from the local police department, permitting him or her to emboss or engrave an identification number on each piece of valuable property in the home. (The identification number used may be the resident's social security number or driver's license number; however, the number must be accessible for later tracing by the police.)
- The equipment is returned to the police department and the resident puts a decal on his or her front door to indicate that property has been marked and that the resident is thus a participant in an Operation-Ident project. The decal is presumed to have some deterrent effect.
- When the police recover stolen property, officers check it for any identification numbers.
- When an identification is successfully traced, the property can be returned to its owner.

The Operation-Ident experience illustrates further the complexity of citizen crime prevention efforts. In contrast to CPTED, which highlighted the complexity of the broader context within which single interventions occur, Operation-Ident should suggest some of the complexities of the implementation process that occur even within a single activity.⁵¹ For Operation-Ident, these involve both the property marking and property-inspecting procedures.

First, there is no adequate means for assessing the effectiveness with which property has been marked. Because there is no standard marking procedure, residents may have put the identification numbers in a place that is awkward or simply difficult to locate. Moreover, there is no monitoring to assure that later purchases by the resident have also been marked. Thus the marking procedure cannot be readily assessed either with regard to the quality of the marking or the proportion of property that has been marked.

Second, within the police department, there is also a critical step--the effectiveness with which the individual police officer conducts the search for identification numbers. In fact, the previous review of Operation-Ident projects⁵² noted that officers were typically "unable to spend more than a token amount of time" in examining stolen equipment for such numbers. Again, the inspection procedure cannot be readily assessed either with regard to the quality of the inspection or the proportion of stolen property that has been inspected. (Note that, to be effective, the officer must inspect all stolen equipment, even though only a small portion of it is likely to be marked in the first place.)

On the surface, Operation-Ident appears to be a fairly simple and straightforward crime prevention activity. Yet, even here, the problems of defining the activity are severe. The illustration shows that assessments of the implementation process must be made at a very detailed level, one that is probably beyond the scope of everyday practicality. Yet, once again, without such detailed assessment, little can be said in interpreting any of the outcomes of a particular Operation-Ident project. Where the project has been ineffective, for instance, the source of the ineffectiveness cannot be traced directly to the property-marking or property-inspecting procedures, because such procedures have not been measured. The results, in short, whether positive or negative, will not be able to distinguish between two fundamental situations: where the basic concepts underlying Operation-Ident are ill-conceived, and where there have simply been flaws in the implementation process. If this is the situation with Operation-Ident projects, one can readily imagine the difficulties with other citizen crime prevention activities that, even on the surface, have a more complicated implementation process.

SUGGESTIONS FOR IMPROVING FUTURE RESEARCH

The previous sections of this paper have described the importance of, and difficulties in, assessing citizen crime prevention activities. The main concern has been with the definition and measurement of the activity itself, whether such activity is part of a formal intervention project or is merely an ongoing experience about which others wish to learn. The argument has been that, without more precise knowledge about the activity, information about the activity's effects or outcomes cannot be usefully interpreted.

To simplistically state the standard research paradigm, some independent action or activity (typically referred to as x) is believed to have an effect or impact on some dependent activity (typically referred to as y). Our claim is this: although researchers may be rightfully concerned with (1) the development of appropriate research designs for linking x with y, or with (2) the definition of various outcome measures--i.e., definitions of y--inadequate attention has been given to the problem of defining x.

A companion paper⁵³ addresses the earlier problems; the suggestions offered below are intended to assist future research in establishing better operational measures of x, which is the citizen crime prevention activity itself.

Most citizen crime prevention activities actually involve at least five subactivities that must all be defined operationally. These include:

- outreach efforts: the efforts made to recruit participants;
- organizational form: the nature of the organization, whether formal or informal, that is responsible for administering the crime prevention activity;
- relationship with the local police: the nature of communication and cooperation between the citizen activity and the local police;
- implementation process: the ways in which specific activities are designed and initiated; and
- the activity itself: the amount and nature of surveillance, preventive patrol by citizens, or other actions that actually constitute the crime prevention activity.

A comprehensive assessment of citizen crime prevention activities must account for all of these factors, as well as independent activities that may be occurring in the same area. The sections below describe some of the questions that must be answered in defining each of the five subactivities; researchers must then apply the same framework to any of the independent efforts that are simultaneously occurring.

One general issue that cuts across all of these subactivities, however, should first be mentioned. This is that citizen crime prevention activities, like other human behaviors, change over time. Some of these changes may be subtle--e.g., the quality of patrol duty by a citizen patrol may vary over time. Other changes may be dramatic--e.g., the crime prevention activity may be dissolved and reconstituted several times over the history of what may have appeared to be the same project. These changes over time all need to be monitored, making the definitional problem even more difficult. The sections below thus assume that, wherever specific issues are raised, the reader will keep in mind the fact that measurements cannot be made at one point in time alone, but must be monitored over several points in time.

Outreach Efforts

It is generally believed that the more citizens who participate in a crime prevention activity, the more effective the activity will be. Although there has been no clear empirical test of this proposition, it does appear to be a reasonable assertion. In addition, for some types of programs, such as Operation-Ident, it is clear that increased participation does facilitate crime prevention: Stolen goods would be more difficult to fence if everyone, rather than a few people, were able to mark their property with a unique identifier. Unfortunately, one of the notable problems with Operation-Ident projects has been their low participation rates, averaging less than 10 percent of the households in the target areas.⁵⁴

How a particular crime prevention activity enlists participants is therefore an important part of the overall activity. Typically, this whole procedure is overlooked in any assessment; the participants are merely assumed to exist. Occasionally, reports will indicate that outreach efforts consisted of the use of media or word-of-mouth communications.⁵⁵ However, such minimal description is clearly inadequate. For any citizen crime prevention activity, an adequate assessment must define:

- the universe of eligible participants and their demographic characteristics (usually defined as the residents in a target area);
- the specific steps taken to contact each member of this universe, whether through door-to-door canvassing, the use of mass media, or announcements at community meetings;
- the content of the outreach materials; and
- the various levels of response--e.g., those who ultimately become participants vs. those who merely make initial inquiries.

The reason that this type of information is so important in assessing a citizen crime prevention activity is that the activity might otherwise be well conceived but suffer from a poor outreach effort. A proper description allows for comparisons among the outreach efforts of various projects, providing the basis for more appropriate guidelines.

Organizational Form

Every citizen crime prevention activity involves some type of organized effort, however formal or informal. A minimum yet essential amount of organization can simply be reflected in the telephone

or mailing lists that are maintained, for these lists form the basis for organizational membership. Other organizational factors may be:

- the formal organizational structure, including the ways in which leaders are selected;
- the source and amount of financial support;
- membership rules;
- the types of files or logs maintained by the organization;
- the existence of an organizational office and staff; and
- the formal affiliation between the organization and other organizations.

These organizational characteristics are also an integral part of any citizen crime prevention activity and must therefore be assessed. One previous study, for instance, showed that citizen patrols that involved some minimum level of formal organization were less likely to be considered vigilante groups than those with no formal organization.⁵⁶ More generally, however, it is necessary to define fully the organizational form of any citizen crime prevention activity just in order to replicate that activity at some subsequent site.

Relationship With the Local Police

No single characteristic of a citizen crime prevention activity appears to be more important than its relationship to the local police. At the same time, this is the same characteristic that is probably the most difficult to define or measure.

Positive relations with the police are important to a citizen activity because the police are the only legal agents when physical intervention is needed. Secondly, the police may choose to augment the citizen crime prevention activities through patrol tactics (especially for a small area), through the provision of communications and equipment support to the citizen group, or through the provision of information from police files. For all these reasons, it is usually recommended that citizen groups maintain a good working relationship with the local police (except in those few cases where the police may be the source of a service problem).

In spite of the critical nature of this aspect of citizen crime prevention activities, little attention has been given to its measurement. Conceptually, the relationship consists of both individual and organizational relations as well as interpersonal relations be-

tween residents and individual police officers. Neither is easy to assess, but the various indicators that could be monitored; depending on the type of activity, might include:

- the number of meetings between the citizen group and the police;
- the number of officers known by the members and, conversely, the number of members known by the beat police officers;
- the amount and type of information provided by the police to the citizen group (e.g., from police records) and, conversely, by the citizen group to the police (e.g., from witnesses' reports);
- the supplies or equipment shared by the police and the citizen group; and
- the amount of time spent by the police in orienting or educating members of the group concerning crime prevention practices.

In contrast to the use of these and other indicators, previous studies have generally been limited to broad assertions such as ". . . the police have become a supporter of the program." This has been the situation even with the evaluation of major crime prevention projects such as the Seattle project, in which police cooperation was considered essential and in which an exemplary project was presumably being described.⁵⁷

Implementation Process

The implementation process consists of the efforts made to get an activity from its design phase to its action phase. The process is important because initially similar designs may result in entirely different actions at different sites. Second, the implementation experience may create permanently antagonistic relationships--e.g., between leaders and members, or between the citizen group and the police--and these may later affect the outcomes of the crime prevention activity. Third, implementation may directly affect the crime prevention activity's potential effectiveness, as in the description of Operation Ident in a previous section of this paper.

Again, little attention has been given to the problem of assessing the implementation portion of a citizen crime prevention activity. Here, conceptualization is needed as much as measurement, for little is known about the implementation process in the first place.⁵⁸ For crime prevention activities, for instance, it is not known whether an activity is better off if implemented with police

leadership, with citizen leadership, or through some joint task force.⁵⁹ Similarly, it is during the implementation process that leaders must strike a careful balance between the amount of volunteer help or other resources available and the coverage of the group's initial agenda; in many instances, even a gross mismatch between resources and expectations is not recognized until implementation occurs.

The Activity Itself

The core of any assessment must obviously deal with the specific nature of the citizen crime prevention activity itself, whether it is a public education project, a surveillance project, a patrol project, or any combination of the projects previously listed in Figure I. Naturally, because of the diversity of these approaches, few generalizations can be made about what needs to be defined and measured. However, one good example, involving police deployment rather than citizen crime prevention, should suggest the types of operations that need to be assessed. This example comes from a review of innovations in team policing, in which detailed sections of the review covered the various operations of the teams: team organization, team responsibilities, patrol workload management, non-crime-related services, and community relations activities.⁶⁰ In principle, citizen crime prevention activities should also be assessed by measuring their routine operations. The information for such operations can be based on logs or other records kept by the citizen group.

A common failing in such assessments has to do with public education efforts. In such projects, the "activity" is the dissemination of information about crime prevention practice. Studies commonly fail, however, to indicate:

- precisely how the information was disseminated;⁶¹
- the frequency and intensity with which information was disseminated;
- the content of the information; and
- the degree to which respondents appeared to understand the material.

Such educational projects, in short, must be treated like educational efforts more generally, not necessarily limited to crime prevention. In order to interpret any behavioral outcomes, the basic curriculum must be defined. At a minimum, research reports should contain examples of the pamphlets or other messages that were used. For example, one report of citizen activities gave sample scripts of the 30-second public service announcements that were part of the crime prevention effort.⁶² Most reports, however, merely state that

materials were distributed or media announcements made, but give no indication of the information content of these messages.

A second common failing of most assessments, not limited to public education projects, is that the measurement of the amount of effort involved in the activity is usually ignored. In some cases, as in security surveys, this effort can be directly translated into police officer time and costs. Yet, a recent review did not even raise the issue of assessing the level of effort.⁶³ In other cases, as in voluntary activities, the number of volunteers and amount of volunteer time can be recorded. Again, most reports do not even attempt to assess the level of volunteer effort.⁶⁴ One exception to these general observations was a report on a Minnesota crime watch program, which covered the total number of crime prevention units and the agencies' hours spent in crime prevention activity.⁶⁵ Future research should give much closer attention to this type of assessment, because the level of effort of any activity will obviously affect the outcomes, regardless of the nature of the activity. Conversely, the same outcomes should not be expected from a well-funded and sizable effort as from a poorly supported effort, even if the two efforts involve the same type of activity.

CONCLUSION

Citizen crime prevention activities are believed to be essential components of effective crime prevention. Yet, such activities are exceptionally difficult to define in an operational manner; the activities themselves benefit from the same voluntary and informal communications behaviors that are the most difficult for social scientists to assess.⁶⁶

It is incumbent on further research, however, to develop improved means of measuring the activities because, without such measures, few conclusions can be drawn about outcomes (even where outcome data are available) and only imprecise replication of the activity can be pursued at other sites. Simply put, if we don't know what happened at a site, little can be said about the causes for its success or failure, and little guidance can be given to those wanting to duplicate the experience elsewhere.

The main purpose of the present paper has been to show that the matter of operationally defining citizen crime prevention activities has received scant attention. Numerous reports exist that use a simple label or otherwise avoid the problem altogether, even though these same reports may provide good coverage of research design issues and outcome measures. Future studies must begin by describing, in an operational manner, what was involved in mounting a crime prevention activity. The present paper has suggested that this involves at least five separate subactivities:

outreach efforts, organizational form, relationship with the local police, implementation, and the activity itself. Future methodological research is needed to show how these and related behaviors can be measured in the most economic fashion.

NOTES

1 The concepts "crime vulnerability" and "crime pressure" were introduced by Arnold Sagalyn et al., Residential Security (Washington, D.C.: U.S. Department of Justice, 1973).

2 Bernard Cohen, Research on Criminal Justice Organizations: The Sentencing Process, R-2018-DOJ, (Santa Monica, Calif.: The Rand Corporation, December 1976).

3 Jane Jacobs, The Death and Life of Great American Cities (New York: Vintage Books, 1961), pp. 31-32.

4 Gerald Suttles, The Social Construction of Communities (Chicago: University of Chicago Press, 1972), p. 234.

5 Ibid., p. 264.

6 Ibid., p. 247.

7 A. M. Rosenthal, Thirty-eight Witnesses (New York: McGraw-Hill Book Co., 1964).

8 Warren Lehman, "Crime, the Public, and the Crime Commission," Michigan Law Review, 66:1487-1540, May 1968. The Genovese case also led to major new lines of social psychological research, which today are collectively regarded as "by stander" research.

9 James Q. Wilson, and Barbara Boland, "Crime," in William Gorham and Nathan Glazer (ed.), The Urban Predicament. (Washington, D.C.: The Urban Institute, 1976), pp. 179-230.

10 James Q. Wilson, Thinking about Crime (New York: Basic Books, 1975).

11 George Kelling et al., The Kansas City Preventive Patrol Experiment (Washington, D.C.: Police Foundation, 1974).

12 There have since been a number of criticisms of the Kansas City experiment, including the point that the variations in patrol deployment may not have been effectively implemented. Vide Richard C. Larson, "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment," Journal of Criminal Justice, 3:267-298, 1975.

13 James P. Levine, "The Ineffectiveness of Adding Police to Prevent Crime," Public Policy, 23:523-545. Fall 1975.

14 David H. Bayley, "Learning about Crime--the Japanese Experience," The Public Interest, 44:55-68, Summer 1976.

15 As recently as 1976, the Congress passed new legislation setting aside increased funds specifically for citizen crime prevention activities. See the Crime Control Act of 1976 (P.L. 94-503), which resulted in the creation of an Office of Community Anti-Crime Program within LEAA and an allocation of \$15 million for such activities.

16 George J. Washnis, Citizen Involvement in Crime Prevention (New York: Praeger, 1976).

17 Robert K. Yin et al., Patrolling the Neighborhood Beat: Residents and Residential Security, R-1196-HEW, (Santa Monica, Calif.: The Rand Corporation, March 1976).

18 Gary T. Marx and Dane Archer, "Community Police: An Exploratory Inquiry." Unpublished paper. (Cambridge, Mass.: Harvard-MIT Joint Center for Urban Studies, June 1972).

19 Leonard Bickman et al., Citizen Crime Reporting Projects (Washington: U.S. Department of Justice, April 1977).

- 20 Erving Goffman, Relations in Public: Microstudies of the Public Order (New York: Basic Books, 1971).
- 21 Oscar Newman, Defensible Space: Crime Prevention through Urban Design (New York: Macmillan, 1972).
- 22 Ibid., p. 206.
- 23 Oscar Newman, Defense Guidelines for Creating Defensible Space (Washington, D.C.: U.S. Department of Justice, 1976).
- 24 Thomas A. Reppetto, Residential Crime (Cambridge, Mass.: Ballinger, 1974).
- 25 Ibid., p. 87.
- 26 Richard A. Smith, "Measuring Neighborhood Cohesion: A Review and Some Suggestions," Human Ecology. 3(3):143-160, 1975.
- 27 For a further description of practice-to-practice (non-R&D-based) versus research-to-practice (R&D-based) improvements, see Task Force on Criminal Justice R&D. (Washington, D.C., 1976).
- 28 Washnis, 1976, pp. 7-32.
- 29 The satisfactory definition typically provides an adequate response to the the seemingly disrespectful question, "How would I know a block organization if I saw one in the field?"
- 30 Peter Rossi and Walter Williams (ed.), Evaluating Social Programs. New York: Seminar Press, 1972); Carol Weiss (ed.), Evaluating Action Programs: Readings in Social Action and Education. (Boston: Allyn and Bacon, 1972); James A. Caporaso and Leslie L. Roos, Jr., Quasi-Experimental Approaches. (Evanston, Ill.: Northwestern University Press, 1973); Marcia Guttentag and Elmer L. Struening (ed.), Handbook of Evaluation Research. (Beverly Hills, Calif.: Sage Publications, 1975); Ilene N. Bernstein (ed.), Validity Issues in Evaluation Research. (Beverly Hills, Calif.: Sage Publications, 1976).
- 31 Wesley G. Skogan, next chapter of this volume.
- 32 Michael D. Maltz, Evaluation of Crime Control Programs. (Washington: U.S. Department of Justice, April 1972).
- 33 Donald T. Campbell and Julian C. Stanley, Experimental and Quasi-Experimental Designs for Research. (Chicago: Rand McNally, 1963).
- 34 James M. Tien et al., Street Lighting Projects. (Washington: U.S. Department of Justice, July 1977).
- 35 INTASA, Combating Felonious Crimes by Citizen Involvement. (Menlo Park, Calif., 1972); R. E. Kirkpatrick, Philadelphia Neighborhood Anti-Burglary Project. (Reston, Va.: BFH Associates, 1974).
- 36 Ron Pennington, "An Evaluation of the Open Garage Door Burglary Program." Paper presented at National Conference on Criminal Justice Evaluation, Washington, D.C., February 1977; Donald Q. Brodie and David I. Sheppard, "Neighbors Against Crime Together." Paper presented at National Conference on Criminal Justice Evaluation, Washington, D.C., February 1977.
- 37 International Training, Research, and Evaluation Council, Crime Prevention Security Surveys. (Washington: U.S. Department of Justice, January 1977).
- 38 For example, see Michael Q. Patton, "Evaluation of Program Implementation, in Utilization-Focused Evaluation (Beverly Hills: Sage Publications, 1978).
- 39 Barbara Dietrick, "The Environment and Burglary Victimization in a Metropolitan Suburb." Paper presented at the annual meeting of the American Society of Criminology, Atlanta, Ga., November 1977.
- 40 Gerald Luedke and Associates, "Crime and the Physical City" (Detroit, Mich., June 1970).

41 Dennis C. Duffala, "Convenience Stores, Armed Robbery, and Physical Environmental Features," American Behavioral Scientist, 20:227-246, November/December 1976.

42 Carol Bevis and Julia Brown Nutter, "Changing Street Layouts to Reduce Residential Burglary." Paper presented at the annual meeting of the American Society of Criminology, Atlanta, Ga., November 1977.

43 Dietrick, 1977.

44 Barbara B. Brown and Irwin Altman, "Territoriality and Residential Crime." Paper prepared for Westinghouse National Issues Center, Arlington, Va., 1977.

45 Larson, 1975.

46 Newman, 1972.

47 C. Ray Jeffery, Crime Prevention through Environmental Design (Beverly Hills, Calif.: Sage Publications, 1971); William Brill Associates, "Controlling Access in Highrise Buildings." Report submitted to U.S. Department of Housing and Urban Development, July 1976; William Brill Associates, "Victimization, Fear of Crime, and Altered Behavior." Three separate reports published by the U.S. Department of Housing and Urban Development, 1976; and Thomas A. Repetto, "Crime Prevention through Environmental Policy," American Behavioral Scientist. 20:275-288, November/December 1976.

48 For example, see Newman's own handbook (1976) and the HUD brochure on "Residential Security" (HUD--440-PDR, August 1976). For a full review of this recent literature, see Margaret K. Gwaltney and Robert K. Yin, "Designing Safe Environments: I. What is a Feature?" Unpublished paper. (Santa Monica, Calif.: The Rand Corporation, 1978).

49 Janet G. Frohman et al., "Special Report on CPTED," Nation's Cities. 15:14-28, December 1977.

50 Nelson Heller et al., Operation Identification Projects: Assessment of Effectiveness (Washington: U.S. Department of Justice, August, 1975).

51 For a thorough discussion of implementation problems in public programs, see Robert K. Yin et al., Tinkering with the System: Technological Innovations in State and Local Services (Lexington, Mass.: Lexington Books, 1977); and Robert K. Yin, Changing Urban Bureaucracies: How New Practices Become Routinized (Lexington, Mass.: Lexington Books, 1979).

52 Heller et al., 1975.

53 Skogan, next chapter.

54 Heller et al., 1975.

55 John D. Carr, "Operation Identification: Interim Evaluation Report." Paper presented at National Conference on Criminal Justice Evaluation, Washington, D.C., February 1977.

56 Yin et al., 1976.

57 Office of Policy Planning, City of Seattle, "Policy Development Report on the Community Crime Prevention Program." October 1976; and Paul Cirel et al., Community Crime Prevention Program, Seattle, Wash. (Washington: National Institute of Law Enforcement and Criminal Justice, September 1977).

58 Yin et al., 1977; and Yin, 1979.

59 For example, see Federal Bureau of Investigation. Crime Resistance (Washington, 1977), for a description of four task force efforts organized to promote citizen-based, self-help crime prevention.

60 William B. Gay et al., Neighborhood Team Policing (Washington: U.S. Department of Justice, February 1977).

61 For example, see the useful distinction between group presentation versus home presentation projects made by Bickman et al., 1977.

62 Federal Bureau of Investigation, 1977.

63 International Training, Research, and Evaluation Council, 1977.

64 E.g., Federal Bureau of Investigation, 1977.

65 John Richardson and Peter Hartjens, "Methodological Problems Facing the Crime Prevention Program Evaluator: The Minnesota Crime Watch Experience." Paper presented at National Conference on Criminal Justice Evaluation, Washington, February 1977.

66 Robert K. Yin, "Evaluating Citizen Crime Prevention Programs." P-5853. (Santa Monica, Calif.: The Rand Corporation, April 1977).

Community Crime Prevention Programs

Measurement Issues in Their Evaluation

Wesley G. Skogan *is associate professor of political science and urban affairs at Northwestern University. He is the editor of Sample Surveys of the Victims of Crime (Ballinger, 1976) and serves on the editorial board of several professional journals.*

INTRODUCTION

The prominent astrophysicist Sir Arthur Eddington once noted,

The whole subject matter of exact science consists of pointer readings and similar indicators. Whatever quantity we are "observing" the actual procedure nearly always ends in reading the position of some kind of indicator on a graduated scale or its equivalent.

All rigorous and empirical research thus shares at least two concerns, the accuracy with which indicators point to the right position on their "graduated scales" and whether those scales measure exactly what the researcher needs to know in order to come to some firm conclusion. Like others engaged in scientific enterprises, those involved in community crime prevention evaluations must confront the twin issues of utility and accuracy. If their measures do not adequately reflect what they really are trying to measure, then their conclusions must remain highly equivocal. In measurement terms, in those cases we are uncertain of the validity of our measure of something. Certainly all who have attempted to measure reductions in the rate of crime using official police statistics have had to contend with such shortcomings. Even if the measures employed could validly represent the real phenomenon of concern, the accuracy with which they do so still must be demonstrated. Recent research on the measurement of victimization conducted through personal interviews suggests, for example, that simple differences between interviewers employed by the Census Bureau for the same project had an enormous impact on the "true" rate of crime in eight large cities.¹ In this case, we are uncertain of the reliability of those estimates, for different interviewers give us different "pointer readings."

The penalties for ignoring the twin issues of reliability and validity are considerable. Measures which are not indicators of our real object of concern (like police counts of victimization) can easily lead us astray. If the linkage between measure and object is cloudy, the results of the evaluation become highly interpretative. One recent report, for example, counted a rise

Copyright (c) 1979 Wesley G. Skogan.

in official statistics as a "success" because (the evaluators presumed) it indicated that the Philadelphia Transit Police were more effectively keeping track of crimes in their jurisdiction.² Another evaluation, conducted in the same year, counted a decrease in official statewide crime counts as an indicator of success because it followed their public-relations program,³ while at the same time yet another researcher⁴ was arguing that effective crime prevention programs probably will both increase and decrease official crime rates, at the same time.

Unreliable measures can lead to as many problems as invalid ones. The problem with unreliable measures is that they do not give us stable readings; one measurement does not agree with another, and repeated measures fluctuate over short periods of time. As a result we may confuse unreliable findings with true change, or be forced to pick and choose among similar measures which point to different conclusions. Changes in the "true score" component may be lost in a welter of sources of extraneous variation, including sampling variation and indicator instability. A report on a recent evaluation of a statewide crime prevention program, which employed pre- and post-intervention surveys, appears to have fallen into this trap. The report reputedly focused only on survey items that showed some change, and ignored the remainder because "they didn't tell anything."⁵

Despite textbook injunctions to pay attention to these issues, it appears that only rarely do criminal justice evaluators do anything about them. In this essay I review a number of recent evaluation efforts focusing on criminal justice problems, but few of them will prove fruitful examples of "what to do." Most are illustrative of what could have been done, or what probably was useful but which remained undocumented. The measurement efforts in these evaluations focused on two general topics: interventions and outcomes. All employed data from one or more of three sources: observations, interviews, and archives. In another chapter in this volume Robert Yin details many conceptual and practical problems involved in answering the questions, "What actually was the intervention?" In this chapter I will review what others have done to measure those interventions and their consequences and tender some suggestions for improving the reliability and validity of measurement in this area.

CLASSES OF THINGS MEASURED

In reviewing evaluations of criminal justice programs it proved useful to classify "things measured" into two general categories: interventions and outcomes. Interventions themselves seemed to be measured in two distinct ways. In one group fell measures which were generated from the perspective (and usually

the data) of agencies or program officers. They often were based on the "official record" of those sources, which seemed to lend them a "presumptive validity." These measures typically purport to represent agency activity, effort, implementation, or service delivery. We shall call these "upstream" measures of intervention to denote their perspective relative to events in the field. They are the "view from headquarters." Some examples include accounts of staff activity, counts of households visited, or "ID" stickers distributed, and reports of expenditures. On the other hand, some studies generated measures of implementation which took more of a "consumer" perspective on interventions; these relied on citizen perceptions, reports of the actual receipt of messages, the adoption of recommendations, program participation, and the behavior of the targets of programs. A good example of these would be a measure of "the number of security survey recommendations actually adopted." These I have dubbed "downstream" measures of interventions, and they reflect the "view from the field." The category of measures employed to gauge the outcomes of those interventions also was varied. Some focused on crime and fear, or employed more indirect measures which presumably were responsive to those factors, whereas others were concerned with the delivery of services and the satisfaction expressed by the presumed beneficiaries of programmatic activities.

In many cases, the difference between indicators measuring the outcome of interventions and those measuring antecedent factors is a difference in the conceptualization of the program under examination. In an example discussed below, the flow of vehicular and pedestrian traffic through an area in Hartford was seen as a source of many problems in that community, so the degree to which it was reduced was employed as a measure of the success of an experimental program there. On the other hand, because the traffic flow seemed to contribute to victimization and community deterioration in the area, its control was seen as an intervention which presumably would have measurable consequences. In fact, the evaluation counted it both ways, arguing that the street-use intervention was the most important element in the success of the program.⁶

Thus, in this essay, we will not be extraordinarily concerned about the logical status of indicators in analytic schemes, but rather will focus on their more frequent uses and misuses.

MODES OF MEASUREMENT

Three kinds of data gathering were employed in these evaluation studies: observations, interviews, and the use of archives. The most common observational technique consisted of looking and counting according to a set of highly structured

rules, and often involved the use of checklists. Some examples of these include counts of people using city streets after dark, counts of street lights, and structured judgments about the security features of doors and locks. No one was dispatched to observe meetings or more fluid settings, or participated in the activities they were observing, thus limiting the range of methodological issues to be considered here.

The interviews were conducted in more heterogeneous fashion, ranging from indepth discussions with program officials to paper and pencil tests and telephone surveys. They dealt with two general classes of topics: self-reports of behaviors and expressions of attitudes or perceptions. People were quizzed about property marking, police patrolling, attendance at meetings, and what they had seen on television. The questions sometimes focused on activities which could have been observed, but as we shall see, there are limitations to observational studies which often make it more effective to gather behavioral data through interviews, albeit at some cost.

Archival sources do not yield another kind of information: rather, they involve unearthing useful records of observations and interviews conducted by others in the past, often in the course of their official duties. As such they suffer from all the limitations of observational and interview data, but typically all records of those difficulties have been lost or obscured. Also, these records often do not contain information about what we really want to know, but instead their existence encourages us to use what the original gatherers of the paper thought was important. Archival data typically generate measures of crime, program participation, and staff activity, along with routine "management" information (such as public housing vacancy rates) which may be interpreted as measures of program outcomes.

There are, in addition, two sources of data which do not appear to be used often in criminal justice evaluations: content analysis and ratings by experts. Again, no new sources of information are being overlooked: all data come from watching things or asking people about them. Content analysis usually involves systematically coding what other people have seen or heard and committed to paper, whereas ratings by experts are based on what they have seen or heard that has qualified them for the role. An example of the latter, outside the context of community crime prevention evaluation, is Wilson and Boland's use of expert ratings of the quality of police departments in a sample of cities as one variable in a model of the deterrent capacity of their criminal justice systems.

The specific measures to be discussed here organized by type of measure (observational, interview, or archival), and by wheth-

er they have been used primarily as measures of interventions ("upstream" and "downstream") or outcomes. In general it is preferable to employ mixed-mode measures (for example, both interviews and observations) to determine the extent to which they triangulate on the same conclusion. As we shall see below, this enables us to gauge the extent to which our measures are free of "method bias," and to assess their validity in rigorous fashion. However, few of the concepts used to evaluate community crime prevention programs have been operationalized in more than one way in the same analysis.

OBSERVATIONAL MEASURES

Introduction

Quantitative observational measures have enjoyed only limited use in most social sciences, and they are only infrequently employed in evaluation research. Most observational data employed in the evaluation of criminal justice programs has been of a relatively simple variety, and the techniques employed to generate them provide the potential for highly reliable measurement. It should be relatively easy to employ at least some observational measures in most community crime prevention evaluations. However, although there are numerous advantages to data gathered by watching things and recording what you see, there are several important limits to observational techniques.

The principal advantage of observational measures is that data are collected on real-world phenomena as they appear or occur, not as they are recorded, remembered, or interpreted. If behavior is being observed, it typically occurs in natural settings and is sparked by naturally occurring stimuli. This is in contrast to interview measures, which may simply generate hypothetical answers to hypothetical questions. Further, the data are recorded "as they happen," freeing them from problems of forgetting and recall error which plague self-reports of behavior. If structural features of an environment or behavior context are being observed, data also may be generated which a respondent cannot supply.

Applications

A good example of an opportunity for observational measurement was Schneider's⁸ study of a burglary reduction program in Portland, Oregon. For that evaluation, interviews were conducted in about 3,900 homes in Metropolitan Portland. One focus of the survey was participation at various levels in a property-marking program (Operation ID) which was being publicized widely in the community. A key element of the program, which was probed

through the survey, was whether householders had engraved any of their valuables with an identification number and whether a sticker proclaiming this fact was visibly posted in order to deter potential burglars. A key evaluation measure in the study was the burglary rate for participating and nonparticipating households. However, rather than instruct the interviewer to look for such a sticker, respondents were asked if they had displayed one (12 percent said "yes"). While there may have been some advantage in knowing if a household had displayed a sticker in the past, but that for some reason it was not in place, central to the concern of the evaluation was the deterrent effect of its prominent appearance. This could have been directly operationalized by asking the interviewer if he or she could see it.

Interviewers were used as observers for other purposes in this evaluation. For example, the survey also was employed to evaluate a special street-lighting program.⁹ Respondents were asked about the adequacy of lighting in their neighborhood, and whether or not it recently had improved. In order to link these data with the lights visible from the front entrance of each respondent's home, interviewers were instructed to count all street lights visible from the front entrance of each respondent's home. In a related survey of burglary victims¹⁰ interviewers also sketched a map of each block face to locate responding households vis others in the neighborhood. This enabled analysts to assess the extent to which neighbors potentially could surveil sample households, which should have affected their burglary rate. Interviewers also were asked to rate the deterioration of each neighborhood and to comment on any unusual design features of the sample households. All of these measures were to be employed at the analysis stage to explain the burglary rate for individual homes in Portland.

In a similar survey, Reppetto employed his interviewers as observers to assess the vulnerability of households to burglary. To this end they counted the number of "entry portals" (doors and windows) to each sample unit, rated the strength of their exterior doors, and noted the type of the building and its location on the block face. He found that residential burglary was encouraged by (a) accessible and inconspicuous windows, (b) "nonstandard" doors (which were thin or had exposed hinges, poor locks, or window glass near their handles), and (c) street-corner locations.¹¹

Certain advantages of the observation procedures utilized by Reppetto and Schneider should be noted. Because their interviewers already were being trained for their role, the incremental cost of adding additional training in the use of an observational form was relatively slight. As they were already searching out randomly sampled households, the cost of positioning the observers to gather the data was minimal. Further, the

data from the observations could be linked to the surveys, enabling analysts to combine the direct observations with attitudinal and self-report data.

This is important, for one of the major weaknesses of observational measures is that they cannot tap perceptions about, or the motives for, behaviors they may record. We can only infer the reasons why people do the things we see, although often the reasons are the point in evaluation research. Thus, observational studies of human activity (such as counts of people on the street, or using facilities) often require a very strong research design. We usually want to say "they are doing this because of crime" (or in spite of crime, or due to fear, etc.). This requires, however, inferences about the motivations or purposes which underlie the observed activity, and they may be various and diverse. What is required is research designs which can systematically rule out alternative interpretations of what we observe, leaving only the proffered explanation. This involves gathering data which test those alternative hypotheses (like, "they were out because there was nice weather").

In addition, observational studies are severely limited by their emphasis on the gathering of direct sense data. As a result, they cannot deal with events or conditions in the past. This is a considerable limitation, for we frequently are interested in what happened before we came on the scene, and explanations for people's present behaviors often may be found in past events like criminal victimization.

Methodological Concerns

The major methodological problem to be overcome in order to carry out the rather simple observations described above is one of observer reliability. The key is to insure that the same observer would give the same rating or count on two different trials, and that pairs of observers would agree on the proper rating or count for a given trial. Evidence of high interobserver agreement can only be obtained by rigorously comparing observations under controlled circumstances. If the resulting agreement scores are high, we at least know that Eddington's pointer will come to rest at the same point on the scale each time, which is the sine qua non of good measurement.

Such accuracy should be attainable for observations like those described here, given adequate observer training. Observational measures may not be reliable, however. For example, it now is clear that judgments about housing dilapidation made by Census Bureau enumerators for the 1960 census of population were highly unreliable. In an effort to increase the reliability of such ratings, the 1980 census of population may ask if there is "a hole in the wall" greater in area than the size

of the paper on which the questionnaire is printed. The latter would provide a common standard for judgment which would enhance greatly interobservational agreement about that aspect of housing deterioration. It also would be an improvement over the 1970 census, which focused almost exclusively on plumbing standards as a measure of housing quality. In that case reliability was enhanced at the expense of validity, for no one was substantively interested in the detailed data on the type and operability of plumbing arrangements gathered in that census. On the other hand, the validity of an observer's notation that he or she could spot an Operation ID sticker should be very high, for the operation involved in making the measurement almost exactly matches the phenomenon under investigation.

Schneider and Reppetto's use of survey interviewers as observers, which yielded "add-on" data for random samples of households, solved another problem of observational studies, that of sampling. Because observers can only count or judge that which lies before them, the positioning of observers in time and space relative to the occurrence of events or the presence of conditions in time and space yields cluster samples of observations with unknown sampling biases. This arises because the observations they make once they are in place are not independent of one another, in a sampling sense. Unlike people or households, we often cannot sample events or behaviors (or, to a lesser extent, conditions) with a priori knowledge which will enable us to judge the randomness or representativeness of the resulting sample. More frequently we stratify our observations using other criteria and hope for the best.

This sampling strategy was employed by Lavrakas et al.¹² to evaluate an ambitious crime prevention program in a commercial sector of Portland, Oreg. One intermediate goal of the program was to increase pedestrian use of the sector's main street after dark. Increased public use of the street was seen both as a measure of the program's success (as a "proximate goal") and as a mechanism for reducing further the incidence of crime and fear of crime (the "ultimate goals") along the shopping strip.

In order to measure the density of pedestrian traffic in the area the evaluators conducted over 70 evening field observations. During each of them an observer was twice driven the length of the 3.5-mile corridor, speaking continuously into a tape recorder. He described each person he saw either on the avenue or down one of its side streets, identifying individuals by sex, age, race, and activity. The tape recording was later replayed and the observations coded. These observations were made on five randomly-chosen evenings each week, and each was conducted at a randomly-chosen hour between 6 and 11 p.m. The data were gathered every other week for 7 months, straddling the implementation of the program. In the analysis they were merged

with information on time, temperature, and rainfall, all of which were substantially correlated with the number of persons observed on the street. A time-series analysis revealed no significant change in street-use levels or patterns during the study period.

A less sophisticated "before-and-after" analysis of observational data on street traffic recently was employed to evaluate the impact of environmental interventions in a high-crime area of Hartford, Conn.¹³ One goal of this program was to reduce the volume of pedestrian and vehicular traffic through the area. (Much of it was "pass through" traffic by strangers from the downtown area adjacent to the target site, and the congestion and unpleasant environmental impact of amounts of traffic appeared to have inhibited community development in the area.)

In order to evaluate the effectiveness of a number of traffic-flow reduction measures, including narrowing and blocking streets, counts were made from vantage points of automobiles and pedestrians passing through the area. The counts were made twice: once before the implementation period and once after. They revealed a slight (4 percent) decrease in vehicular traffic and no change in foot traffic. Not surprisingly, most of the decrease in the flow of automobiles was accounted for by the blocked streets.

Potential Applications

These examples of simple observational measures should suggest others. For example, one visible sign of community decay is the incidence of graffiti in public places. My own (survey-based) research suggests that the presence of graffiti and visible vandalism underlays at least some of the "fear of crime" which haunts urban neighborhoods. Richard Taub¹⁴ is employing systematic counts of the incidence of graffiti as one of several indicators of neighborhood deterioration in a study of crime in Chicago, while Oscar Newman¹⁵ is using observational data on defacement and vandalism to characterize public housing projects. For Taub the incidence of graffiti is one determinant of housing values, whereas for Newman it is an outcome measure which should be influenced by building management policies.

Other downstream observational measures of interventions could be employed by criminal justice evaluators. In the area of shoplifting one could dispatch observers to note the extent to which store owners employ various strategies to combat theft. This would involve observing the deployment and coverage of mirrors, cameras, and other surveillance hardware, as well as the presence of menacing warning signs. These measures could be used to evaluate either the adoption of theft-prevention pol-

icies (in which case they would be employed as outcome indicators) or the effectiveness of those programs (for which they would be intervention measures).

Robert Yin¹⁶ proposed observational measures of interventions that have been characterized here as "upstream" data, measured from the perspective of service providers. He wanted to record, for example, how frequently the police passed by sample observation posts, in order to measure the density of police coverage of a site. He also proposed traveling with citizen groups which patrol their neighborhoods in order to observe the frequency of "vigilante" tactics on their part, and against whom they are employed. Perhaps the best known work of the latter genre is Albert Reiss and Donald Black's¹⁷ observational study of the police. They deployed observers in squad cars patrolling high-crime areas in Boston, Washington, and Chicago. Their agents monitored police-citizen contacts and various aspects of officer behavior. They uncovered, among other things, high levels of official illegality and frequent expressions of racial prejudice, neither of which are attributes of police work easily recoverable from official archives or through standard interviews.

Summary

In sum, observational procedures offer great potential for capturing data which are both reliable and valid indicators of interventions and outcomes. Gathered in naturalistic settings, and free of much of the recall error which plagues self-report data, highly structured observations may best reflect "the real world." However, there are several substantial limitations to this sort of data. They cannot gauge the motives underlying observed actions, nor can they probe the past. Further, the less structured and rigid the observational procedure (the "richer the data"), the less reliable it is. Gaining high inter-observational agreement outside of a checklist data-gathering format can be extremely difficult, but as the resulting data become simpler we often begin to sacrifice validity (what we want to measure) for confidence (what we believe we can measure). The trick to generating useful observational data is to strike some balance between the two.

INTERVIEW MEASURES

Introduction

After official police reports, interviews are the most commonly used source of data for evaluations of community crime prevention programs. Structured interviews have been the a prominent source of data for most social sciences for close to 30

years. During those 3 decades researchers have gained a deep appreciation for the utility of the interview, if not for its substantial limitations. As we shall see, revealing as they often are, the interview-based measures used in most community crime prevention evaluations have been generated by quite primitive methods. With few exceptions they have no known reliability and little evidence of their validity ever is proffered. Probably the result has been to underestimate the apparent impact of most of those programs (but occasionally to overestimate the effects of others) and mistakenly report that "it didn't work."

Measures of Interventions

One of the most important contributions of citizen-based interview data on crime prevention activities has been to make us suspicious about claims of what has been done. Interviews usually are conducted with the targets of or participants in programs, and whenever they focus on program implementation and the extent to which "upstream" activities by agencies have had any discernable "downstream" impact among the citizenry they tend to reveal that there is a considerable gap between official activity and reports that anything actually has happened.

Nowhere is this more evident than in the case of security programs. These projects, usually conducted by police departments, involve inspecting homes or commercial establishments and then recommending changes which would increase their physical security. The inspectors may suggest better locks, alarms, or trimming nearby shrubs to make persons lurking among them more conspicuous. Programs which count themselves a success by accounting for their activities often report a formidable number of such inspections; those which follow them up to see which of their recommendations actually were implemented usually are more humble about their efforts.

For example, a Wilmington "Mr. Victim" program was evaluated by surveying 800 homes and 245 businesses to check on compliance with security survey recommendations: only 29 percent of households and 26 percent of businessmen reported undertaking any of those suggestions.¹⁸ In Seattle, a 20 percent followup of home inspections found only 38 percent even in partial compliance.¹⁹ In Eden Prairie, Minn., 62 percent of households reported making some change, but only 29 percent of the specific recommendations preferred were taken to heart. In Eden Prairie those tended to be the less expensive ones, and changes that householders could make for themselves.²⁰

Similar interview-based downstream evaluations of program implementation have been conducted for a variety of other citizen and business-oriented projects. Like security-survey evaluations, they typically reveal modest levels of program adoption,

participation, or impact. Schneider's²¹ study of a burglary reduction program in Portland found that about 10 percent of households were represented at some community meeting about crime, and that about 12 percent reported displaying an Operation ID door-or-window sticker. In her study of Portland's intensive street-lighting program²² only 25 percent of the residents of the newly lighted district thought that there were more lights and fewer thought their area was well-lit than did residents of other parts of the city. Moreover, despite substantial ballyhoo about the program, only one of her 3,900 respondents brought it up when asked in an open-ended question if they knew about any special crime prevention programs in the city. Considering the low visibility of the lighting effort, it is not surprising that the evaluators found little impact of the program on perceptions of neighborhood safety.

More success in delivering services on the downstream end of the intervention process was reported in a recent evaluation of a crime reduction program in a public housing project in Chicago. One element of the program was the stationing of a Resident Safety Aide in each experimental building in the project. The Aides were to talk to building residents about safety programs, encourage crime reporting, and engage in security patrols. They were dressed in uniforms and carried walkie-talkie radios while on patrol. As part of the evaluation, building residents were asked how often they saw the Aides, and whether they were satisfied with their assistance. Fully 42 percent of the residents reported seeing the Resident Safety Aide "every day," and another 20 percent at least "a few times a week."²³

Problems of Interview Measures

Interesting as such findings are, it is not clear to what extent they adequately represented the true extent of program implementation in these jurisdictions. All of these studies relied on self-reports of behavior or experiences by the respondents, and they probably are affected in various ways by the "social desirability" problem in survey research. We know that in interviews, as in life, people generally prefer to present themselves and their actions in the best possible light. Studies have revealed that people are likely to substantially overreport doing such "good" things as getting a library card, voting, and contributing money to a charitable organization.²⁴ For most, participating in civic programs and complying with the recommendations of visiting police inspectors presumably would fall into that category. We also know that it is risky to employ a single member of a household as an informant about family activities. This procedure was used in the first victimization surveys conducted for the Crime Commission,²⁵ and it resulted in the substantial undercounting of criminal events. It is apparent that even adults are not fully informed about the doings and experiences of others in their household, and cannot be

expected to report on them with accuracy. This should lead us to be suspicious of surveys which attempt to characterize households as being represented or not at community meetings about crime, or as participating in crime prevention programs.

Invalid responses to questions about programmatic interventions also are encouraged by the great distance between program offices and the lives of most citizens. This often calls for a thorough prompting of respondents before responses about those programs are forthcoming. It is probable that they can guess what the interviewer "wants" to hear. In more general terms, interviews have "demand characteristics" which encourage certain kinds of behavior by respondents in order to satisfy that demand; pleasing the interviewer is one of them, and the puzzle presented to respondents usually is that of discerning the questioners' needs. For example, to evaluate a property-marking program in Denver, 2,000 residents were telephoned and asked if they had heard of the program. Approximately 80 percent said they had, but few recognized the name of the project and most had to be prompted with a description of the program before they managed to recall that they had heard of it.²⁶

When respondents are not filled in with the details, fewer seem to recall hearing of programs. In a statewide survey conducted to evaluate a crime prevention program, respondents were asked, "Do you know if a security check program is available in this community?" Not only did a substantial number not know, but in addition between 25 and 47 percent of the respondents in various subgroups being analyzed were "not sure" if they knew or not.²⁷

Errors of a different kind are encouraged by questions which place impossible demands on respondents. Usually they will do their best to answer them, but it is not clear what those answers denote. One of the most common offenses in this regard is to ask a question which depends heavily on the ability of respondents to recall past events. Those memories may be of great interest to program administrators, for often they are evidence that bureaucrats are doing their jobs, but the actions probably were not of much interest to the respondents. As a result, we should expect the forgetting rate for such events to be very high. For example, in repeated statewide surveys evaluating a public information program, respondents were asked, "Have you ever seen or received any information about protecting your home from burglary?"²⁸

In addition to the vagueness of the information being sought, the "reference period" for this question--the respondent's entire lifetime--presents impossible demands on the person being interviewed.

The same is true of the demand imposed by another question in those surveys, "Have you ever been the victim of a crime?" As we shall see below, victims of crime often forget or make errors

when recalling information about their experiences, and those difficulties increase sharply with the passage of time. Studies of the recall of "objective" events or activities as diverse as making home repairs and visiting a doctor indicate that recall errors are enormous even when the elapsed time is quite brief.²⁹

Further, because respondents have a story to tell and because they like to give interviewers the information they "want," they (apparently unconsciously) shift the placement of those events in time if necessary so that they can report them in the interview. (This is called "forward telescoping.") It is necessary to probe carefully the temporal placement of events, setting up clear boundaries for the reference period of the question (like, "since the first of the year") in order to control this quite powerful impulse.

The general message is that evaluators gathering data on interventions through interviews must be content with information of the sort that respondents can supply reliably. These are likely to be data on events which were salient when they occurred, which they were likely to be informed about, which recently occurred, which took place within salient time points, which they are likely to identify and label in the same terms as those seeking to elicit the information, and which are not so desirable that they would be embarrassed to admit not participating in them.

Estimating Reliability and Validity

Evaluation researchers must expend considerably more effort developing measures of interventions which have some demonstrable reliability and validity. Again, reliable measures will give the evaluator consistent readings from indicator to indicator, and over time. As this suggests, the reliability of a measure is estimated from the agreement between two concurrent measures of the same thing ("internal consistency"), or from the agreement of two measures repeated over some reasonable length of time ("test-retest stability").

An example of a survey-based indicator with relatively good reliability is Baumer's³⁰ measure of individual precautionary activity. He combined responses to four survey items which were described to respondents as "things that some people do to protect themselves from being attacked or robbed on the street." The precautions included "getting someone to go with you," "going by car rather than walking," "taking something with you that could be used for protection," and "avoiding certain places in the neighborhood." In every case respondents were asked whether they did those things "most of the time," "sometimes," or "almost never," at night. The average correlation between these items

was about +.37, and when the responses were factor-analyzed they proved to be unidimensional. Added together to form a "personal precaution" scale, they had a reliability of .70. This indicates that about 70 percent of the total variance in that scale can be attributed to consistent responses to the four individual survey questions, which is substantial.

The validity of survey measures of interventions probably could best be estimated by employing the "reverse record check" procedure widely used in health research. This involves developing questionnaires by pretesting them on samples of subjects known to have done certain things or to have certain attributes. For example, survey items measuring health-care consumption have been tested on persons drawn from lists of hospital patients and physicians' records. This enables researchers to estimate the effectiveness of the survey in accurately eliciting the desired information, such as the cost of medical care or the exact duration of these visits. In the community crime prevention area, Lavrakas³¹ is now evaluating questionnaire items by interviewing samples of participants in various crime prevention programs. The purpose of the interviews, as well as the source of the respondent sample, is "blind" to the interviewers in order to control any motive on their part to probe for the "correct" answer.

By extension, reverse record checks (which are "criterion validations" of measures) could be employed to determine if other measurements in the criminal justice field were yielding the intended information. Thus, Schneider³² and the evaluators of the Minnesota Crime Watch³³ could have demonstrated that their questions on property marking, home security surveys, meeting attendance, and receiving crime information had high "recovery power" when employed under field conditions.

This is important, for studies of the relationship between survey measures of implementation or service delivery and other measures of the extent to which those services have been deployed have not all shown that survey measures and "objective" counts reveal the same distribution of benefits. Ostrom³⁴ found that survey questions and electronic light meters tended to produce relatively similar measures of the adequacy of street lighting in small areas, but Schneider and Reiter³⁵ found only a low correlation between their doorstep count of street lights and householder ratings of how well lighted the area was, and that residents of their experimental lighting area did not seem to have noticed that their area had changed. As we have seen, there are a number of reasons why apparently valid measuring scales (those which make perfect sense to the investigator) may not in fact recover the intended information, and therefore some evidence of the actual validity of important measures should be presented in any sophisticated evaluation report.

Measures of Outcomes

Because they so easily lend themselves to gathering data at the "grass roots," interviews frequently are employed to measure the outcomes of interventions and the consequences of policies and programs. Because they focus on people and their experiences, perceptions, and attitudes, interview data potentially are more reliable and valid indicators of outcomes than they are of interventions, which often involve things about which individuals are not well informed. While many interview-based measures of outcomes have been employed in community crime prevention evaluations, in this section we will focus on only four: victimization, fear, satisfaction with programs or services, and reporting crimes to the police.

A. Victimization

Developed to the point where it is a relatively reliable and valid survey research tool only during this decade, the victimization survey has become an important instrument for criminal justice evaluators. Victim surveys have been utilized to evaluate a number of notable crime prevention experiments, including those in Seattle,³⁶ Portland,³⁷ Hartford,³⁸ Kansas City,³⁹ and a Chicago public housing project.⁴⁰ Properly conducted, these surveys have several advantages over other sources of information about crime. They can gather information about crimes which are not reported to the police, probe for details about incidents and their consequences in ways the police routinely do not, and gather these data within a larger context of information about the experiences, perceptions, and attitudes of victims. Moreover, unlike the police, they can gather much of this information from people who have not been victims of crime as well.⁴¹

Unlike most evaluation measures reviewed here, some evidence about the reliability and validity of measures of victimization has been produced through carefully executed methodological studies. Methodological investigations conducted in Great Britain⁴² and the United States⁴³ have led to the development of techniques for gathering reports about crimes from their victims that have relatively high "recovery power."

The basic tool in this developmental research has been the reverse record check technique described above. In the case of victimization, questionnaire items and interview procedures have been developed using samples of known crime victims, enabling researchers to evaluate what kinds of crimes are being missed in their questioning. In a typical reverse record check study⁴⁴ interviewers are given lists of households or persons to be covered with the current version of the survey instrument, without being told that imbedded within this sample are crime victims who have been selected from police records. After these interviews have been completed, they

are compared with data from the police reports. Two important questions can be addressed using this technique: does the survey recover the incident, and is it described accurately?⁴⁵

Using these methods, the Census Bureau developed some of the procedures it currently employs in the National Crime Survey. That survey seems to do fairly well at recovering accurate information about serious crimes such as robbery, burglary, auto theft, and even rape. It appeared that victims began to forget details about their experiences fairly rapidly, so the Bureau currently asks only about crimes which have occurred "in the last 6 months." It is important that this time frame (known as the "reference period" for the question) be bounded in some way with a salient date (like, "since the first of the year") or by an event. The Bureau uses a previous interview with the respondent as the reference point, taking advantage of the fact that the NCS is a panel study. The Census Bureau's victimization questionnaire also is experience based. That is, rather than ask respondents if they have "been a victim" of some crime, they are asked a series of questions about things which may have happened to them (like, "someone trying to hit you with a rock or bottle"). This resolves the definitional issue of "what is a crime?" which may vary greatly from group to group in the population, by relying on a conceptualization of victimization which is behaviorally based. Based on 30 years of experience in the measurement of unemployment, this approach enhances the reliability of the measure. Reliability also is constantly monitored by the Bureau through a vigorous reinterview program which yields test-retest estimates of the stability of their measures.

Despite all of this development effort, the methods employed in the National Crime Survey still have some problems. It appears that the survey does not adequately measure attempted assault and less serious aggressive crimes against blacks, and reverse record checks indicate that assaults and other crimes which involve people who are related or who know each other are being undercounted by a substantial margin. In general, crimes which do not usually involve strangers are poorly represented in current NCS victimization data.

It is also clear that victimization surveys, which depend on a series of "screener" questions to jog respondents' memories about past events, rely heavily on the skill and dedication of interviewers. Because this varies, there appears to be a substantial "interviewer effect" in victimization data. A recent report⁴⁶ documents the strength of this effect for eight cities surveyed by the Census Bureau in 1974. Interviewer variance in those data seems to be of about the magnitude of sampling error, which is considerable. Unpublished tabulations from the Bureau indicate also that rates for certain hard-to-measure types of crime ("series" incidents) dropped during the first year of the national program

as interviewers learned to cope with their new duties for this survey. Interviewer training and supervision must be managed carefully to combat these problems.

The dominant methodological issue affecting the use of surveys to measure crime is really a "substantive" one, however. The difficulty stems from the fact that crime is a rare event. Even in high-crime areas most people are not victimized during any brief period of time. As a result, within the reference period for a victimization survey, few crimes will be described. In the Census Bureau's city surveys, which asked about events during the previous year, only about 4 percent of all respondents recalled being robbed, and even fewer women told of a sexual assault. Thus, when those data are used to make population estimates of the victimization rate the error in those estimates is potentially quite large, and the "confidence interval" we must draw around them covers a broad range. This range is so great that evaluators who plan to compare the results of two surveys to look for the effect of a program must hope for a very large change in the victimization rate so that it will fall outside the boundaries of chance fluctuation. For any reasonably sized survey the required effect often is greater than any planner would ever realistically project for a crime prevention program.

For example, victimization surveys conducted to evaluate the Seattle Community Crime Prevention Program⁴⁷ produced "before and after" victimization rates for their experimental areas which indicated a 36 percent reduction in burglary--but this difference was not large enough to be statistically significant. This was true in spite of the fact that almost 1,500 households were questioned in the first round of interviewing and 1,200 in the second round, that these interviews all were concentrated in only five census tracts, and that the survey examined the incidence of the most frequent major crime. Crime prevention programs that focus on personal crimes or on small subgroups of the population like the elderly will quickly discover that the laws of sampling and statistical inference make victimization surveys extremely expensive to conduct, and that even more extreme effects must be projected for them to yield any evidence of positive outcomes.

B. Fear

Reports of fear of crime often are employed as ancillary measures of the success of crime prevention projects, whereas programs aimed at community renewal, commercial revitalization, and the elderly usually see reductions in fear as one of their most important goals. The evaluation of a crime reduction intervention in Hartford used measures of both fear and estimates of victimization rates in target and control areas. The evaluators found that those living in the experimental area perceived a significant drop in their chances of being victimized by burglary, while estimates of the probability of this victimization rose in the rest of the city.⁴⁸

A continuing evaluation by the National Council of Senior Citizens of a multicity project to assist elderly crime victims and decrease victimization of those over 60 years of age is employing survey measures of concern about crime and crime-related behaviors to assess their progress.⁴⁹

Several conceptual definitions and many different indicators have been employed under the general rubric of "fear." Most evaluations have fielded measures either of concern about crime or estimates of risk of victimization under that heading. Measuring risk involves asking respondents to rate their chances of falling victim to specific crimes, usually on a 0-to-10 scale. These indicate "how likely it is to happen." These were employed in the evaluation of the Kansas City response time evaluation.⁵⁰ Measures of concern, on the other hand, focus on what could happen, or how bad it would be if one were a victim. In their evaluation of a crime-reduction program in a public housing project, Arthur Young & Co.⁵¹ asked residents:

Thinking about your personal safety when you are outside after dark in this development, would you say you are very concerned, a little concerned, or not at all concerned?

The fear-of-crime measures most frequently employed in public opinion polls are of this genre. One, which is popular, was used by the Census Bureau in its city victimization surveys:

How safe do you feel, or would you feel, being out alone in your neighborhood at night--very unsafe, somewhat unsafe, reasonably safe, or very safe?

Another, employed by the Gallup organization, asks if there are "places in your neighborhood" where respondents "feel afraid to walk at night," and employs a "yes-no" response format.

Note that, unlike estimates of risk of victimization, concern questions clearly focus only upon personal attack and injury. This may be advantageous, for many suspect that a visceral sense of fear is at issue only in the case of crimes against the person. However, this conceptualization of the problem cannot yield comparable information about property crimes, which are the focus of most community crime prevention programs.

Measures of both concern and risk have been used with some effectiveness to evaluate crime programs. In the public housing project experiment residents were asked how "fearful" they were in a number of specific building environments, including elevators, stairwells, and the lobby. Information like this could be used by planners to determine where to locate television cameras and other surveillance or alarm devices. In Kansas City, concern measures were employed to evaluate whether a reduction in police patrol

frequency would affect citizen's perceptions of their personal safety.⁵² In Cincinnati, the Police Foundation used measurements of fear to evaluate a neighborhood team policing experiment in one police district there.⁵³ Schneider and Reiter⁵⁴ report that there was only a low correlation between how people residing in well-lit areas of Portland thought their neighborhoods to be and how safe they felt alone on streets in the area after dark.

None of these reports have concerned themselves with how reliable and valid these measures of fear have been. We usually have, at best, only evidence of their "construct validity." That is, people who "should" be afraid are. In a re-analysis of survey data from Minneapolis by neighborhood, McPherson⁵⁵ demonstrates that people who live in high-crime areas (as measured by police statistics) are indeed more likely to report being afraid. A recent analysis of the Gallup-poll fear-of-crime measure using national survey data collected by the National Opinion Research Center has provided evidence of the "discriminant validity" of that indicator. That analysis demonstrated the independence of "fear of crime" from other concerns and worries, including suspicion, distrust, and generalized anxiety about social change. The discriminating power of the fear-of-crime measure indicates that attitudes about the issue are not simply manifestations of other, more generalized concerns.⁵⁶

A recent survey provides evidence of the reliability of a scale constructed from commonly-employed measures of risk of victimization. In a random-digit-dialing telephone survey of three major cities respondents were asked:

For this next question I'd like you to think of a row of numbers from 0 to 10. Now, let the 0 stand for NO POSSIBILITY AT ALL of something happening, and the 10 will stand for it being EXTREMELY LIKELY that something could happen. On this row of numbers from 0 to 10 how likely do you think it is that...⁵⁷

This introduction was followed by capsule descriptions of the crimes of rape, robbery, burglary, and assault by a stranger.

Responses to these questions were used to form a two-item additive scale measuring risk of personal victimization. Estimates of the probability of being raped had to be dropped, for that question was asked only of women. Perceived risk of burglary, on the other hand, appeared to be relatively independent of that for personal crimes. Estimates of risk of victimization for robbery and stranger assault were strongly correlated, and together they formed a scale with a reliability (Cronbach's Alpha) of .83.

C. Satisfaction

Another common tool used in program evaluation is to focus on the level of satisfaction expressed by participants or the targets of those efforts. The bulk of these evaluations have focused on government agencies and the delivery of public services. In Kansas City, residents of experimental and control areas were asked to generally rate the quality of police service, and those who had reason to call the police were asked how rapidly they responded and how well they did their job.⁵⁸ One component of the neighborhood revitalization experiment in Hartford was to be a neighborhood policing team, so residents there were quizzed extensively about police activity. (As is sometimes the case, the program never really was implemented.) In Cincinnati a team policing experiment was evaluated by interviewing a panel of businessmen about police service.⁵⁹

A few evaluations have focused on services other than policing, and on community or housing-project activities. The evaluation of crime prevention activities in a Chicago housing project included questions about resident satisfaction with 17 specific kinds of services, including the Resident Safety Aide program, building maintenance, youth services, and elevators.⁶⁰ In Seattle, evaluators asked samples of respondents in target and control areas if they had used any of the services of the Community Crime Prevention Program. Those who had were asked to rate the "helpfulness" of each of several specific activities, and if there were any forms of assistance which they did not feel they had received.⁶¹

Focusing on citizens as consumers of a program's product fosters the development of downstream rather than upstream measures of program success. As such, it reflects concern about the responsiveness of (especially public) organizations to citizens' needs, and the necessity of developing channels for the systematic feedback "upstairs" of information about the effectiveness of programs developed at the grass roots. This concern is reflected in the increasing use of a new tool for the general evaluation of municipal services, the omnibus service delivery survey. From St. Petersburg to Eugene, communities are sponsoring periodic surveys of their residents to inquire about their needs and probe their assessments of the quality of municipal services. These surveys usually include questions concerning the criminal justice system, and there is no reason why the scope of those questions could not be expanded on occasion to cover community crime prevention activities as well.

A service-delivery focus for evaluating community crime prevention programs also reflects trends in the operation of community organizations. One impact of extensive Federal and State funding of community organizations engaged in anticrime efforts has been to bureaucratize those groups and transform them from "bottom up" movements to "top down" vendors of services. Groups now get money to hire staff to sell people whistles, mark property, and organize

patrols in their "catchment area." Crime prevention is a moderate-sized industry, but operating as it does in a nonmarket grant economy, those who wish to keep it responsible to its neighborhood constituencies will have to support the development of new ways to feed citizens' perspectives on the progress of the War Against Crime back into the system.

D. Reporting to the Police

Many crime prevention programs share the goal of increasing the rates at which victims report their experiences to the police, neighborhood residents call the police quickly to deal with suspicious persons, and witnesses step forward to assist the police with their investigations. This is seen as desirable for a number of reasons, including that it is good citizenly conduct. Further, it is presumed that rapid crime reporting and intervention by bystanders to summon the police will increase the rate at which offenders are apprehended, and that citizen cooperation facilitates the conviction of those who do get caught. As a spinoff, increased crime reporting would improve the quality of police records on crime (see below). This would enable police to plan and evaluate their activities more effectively.

Several evaluation surveys have focused on the effects of programs on citizens' reporting of crime, including those of policing experiments in Kansas City and Cincinnati. In the Community Crime Prevention area, evaluators in Seattle found that (a) there were no differences between reporting rates in experimental and control tracts before their intervention, (b) reporting rates rose substantially in the experimental tracts after the intervention (from 51 percent to 76 percent), and (c) reporting rates did not change significantly in their control areas.⁶²

In Portland, Schneider examined the impact of program participation on burglary reporting rates. She found in a correlational study that participants in meetings and those who marked their property and mounted an ID sticker on their door were more likely to report victimizations to the police, and the effect of these activities was additive and cumulative. Those who were victimized but knew nothing about those programs reported only 65 percent of all burglaries, while those who had been involved in all three activities reported 87 percent of them. Thus, although their burglary rates were lower, participants generated more reports for the police than did nonparticipants.⁶³

These differences are impressive, for analysis of the Census Bureau's victimization survey data indicates that the primary determinant of whether burglaries are reported is the amount stolen and whether or not the household was insured for the loss. However, the incremental effect of program participation does not seem to be marginal, but leads to substantial increases in reporting rates. In fact, in Portland, the impact of that community's antiburglary

program was so dramatic that it apparently increased the burglary rate during the postintervention period. However, Schneider demonstrated that the burglary rate actually went down when she instituted in her study controls for shifts in the reporting rate.

Fowler⁶⁴ observed the same phenomenon in Hartford. During the 5-year evaluation effort there, the official crime rate in the experimental area was seen to go both up and down as a result of shifts in reporting rates: both shifts were the result of police politics and policy. Only because the Hartford Institute was continually monitoring test and control sites via sample surveys could they detect these changes, and appropriately discount shifts in official crime counts.

As this suggests, it is difficult to envision using official crime reports to evaluate many community crime prevention programs without gathering parallel information on the crime reporting practices of the community. The problem is that good measures of the reporting rate depend on the same methodology and procedures which produce good estimates of the victimization rate. This means, among other things, that the samples involved must be very large. In the Seattle evaluation it was impossible to compare reporting rates for participating and nonparticipating households, ". . . due to the small number of burglary cases. All of the six burglaries to CCPP residences were reported, however." It is for this reason that Schneider confined her analysis to burglary, although she collected data on a number of crimes. Burglary was the only major crime frequent enough to be measured well in a survey with "only" 3,900 respondents.

Further, we also have no inkling of the validity of survey measures of citizens' crime reporting. There is considerable evidence that a "bandwagon" effect exists for socially desirable responses to questions about citizens' behavior,⁶⁵ and crime reporting may be among that set. In her "forward record check" Schneider⁶⁶ took reports of crimes which were said to have been reported to the Portland police, but she could only find 50 percent of them in department files. It is uncertain what proportion should be written off to police nonrecording and what percentage can be attributed to invalid responses to the original survey question.

Summary

As we have seen in this section, data gathered through interviews are very attractive to evaluators. They are perhaps the only practical way to gauge the impact or the implementation of programs among mass publics. They give us an alternative perspective to what officials say they did. They can probe the past, gathering data on events or conditions which were obtained even before evaluators came on the scene. Surveys can employ respondents as observers by asking them to recount things they have seen or experienced;

they can also be used as evaluators to make judgments about those events. However, interview data typically are replete with error. Few community crime prevention evaluations have focused on the technical characteristics of their data in this regard. People do not always reliably report things they have seen: they forget, misremember, lie, and strive to meet the (often misinterpreted) "needs" of the interviewer. On the other hand, evaluators also, too often, place impossible burdens on their respondents, or quiz them on things about which they have no information or opinion. There are specific techniques for estimating the reliability and validity of measures, and knowing these characteristics of one's data should play an important role in its analysis. Too often it does not.

ARCHIVAL MEASURES

Introduction

Archival measures are traces of programs or their consequences which are found in existing banks of records, files, reports, or accounts. Because these are kept almost exclusively by organizations and government agencies, data from these sources inevitably reflect upstream rather than downstream perspectives. Further, because the information is drawn from things which administrators thought important to commit to paper (and now to computers) the data have a strong tendency to reflect what they felt to be significant, and to exclude information which was uninteresting to them, difficult to gather, spoke poorly about their efforts, or was potentially embarrassing and thus prone to getting "lost." Good examples of these difficulties would be encountered if one were to follow Yin's⁶⁷ suggestion to make use of agencies' records of complaints against themselves.

As a result, the fundamental problem with most archival measures is one of validity--do indicators reflect "what the evaluator really wants to know?" More often they reflect what is available. The most frequently cited rationale for the use of archival data is that it is "unobtrusive," or that its collection flows naturally from an organization's activities and does not intrude "unnaturally" in the process being evaluated. As this criticism suggests, however, intervention and outcome data can be "obtrusive" at the time it comes in the door.

Measures of Interventions

Because intervention indicators usually employ data collected by and for project administrators, indicators of implementation collected from archival holdings typically reflect upstream intervention processes. They detail how project personnel and resources were allocated, how people spent their time, and the extent of

their "outreach" into the community. As we have seen above, these measures may not square with those generated from the standpoint of the targets of those efforts, but archival measures like these are employed frequently because they are available, cheap, and reflect the interests and internal accounting needs of sponsoring organizations, and thus, almost by definition, are of concern to them. Under those circumstances the validity of the measures seems to take on secondary importance.

The simplest archival measures of implementation are activity counts. These describe what an organization thinks it did with its personnel and resources that could be counted as "services rendered." For example, an evaluation of the Miami Robbery Control Project noted the number of witnesses and victims interviewed, the number of publications distributed, the number of stakeouts and field interrogations carried out, and how much the project cost the police department. As is often the case, the Federal money which sparked the program also paid for the first systematic collection of most of this information, so there also were no "before" measures of those efforts in an archive.⁶⁸

More sophisticated activity counts are used to account for the attainment of specified project goals. The Oregon Law Enforcement Council thus reported that the Portland Crime Prevention Bureau held 50 percent of its targeted number of meetings and marked property in only 27 percent of the desired number of households. The evaluative potential of these sorts of measures was reflected in the conclusion to the report: "Either the number of visits performed will have to be increased or the objective will have to be revised and made more realistic."⁶⁹

Agency "outreach" measures estimate the number of persons contacted or assisted by a program. When they are combined with projections of the number of qualified recipients of the program, those counts may be converted by long division into measures of "program penetration." The central staff of the Minnesota Crime Watch program gathered figures from police and sheriffs departments on the number of persons who had enrolled locally in an Operation Identification program. They then estimated the number of households located in jurisdictions which were served by agencies that made this participation possible by cooperating in their program. This constituted their universe of "targets to be served," of whom they enrolled about 9 percent. In addition, they then recalculated their program participation figure as a proportion of the rate achieved by the Operation Identification program in Monterey Park, California, which they argued is the "benchmark" American program. This placed them at 38 percent of "success" as defined by the Monterey Park project.⁷⁰

The unknown quantity in all of these measures is what they "mean," which is a validity question. They are, at best, descriptions of the operation of programs, and their analytic utility for

any but the most mundane accounting purposes remains problematic. Further, they may not even be accurate indicators of the operations they purport to describe. Because they largely are "management" data describing the activities of subordinates, there doubtless is a tendency for them to overstate levels of effort and effectiveness. People who are held accountable at various levels within organizations, as well as organizations as a whole, usually attempt to present their activities in the best possible light, and evaluators--like top managers--usually find it difficult to ferret out what is "really going on."

Outcome Measures: Crime Rates

We have indicated that the most commonly employed set of measures for judging the effectiveness of local community crime prevention programs remains official police reports about crime. The designs in which they are employed vary tremendously in sophistication, ranging from "before-and-after" to "multiple interrupted time-series with control groups." Those designs also vary in the detail of the data analyzed, which ranges from simple Uniform Crime Report (UCR) category frequency counts to detailed analyses of the characteristics of individual offenses reports. However, in the end, all use police records as a source of information to talk about crime and what happened to it, which in turn depends on some quite unsupportable assumptions about the quality of the data.

At a highly aggregated level, official police reports are frequently employed to evaluate crime reduction programs using "before and after with control areas" designs. For example, the Harrisburg Street Lighting Program was evaluated by calculating percentage changes for Part I crimes from the pre- to post-intervention period in "control," "experimental," and "potential displacement" areas.⁷¹ In Seattle, Mathews⁷² examined monthly percentage changes in burglary for target areas and the remainder of the city, before and after (or more accurately, during) their program.

As more sophisticated statistical analysis of the frequency of crime was reported by Lavrakas et al.,⁷³ for a Portland commercial tract. They analyzed fluctuations in a 36-month series of burglary and robbery counts, examining interruptions in the trend caused by a commercial security survey program and a high-intensity street lighting project along the avenue. They made impressionistic use of residential crime counts in the area as "controls," for they should not have been affected by the program.

Arthur Young & Co.⁷⁴ made extensive use of more detailed breakdowns of reported crime in an evaluation of a crime-reduction experiment in a Chicago housing project. In addition to comparing crime counts over a 3-year period in experimental and control buildings, the company also analyzed them by type and location. For example, contrasts were drawn between trends for crimes within

apartments (which should not have been affected by the interventions) with those in public areas, and those which took place inside of the buildings (which should have declined) with those from outside. They also could have examined such indicators as trends for stranger and nonstranger violence, and arrests of offenders who lived outside as opposed to inside the buildings (in each case only the former should have been affected).

Criticisms of Crime Rates

General criticisms of the official crime reports used in these evaluations abound, and will not be repeated here.^{75, 76} A very insightful inventory of things for which they are often useful can be found in Inciardi.⁷⁷ Deficiencies in the data can be attributed to two major sources: citizen crime reporting and police crime recording. At the reporting end, it is apparent that the majority of crimes are never brought to the attention of the police. In major crime categories (robbery, rape, burglary, and the like) victims appear to report about 50 percent of their experiences, but that figure drops to as low as 25 or 30 percent for petty thefts. Thus, there is a large pool of events which never come to the attention of the criminal justice system.⁷⁸ The police deal with citizens' reports in a variety of ways. At various times police have been charged with underrecording, overrecording, misrecording, and nonrecording crimes. Although victimization surveys reveal a great deal about victims' reporting habits, we know much less about how the police generally deal with those reports. A comparison of what people say they told the police about and official crime counts for 26 cities indicates that the authorities properly record about 60 percent of what is brought to their attention;⁷⁹ this is roughly of the magnitude of Schneider's ability to recover police reports of crimes which were described in her victimization survey as reported to the Portland police.⁸⁰

For victims it appears that the major determinant of reporting to the police is the seriousness of the incident. The police are most likely to be notified about crimes involving injuries, large financial losses, or the use of a weapon. The pool of reported crime is more serious than the universe it "represents." In addition, burglary and larceny victims whose losses were insured are much more likely to report them to the police. In the National Crime Survey data for 1973 the correlation between having insurance and reporting these two crimes to the police was +.57 and +.56, respectively. It also is widely suspected that victims are much less likely to report crimes that involve relatives, friends, and other "nonstrangers" as the perpetrators. Finally, reporting rates are very low for youthful victims, who suffer most frequently from assaultive violence.⁸¹

The recording of crime may vary enormously among police departments, depending on organizational, leadership, and political

factors. At the level of the individual officer, observational studies of the decision to "write up" an offense indicate that the seriousness of the complaint, the preferences of the victim, the relational distance between victim and offender, the victim's deference to the officer, and the victim's social class, all affect that action.⁸² In the aggregate, more innovative and resourceful departments seem to record a greater proportion of the crimes brought to their attention.⁸³ As the Crime Commission and others have noted, however, crime recording practices vary considerably within cities and are subject to rapid fluctuations as political and administrative factors change.

It is precisely change, or instability in citizen reporting and police recording practices generally, which constitutes the principal threat to the reliability of official measures of crime. In measurement terms these result in "changes in instrumentation" which can become confounded with changes in the variables under scrutiny. As the evaluators overseeing Seattle's Community Crime Prevention Program noted:

On the surface it would appear to be simple to assess the impact of the Community Crime Prevention Program upon burglary. A researcher would simply need to check police records of burglaries before and after the program delivered services to an area and see if the number of burglaries was reduced. This approach is severely flawed, however, because in addition to reducing burglaries the program has a second goal of increasing citizen reporting of burglaries to the police. . . . Program success in increasing citizen reporting of burglaries would mask the crime reduction impact and might even produce an increase rather than decrease in burglary reports in neighborhoods receiving the services of the CCPP.⁸⁴

As we have seen above, the problem of change in instrumentation is reflected in Schneider's⁸⁵ findings about the relationship between program participation and burglary reporting in Portland, and in the (presumed) reason why the inauguration of the Philadelphia Transit Crime Program signaled an increase in the official crime count there.⁸⁶

In each case, official crime reports measured things differently after the intervention, and did not merely reflect changes in crime rates.

It has been recognized for a long time that there is a substantial "dark figure" of unrecorded crime. The serious blow, however, is the apparently unstable relationship between reported crime and the "true" crime rate. "Crackdowns" and other special enforcement programs may change the proportion of reported events which surface as official reports, whereas programs and longer-term factors like the spread of property-crime insurance may affect

citizen reporting practices. Booth and others⁸⁷ have shown that official crime statistics and the results of victimization surveys can lead evaluators to quite different conclusions.

Outcome Measures: Consequences of Crime

Unlike survey research, reliance on archival data to measure the consequences of crime-reduction programs does not seem to have encouraged the use of diverse indicators of outcomes. Although surveys have focused on fear, behavior, and satisfaction with services as frequently as they have personal victimization, only a few archival ventures have strayed beyond the analysis of crime data.

Several of those evaluations have been studies of programs in public housing projects in which vacancy rates have been used as indicators of residential satisfaction. In their analysis of activities in the Cabrini-Green complex in Chicago, Arthur Young & Co.⁸⁸ used "move-ins" and "move-outs" to chart the progress of their experimental and control buildings. Oscar Newman⁸⁹ employed two archival measures, apartment turnover rates and vacancy rates, in a composite index of the instability of each of 63 project sites (the two were correlated +.68, and about +.50 with the third component of the index, the proportion of residents who said they wanted to move out). He found that residential instability was higher in highrises, in buildings next to bad buildings, and in buildings with high levels of crime, vandalism, and fear.

In a Minneapolis study, the Minnesota Governor's Commission on Crime⁹⁰ examined statistically the impact of crime on property values, and the consequences this reduction in value had in turn for property tax assessments and city revenues. They estimated that increases of 1 percent in residences burglarized in a year lowered residential property values in the vicinity by \$553, whereas every reported act of vandalism per 1,000 residents reduced property values by \$172. Across the entire city, these decrements reduced property tax collections by about \$17,000,000.

Finally, Lavrakas et al.⁹¹ examined the impact of the Portland commercial CPTED demonstration on the vitality of business in the area. They analyzed new business openings and the rate at which old businesses on the avenue closed down, the average "life-time" of businesses in the area, and their gross receipts from sales. All of these data were gleaned from license applications and sales tax records on file at the Portland City Business License Division.

As this suggests, municipalities can be treasure troves of information which trace the impact of crime in the community. On file in various offices can be found data on real estate assessments, bus fare collections, telephone connects and disconnects,

school vandalism costs, street sign thefts, residential sales, "going out of business" sale permits, building abandonments, arson complaints and reports, and building permits. All of these can be used as indicators of the consequences of crime and can be employed to track the impact of intervention on the quality of urban life.

Summary

Archival measures often are employed because they are available and inexpensive to collect. Sometimes they validly reflect what evaluators really want to know about interventions and outcomes. More often they reflect the management and political needs of cooperating agencies, who were already interested enough in the data to keep it on file. Except for official counts of crime, there has been little use of archival data beyond those of an administrative nature. Crime data are themselves so plagued with contingencies in citizen reporting and police recording practices that they are difficult to interpret. However, as indirect indicators of the consequences of crime, archival measures have some potential, especially when multiple measures of outcomes are employed.

CONCLUSION

In this chapter I have reviewed many recent evaluation efforts illustrating measurement problems involved in making accurate and credible conclusions about the impact of community crime prevention programs. As we have seen, evidence regarding the reliability and validity of measures employed in those evaluations was sparse, and what we know about the difficulties of measurement in social research and evaluation should not lead us to be optimistic about the accuracy or credibility of those conclusions. This is not, however, cause for despair. In fact, because so many evaluations of programs in criminal justice fail to find a posited effect or are able to discern them only at the margins, and the fact that those evaluations often are founded on an inadequate measurement base should be seen as a positive sign. It is useful to think of an evaluation as a contest between the effects of a program and noise: programs appear to be "winners" only when they can shout the opposition. More systematic appraisal of our measures would enable us to reduce substantially the volume of this noisy chorus, increasing the apparent impact of our efforts. From a purely technical standpoint, it is possible that efforts to improve measurement procedures could alone have a great impact on the policy-making process.

Efforts to increase the reliability of our measures would substantially increase the statistical relationship between criminal justice interventions and most of the outcome measures reviewed

here. This payoff is a result of the fact that the maximum possible correlation between two measures is determined by their reliabilities. If an indicator is an unreliable reflection of a variable, its correlation with another measure would fall even if the "true" correlation were perfect. (More precisely, r_{\max} equals the square root of the product of the reliabilities of the two measures involved.) Thus, the more noise there is in a measure the stronger the "true" effect required to reflect it in the data. As a result of unreliability we probably are rejecting hypothesized effects of community crime prevention programs more often than we should.

Procedures for estimating the reliabilities of measures, and thus for assessing improvements in them, are relatively straightforward, and many of them have been discussed above. They all involve some measure of the agreement between similar measures of a concept. Again, this reflects Eddington's observation that scientists read the positions of pointers, and the corollary requirement that the pointer swing twice to the same position when tested. In the case of observational studies we are concerned that two or more observers can agree on the reading for an event or a condition, and about the ability of observers to make consistent judgments over time. In survey research the analogous standards involve the internal consistency of similar measures gathered at the same time (do they point to the same thing?), and the stability of measures repeated over time. In archival projects we must insure that data-gatherers follow the same procedures.

The rigorous pursuit of standards to assure reliable measurement should be coupled with thorough documentation of the results in evaluation reports. This documentation should play an important role in establishing the credibility of the conclusions they reach, as well as providing a useful diagnostic tool for those conducting "next generation" evaluations based on those findings.

The credibility of the conclusions of an evaluation also are bound up with the validity of the measures it employs. We need to demonstrate the validity of our measures to strengthen the argument that the evaluation was indeed aimed at the point of the program. Thus, checks of the validity of our measures aim at increasing the fit between those measures and the concepts they represent. Like reliability checks (which are aimed at increasing the fit between similar measures), validation involves the use of multiple measures. In this case, however, they are different (rather than maximally similar) measures of the concept. To the extent to which these indicators point to the same position on different scales we know that our evaluation findings are robust and generalizable.

We have seen several examples of the validation of evaluation measures above. In the case of fear-of-crime measures we know that people who live in high-crime areas are more likely than others to score in the high range, and that fearful responses about

crime are independent of (do not point in the same direction as) other expressed fears. As a result of methodological investigations in the area of victimization we know that the results of personal interviews about their experiences with victims of many kinds of crime correspond closely with police records, but that for others the two do not agree. In Hartford, the results of mechanical vehicle counters, traffic observers, and interviews with residents of the target community all pointed to the success of an effort to reduce congestion there.

In each case, evaluation findings were more credible because we had some assurance that the conclusions were "method free," and not the artifact of some particular way of looking at a program. In general, when different observers representing different viewpoints agree on something, their judgment becomes more difficult to dismiss as the result of some "bias." The development of valid measurement techniques simply is a rigorous approach to this standard.

NOTES

1 Leroy Bailey, Thomas F. Moore, and Barbara Bailar, "An Interviewer Variance Study For The Eight High Impact Cities of The National Crime Surveys Cities Sample," Journal of the American Statistical Association, 73:16-23 (March 1978).

2 Police and Security Management Consultants, "Public Transit Crime Reduction Program--Philadelphia Police Department" (Syracuse, N.Y.: Career Co., Inc., 1975).

3 Minnesota Governor's Commission on Crime Prevention and Control, Minnesota Crime Watch (St. Paul: 1976).

4 Anne L. Schneider, "Victimization Surveys and Criminal Justice System Evaluation" in Wesley G. Skogan, ed., Sample Surveys of the Victims of Crime (Cambridge, Mass., Ballinger, 1976), pp. 138-150.

5 John Richardson and Peter Hartgens, "Methodological Problems Facing the Crime Prevention Evaluator," (Washington: National Conference on Criminal Justice Evaluation, 1977).

6 Floyd J. Fowler, Jr., Mary Ellen McCalla, and Thomas Mangione, "Reducing Crime and Fear in an Urban Residential Area: The Planning and Evaluation of an Integrated Approach to Opportunity Reduction" (Boston: Center for Survey Research, University of Massachusetts-Boston, 1978).

7 James Q. Wilson and Barbara Boland, "Crime," in William Gorham and Nathan Glazer, ed., The Urban Predicament (Washington,: The Urban Institute, 1976), pp. 179-230.

8 Anne L. Schneider, "Evaluation of The Portland Neighborhood Based Anti-Burglary Program" (Eugene, Oreg: The Oregon Research Institute, 1975).

9 Anne L. Schneider and Paul Reiter, "Citizen Perceptions of Street Lighting" (Eugene, Oreg: The Oregon Research Institute, 1975).

10 Portland Urban Studies Center, "Robbery and Burglary Victimology Project" (Portland, Oreg: Portland State University, 1972).

11 Thomas A. Reppetto, Residential Crime (Cambridge, Mass.: Ballinger, 1974).

12 Paul J. Lavrakas, Janice Normoyle, and J.J. Wagener, "CPTED Commercial Demonstration Evaluation Report" (Evanston, Ill., Westinghouse Evaluation Institute, 1978).

13 Fowler et al., cited note 6.

14 Richard Taub, "Crime, Fear of Crime, and The Deterioration of Urban Neighborhoods" (Chicago: National Opinion Research Center, University of Chicago, 1978).

15 Oscar Newman, "Crime and Instability in Federally-Assisted Housing" (New York: Institute for Community Design Analysis, 1977).

16 Robert Yin, Patrolling The Neighborhood Beat: Residents and Residential Security (Washington: National Technical Information Service PB-262-333, 1976).

17 Albert J. Reiss, Jr., and Donald J. Black, "Pattern of Behavior in Police and Citizen Transactions," in Studies of Crime and Law Enforcement in Major Metropolitan Areas: A Research Report to the President's Commission on Law Enforcement and Administration of Justice. (Washington, United States Government Printing Office, 1967).

18 Thomas R. Decampoli, "Wilmington Crime Specific Program: A Project Evaluation," (Wilmington, Del.: Delaware Agency to Reduce Crime, 1975).

19 Kenneth E. Mathews, "Evaluation of First-Year Results of Community Crime Prevention--Burglary Reduction" (Seattle: Law and Justice Planning Office, City of Seattle, 1974).

- 20 John Richardson, "The Eden Prairie Premise Survey Evaluation" (Washington: National Conference on Criminal Justice Evaluation, 1977).
- 21 Schneider, "Evaluation. . .," note 8.
- 22 Schneider and Reiter, note 9.
- 23 Arthur Young & Co., Second-Year Evaluation Report For the Cabrini-Green High Impact Program (Chicago: City Department of Planning, City, and Community Development, 1978).
- 24 Hugh J. Parry and Helen M. Crossley, "Validity of Responses to Survey Questions," Public Opinion Quarterly 14, 1: 61-80.
- 25 Phillip Ennis, Criminal Victimization in the United States (Washington: United States Government Printing Office, 1967).
- 26 John D. Carr, "Operation Identification Interim Evaluation Report" (Washington: National Conference on Criminal Justice Evaluation, 1977).
- 27 Minnesota Governor's Commission, cited note 3.
- 28 Ibid.
- 29 Albert D. Biderman and Albert J. Reiss, Jr., "On Exploring the Dark Figure of Crime," Annals of the American Academy of Political and Social Science 374 (November, 1967):1-15.
- 30 Terry Baumer, Urban Crime and Precautionary Behavior, (diss., Chicago: Loyola University, 1979).
- 31 Paul J. Lavrakas, "Citizen Participation and Community Crime Prevention," (Evanston, Ill: Center for Urban Affairs, Northwestern University, 1978).
- 32 Schneider, "Evaluation. . .," note 8.
- 33 Minnesota Governor's Commission, note 3.
- 34 Elinor Ostrom, "Multi-Mode Approaches to Measuring Government Productivity," in R. Gage and R. Sloan, ed., Delivery of Urban Services (Boulder, Colo., Bureau of Governmental Research and Services, University of Colorado, 1976).
- 35 Schneider and Reiter, note 9.
- 36 Paul Cirel et al., Community Crime Prevention Program--Seattle, Washington (Washington: United States Government Printing Office).
- 37 Schneider, "Evaluation. . .," note 8.
- 38 Fowler et al., note 6.
- 39 George Kelling et al., The Kansas City Preventive Patrol Experiment (Washington: The Police Foundation, 1974).
- 40 Arthur Young & Co., cited note 23.
- 41 Wesley G. Skogan, Victimization Surveys and Criminal Justice Planning (Washington: United States Government Printing Office, 1978).
- 42 Richard F. Sparks, Hazel G. Genn, and David J. Dodd, Surveying Victims: A Study of the Measurement of Criminal Victimization (New York: John Wiley, 1977).
- 43 Michael Hindelang, Criminal Victimization in Eight American Cities (Cambridge, Mass., Ballinger, 1976).
- 44 Carol Kalish, Crimes and Victims: A Report on the Dayton-San Jose Pilot Survey of Victimization (Washington: Law Enforcement Assistance Administration, 1974).
- 45 Sparks et al., cited note 42.
- 46 Bailey et al., cited noted 1.
- 47 Cirel et al., cited note 36.
- 48 Fowler et al., cited note 6.
- 49 Victoria Jaycocks, "The Elderly's Fear of Crime: Rational or Irrational," Victimology 3 (Winter, 1979), forthcoming.
- 50 Kelling et al., cited note 39.

- 51 Arthur Young & Co., note 23.
- 52 Kelling et al., cited note 39.
- 53 Alfred I. Schwartz et al., Evaluation of Cincinnati's Community Sector Team Policing Program: A Progress Report (Washington: The Urban Institute, 1975).
- 54 Schneider and Reiter, note 9.
- 55 Marlys McPherson, "Realities and Perceptions of Crime at the Neighborhood Level," Victimology 3 (Winter, 1979), forthcoming.
- 56 Wesley G. Skogan, "The Independence of the Fear of Crime from Other Concerns." Appendix to testimony presented before Joint hearings of the Select Committee on Aging and the Committee on Science and Technology, United States House of Representatives, (95th Congress, 2d Session), Feb. 1, 1978.
- 57 Wesley G. Skogan, The Center for Urban Affairs Random Digit Dialing Telephone Survey (Evanston, Ill., Center for Urban Affairs, Northwestern University, 1978).
- 58 Kelling et al., note 39.
- 59 Schwartz et al., note 53.
- 60 Arthur Young & Co., note 23.
- 61 Cirel, note 36.
- 62 Ibid.
- 63 Schneider, "Victimization . . .," note 8.
- 64 Floyd J. Fowler, Jr. "The Evaluation of the Hartford Experiment: A Rigorous, Multi-Method Effort to Learn Something" (Washington: Second National Workshop on Criminal Justice Evaluation, 1978).
- 65 Anne L. Schneider, "The Portland Forward Records Check of Crime Victims" (Eugene, Oreg., Institute of Policy Analysis, 1977).
- 67 Yin, cited note 16.
- 68 Paul M. Whisenand and Judith A. Hails, "Exemplary Project Validation: Miami Police Department Robbery Control Project" (Washington: Technology Transfer Division, National Institute of Law Enforcement and Administration, 1973).
- 69 Decampii, cited note 18.
- 70 Minnesota Governor's Commission, note 3.
- 71 Harrisburg Police Department, "Final Evaluation Report of the High Intensity Street Lighting Program" (Harrisburg, Pa.: Planning and Research Division, Harrisburg Police Department, 1976).
- 72 Mathews, cited note 19.
- 73 Lavrakas et al., "CPTED . . ." cited note 12.
- 74 Arthur Young & Co., note 23.
- 75 Marvin E. Wolfgang, "Uniform Crime Reports: A Critical Appraisal," University of Pennsylvania Law Review, 30 (April, 1963):708-738.
- 76 Sophia M. Robinson, "A Critical View of the Uniform Crime Report," Michigan Law Review 64 (April, 1966):1031-1054.
- 77 James A. Inciardi, "The Uniform Crime Reports: Some Considerations on Their Short Comings and Utility," Review of Public Data Use 6 (November, 1978): 3-16.
- 78 Wesley G. Skogan, "Crime and Crime Rates," in Wesley G. Skogan (ed.) Sample Surveys of the Victims of Crime (Cambridge, Mass.: Ballinger, 1976.)
- 79 Ibid.
- 80 Schneider, "Portland Forward Records Check . . .," note 66.
- 81 Wesley G. Skogan, "Citizen Reporting of Crime: Some National Panel Data," Criminology 13 (February, 1976):535-549.

- 82 Donald J. Black, "Production of Crime Rates," American Sociological Review
35 (August, 1970):733-747.
- 83 Skogan, "Citizen Reporting," note 81.
- 84 Cirel et al., note 36, p. 47.
- 85 Schneider, "Evaluation . . .," note 8.
- 86 Police and Security Management Consultants, note 2.
- 87 Allen Booth, David R. Johnson, and Harvey M. Choldin, "Correlates of City
Crime Rates: Victimization Surveys versus official Statistics," Social Problems
25 (1977):187-197.
- 88 Arthur Young & Co., note 23.
- 89 Oscar Newman, "Crime and Instability in Federally-Assisted Housing" (New
York: The Institute for Community Design Analysis, Vol. II, 1978).
- 90 Minnesota Governor's Commission, note 3.
- 91 Lavrakas, et al., note 12.

The Impact of Innovation

Assessing the Evaluation of Courts

Malcolm M. Feeley is a member of the political science department at the University of Wisconsin/Madison and formerly of the Yale University Law School.

INTRODUCTION

During the past few years American courts have undergone a major transformation. Increased crime rates and the expanded right to counsel have forced the courts to grow in both size and function. During this period the combined dollar amount allocated to the courts by local, State and Federal governments has also grown, increasing almost fivefold, and keeping abreast if not outpacing crime rates and inflation.¹

This increase in resources has meant not only an expansion in size but also the adoption of a host of new and different programs. Despite recurring complaints that "courts never change," they are constantly changing and are in fact considerably different now than they were 25 years ago. By expanding the right to counsel, first in Gideon v. Wainwright² and later in Argersinger v. Hamlin,³ the Supreme Court precipitated major changes in the operations of many of the nation's criminal courts, dramatically increasing the size of public defense systems and altering the way courts handle cases.

The bail reform movement enjoyed a resurgence in the late 1950's and 1960's, spawning a host of new legislation and new types of pretrial release programs which by now have firmly taken hold in many jurisdictions. This in turn has increased the concern with preventive detention, which has already been translated into formal policy in the District of Columbia.

Increased concern over the use of narcotics has led to a host of drug treatment programs as alternatives to prison. Problems with the masses of petty cases that clog the courts have given rise to the idea of formalized pretrial diversion, endorsed by the President's Crime Commission in 1967 and subsequently fostered by the Department of Labor and LEAA.

Other more recent responses to caseload pressures include creation of informal dispute settlements centers, expansion of the duties of U.S. magistrates in the Federal courts, and increased reliance on parajudicial officers.

Copyright (c) 1979 Malcolm M. Feeley

At the adjudication stage there have been a number of important changes within recent years. Despite continuing criticism, plea bargaining has been held constitutional and is gaining support as a legitimate means for disposing of criminal cases expeditiously. Increasingly, concern is shifting from ways to eliminate plea bargaining to ways to reduce, structure, and control it. A number of experiments now underway are pursuing these possibilities.

The widespread use of defense attorneys is itself a new and novel idea of unprecedented dimensions in modern industrialized societies, and so too are the variety of ways public defense services are provided. The last few years have also seen an increase in concern for victims and witnesses, and a variety of programs have sprung up to provide these people with increased opportunity to be kept abreast of and participate in the criminal process.

Prosecutors' offices have not only expanded dramatically during recent years, they have also become better managed. Most of the larger and many of the smaller offices have phased out part-time staff and, in its place, engaged full-time, career personnel. These offices are better staffed and better organized than ever before. This means better training, greater specialization, more effective support services, and more and better information.

Prosecutors have innovated in other ways as well. Early case screening and assessment by skilled prosecutors have been promoted in the hope that they will save time and effort at later stages of the criminal process. "Career criminal units"--special prosecutorial sections designed to focus exclusively on serious cases against allegedly dangerous offenders--have been implemented in many jurisdictions in an effort to move quickly and confidently in especially serious cases.

The courts themselves have experimented with a variety of administrative innovations. Chief judges are increasingly assuming administrative duties, overseeing the work of their colleagues. Full-time court administrators have now become permanent officials of many courthouses. Calendaring practices, pretrial conferences, motions practices, and court-reporting practices are all subjects of increased concern and experimentation. Speedy trial rules have been adopted in many jurisdictions in an effort to expedite the handling of criminal cases.

Within recent years sentencing has reemerged as a major issue of public policy. Some believe that criminal sentences are too harsh, others contend sentences are too lenient, and still others are concerned with the problem of disparity. This debate has sparked a lively literature on the philosophy of sentencing, which in turn has begun to influence sentencing policies. New York, Massachusetts, California, Oregon, and Maine have already under-

taken major sentence revisions, and many more States are considering them. The United States Congress is also in the midst of considering major revisions in the U.S. Criminal Code and, if made, these changes will significantly alter the nature of sentencing in the Federal courts and have an important secondary effect in the States.

Have all these changes made any difference? Clearly the answer is yes. Courts look different and do things differently today than 10, 15, or 25 years ago. Have, on the whole, these changes improved the quality of justice? Overall the answer is probably yes, but the elusive nature of justice and the complexities of the criminal process do not permit easy and sweeping answers. While there is clearly less corruption, less flagrant abuse of the criminally accused, and greater visibility and concern in the criminal process than there was just a few years ago, expectations and standards have also risen. We expect more of courts today than we once did.

Have individual changes accomplished their limited goals? Here the answer is even more tentative and mixed. Some have, others have not, and many have precipitated unanticipated and undesirable consequences. In most instances, however, we must throw up our hands and admit that we do not know. From a broad perspective, it is clear that there have been dramatic changes in the criminal courts, but once we begin to trace the specific influences of individual programs we must admit near-failure.

It is extremely difficult to determine if and how and why a program has had an effect on a complex system and it is even more difficult to assert with any confidence whether or not any individual program has accomplished its goals in a cost-effective manner. While this problem is in part an inevitable consequence of the complexity of the criminal courts, it is exacerbated by the paucity of careful evaluation research in the area. Careful assessments are time-consuming, expensive, and intrusive. Court officials, strapped for resources, and traditionally isolated from scrutiny and the need to justify their practices, are often unable and unwilling to devote their limited funds or yield their autonomy for research. As a result few of the many innovations cataloged above have been subject to careful scientific scrutiny. Despite their often being labeled experiments, despite the benefits claimed by their proponents, and despite the source of funding (often evaluation-conscious LEAA), few of them have been subject to careful evaluation.

The paucity of systematic evaluation during this period of increased "experimentation" itself speaks volumes about the prob-

lems of the courts. Court systems are isolated, fragmented, and lack accountability and central control. They possess a limited feedback capacity and little incentive to create one. Even when evaluations are undertaken, they usually focus on only selected aspects of changes and do not come to grips with the system-wide effects of the effort.

But to say that there has been little first-rate evaluation of court innovations does not mean that there have been no useful assessments. There is a substantial body of material which has documented, described, and assessed the impact of new programs. Much of it is descriptive, little more than project monitoring. While the quality of this research is often low, the situation is not entirely disillusioning. If regarded as "definitive," a single evaluation might take on undue significance and be used to promote prematurely or terminate a fledgling idea in need of further nurturing. And much can be learned from careful descriptions and partial and tentative findings of flawed and imperfectly designed studies. The cumulative impact of several such studies may eventually yield greater insight and confidence than the results of one or a few well-executed studies. In particular, studies of process which emphasize context and personality may eventually yield greater insights about the program consequences and impact than technically well-executed evaluations which often dwell on quantitative indicators at the expense of other "softer" factors. As such, quantification can and often does yield a false precision, and methodologically "weak" descriptive studies may be of greater value than they are usually credited with being.

In this review, I have construed the notion "evaluation" quite broadly and considered studies which only in the broadest sense can be termed "evaluations." This expands what might otherwise have been a review of only a small handful of studies or a long recitation of methodological weaknesses in a library of studies claiming to be evaluations. But in construing my task this broadly, I was in danger of being swamped with materials. As a consequence, I have been quite arbitrary in what I have chosen to discuss. In no way does it represent a sample of evaluations or a discussion of the best or the worst examples of evaluation research in the courts area. Rather, I have focused on research that deals with what I consider to be some of the more innovative and important ideas which have been implemented in the courts in recent years.

I have also taken certain liberties and excluded a number of topics that reasonably could have been included in this discussion. With but one exception I have not touched on the enormous literature on juvenile courts (which includes some of the best examples of court-related evaluation research), because this topic merits separate treatment. Also, with but one exception, I have not looked at appellate courts, although here, too, there are innovations

and a large body of literature assessing them. Other areas I have not touched on are probation, juries, court unification efforts, and a number of administrative innovations, all of which are of importance and are yielding a rapidly growing, if fugitive, body of evaluation literature. I have also slighted the vast literature on the Federal courts, only referring to some of it from time to time.

Specifically I have limited my review to some of the work on pretrial release and pretrial diversion, some efforts to restructure the process of adjudication, and studies of efforts to reshape sentencing. Part Two focuses on the pretrial process; Part Three the process of adjudication; Part Four, posttrial processes. Part Five briefly examines the evaluation findings on several miscellaneous innovations and offers a few tentative conclusions.

THE PRETRIAL PROCESS

In the United States money bail has been the most widely used condition for securing pretrial release. This has usually meant that the criminally accused has had to turn to a bondsman who, for a fee, will post bail. While this facilitates the release of large numbers of people with limited means, it also works a hardship on the criminally accused, particularly those who are either too poor to afford a bondsman's commission or with whom for any of a variety of reasons a bondsman refuses to do business. This initial hardship has a multiplier effect. There is considerable cumulative data to suggest that those not released prior to trial stand a greater chance of being convicted and, if convicted, receiving stiffer sentences.⁴

While many of the defects of money bail were pointed out by reformers of the 1920's and 1930's,⁵ they were rediscovered in the later 1950's and 1960's. Stimulated by these earlier writings and the work of Caleb Foote,⁶ Attorney General Robert Kennedy launched a drive for bail reform in the mid-1960's. Sharing his thinking in this effort were Daniel Freed and Patricia Wald of the Department of Justice,⁷ and Herbert Sturz, who founded the Vera Foundation to pursue pretrial release reform in New York City.⁸

Robert Kennedy began a quiet revolution in pretrial release in the Federal courts when he instructed U.S. Attorneys to agree to release arrestees on their own recognizance whenever possible, and Herbert Sturz and the Vera Foundation began a major trend in State courts by instituting a release on recognizance (ROR) program in the Manhattan criminal courts in 1961. Since then, this program and variations of it have been adopted by scores of communities across the country. While many of these programs have continued to struggle along on undependable sources of funding,

others are by now fully institutionalized and firmly ensconced in their court systems.⁹

While structured in a great many ways and located in different agencies, the principal aim of these programs is to minimize the inequalities of financial conditions for pretrial release, not only by promoting for ROR those with close ties to the community, but also by pressing for court-administered 10 percent bond programs (which return to the arrestee most of the money which otherwise would go to a bondsman), supervised release, defendant-notification programs, and police citation programs.

During the decade from 1962 to 1972, there was a marked increase in use of ROR and a corresponding decrease in the proportion of arrestees detained because of an inability to post bond.¹⁰ This liberalization of pretrial release can be attributed directly to the efforts of those mentioned above and to the programs they promoted.

While it is clear that these people and their efforts have made a substantial difference, the evaluator's question is: What in particular accounts for these rather dramatic changes? Is it due to publicity by important people? Is it the introduction of specialized pretrial release agencies in some courts? Is it the validity and reliability of scientific predictors of likelihood of appearance based upon ties to the community? Is it the limited capacity of the jails in the face of a rising crime rate? Or even more generally, is it a liberalized attitude about pretrial release? Despite two decades of bail reform "experiments," answers to the questions are difficult to supply.

Perhaps the most dramatic liberalization of pretrial release came about not by an elaborate program or a scientific procedure designed to predict likelihood of future appearance in court, but by the stroke of a pen. When the then-Attorney General Robert Kennedy signed a directive to U.S. Attorneys to agree to the release on recognizance of those defendants in their jurisdictions who appeared to pose no special likelihood of fleeing, he, in effect, opened the doors of the Federal jails and cut deeply into the business of bondsmen.

The first formal program systematically to base the release recommendation of information about arrestee's ties to the community and present this information to the court was the Vera Foundation's Manhattan Bail Reform Project. Established in 1961, this project had volunteers interview arrestees shortly after arrest, assign them "points" according to the number and types of their ties to the community (e.g. length of time in the city, permanent address, employed, living with family, etc.), and then pass on to the court this information, and, when appropriate, ROR recommendations.

This basic model continues to operate in New York City and a great number of other jurisdictions, although with considerable variations. In California, for instance, the Legislature has mandated that all courts establish a "bail schedule" so that arrestees, if able, can post bond almost immediately upon completion of booking. Formal pretrial release agencies intervene at a later stage, just prior to arraignment, which may be up to 2 or 3 days after arrest, and after a number of people have already secured release on bond.

There are, of course, a host of ways programs can be structured and a host of indicators of community ties that might be used. The central questions for evaluation researchers in this area are which structures are most likely to facilitate pretrial release and which indicators are best able to predict subsequent appearance in court. There are other important questions as well: At what stage and how quickly is someone released? How effectively do the programs minimize the importance of such invidious factors as wealth and race? At what price--in terms of nonappearance rates--is release increased? What is the effect in terms of additional criminal activity of increased pretrial release?¹¹

Despite an immense literature assessing the operations of various pretrial release programs, there is no evaluation of which I am aware that provides firm answers for any of these questions. While there has been some excellent research assessing a number of different types of programs and their problems, most of it either possesses major defects or ignores basic questions.¹² Despite the purported experimental nature of many pretrial release programs, with but one creative exception none of them has been conducted as a true experiment, in which arrestees have been randomly assigned to experimental and control groups, i.e., the pretrial release program and the regular court processes.¹³ And even when a program has established clear and objective criteria for eligibility for release, these criteria are invariably filtered through additional discretionary judgments of program staff and judges in a way that undercuts the capacity of researchers to determine how powerful these "objective" factors are in predicting future appearance and recidivism.

While a number of studies have attempted to measure the efficacy of pretrial release programs by comparing release and appearance rates of those going through the program with arrestees who have not, the comparisons are invariably flawed. For instance, the report on the initial Vera project compared the subsequent appearance rates of those the project recommended for release and who were subsequently released by the court with all other arrestees who were released. The report found that project-recommended releasees had a higher reappearance rate than the others. While this suggests that the project was successful, it does not answer

the central questions. Was the differential due to the fact that the project selected the "good risks" and left the poorer risks to their own devices? If so, might not these "good risks" have been spotted by the court anyway, and released without the intervention of the project? Were the differentials due to the fact that the project identified those with close ties to the community or because highly motivated volunteers continued to remind releasees under their watchful eye that they had to return to court? If community ties are good predictors, which indicators are most effective? None of these and similar questions were asked or answered in the study, which nonetheless concluded with an enthusiastic endorsement of the project.

A second flaw in the initial Vera Project was that it did not carefully identify objective criteria for eligibility for pre-trial release and then scrupulously adhere to them. The decision allowed project staff to exercise considerable discretion in making recommendations and permitted judges freedom to accept or reject them. As long as subjective, discretionary factors influence decisions in unknown and uncontrollable ways, it is impossible to make inferences with confidence about the importance of the known, "objective" factors. In short, it is impossible to attribute relative importance of some factors without first identifying the entire set of influencing factors.

In an extension of the inquiry begun by the Vera Foundation and others, Michael Gottfredson worked with the courts in Los Angeles to construct the closest thing to an experiment that there has been in this area. He monitored the post-release activities of a sample of defendants who had been released on their own recognizance through the auspices of the Los Angeles OR (on recognizance) Unit and also watched a group of defendants who had been deemed ineligible for OR release by the unit but nevertheless had been ROR'd by special arrangement with the court. He compared the appearance and rearrest rates of these two groups to determine how accurate the Vera-promulgated community ties factors were in predicting future nonappearance and rearrest. To advocates of a scientific release procedure, his findings are disillusioning. While those who would not normally have been released did have a higher failure to appear (FTA) and rearrest rate, Gottfredson found that the overwhelming majority of people in both groups eventually appeared in court and were not re-arrested on serious charges. He found that the conventional predictors had little, if any, real power.

Put another way, he found that in order to prevent one FTA or rearrest, a great many people who would not "lapse" if released would have to be detained as well. The difficulty, Gottfredson pointed out, lies in the fact that the overwhelming majority of all arrestees are likely to reappear and not commit serious offenses if released. Developing a reliable model capable of selecting

the 1 prospective deviant out of every 15 or 20 people is much more difficult than trying to predict behavior engaged in by, say 1 out of every 2 people.¹⁴

Despite these difficulties, it is disappointing--although perhaps not surprising--that no truly experimental projects have been designed and assessed and such little first-rate research has been conducted. While the Vera study marked an important beginning, and Gottfredson's study is an important effort, unfortunately neither marked the beginning of an increasingly sophisticated evaluation methodology.¹⁵

But if the past is disappointing, the future need not and, in fact, may not be so. Recognizing the limitations on past "experimental" pretrial release projects, the National Institute of Law Enforcement and Criminal Justice (NILECJ) has recently supported a major effort by the Lazar Institute to help plan, implement, and evaluate demonstration pretrial release programs which one may hope will overcome the problems I have identified here. After almost two decades of "experimentation" it is time for a true experiment.

PRETRIAL DIVERSION

Pretrial diversion is a term commonly used to refer to formally constituted programs in which the criminally accused is removed from the traditional adjudication process and instead offered a short-term, community-based program of supervision and supportive services in anticipation of having criminal charges dropped upon successful completion of the program. While there are no accurate figures on the number of these programs in this country, they have grown from a handful of programs funded by the Department of Labor in the late 1960's to several dozen by the late 1970's. During this interval they caught on like wild-fire and only now is their blaze beginning to fade. It is still unclear whether the idea will burn itself out completely or continue to burn at a steady but reduced level. Whatever the case, pretrial diversion remains as one of the major "new" ideas of the decade and as such is worthy of careful attention.

Diversion programs vary immensely in structure and purpose.¹⁶ Some focus on young persons, some on first offenders, others on those charged with misdemeanors, and still others on those charged with felonies. Some provide employment counseling, training, and placement, while others offer psychological counseling. Some are designed for those with drug or mental problems, while others categorically exclude such people. Despite these differences, diversion programs share a number of features in common. They are justified on the basis that they provide a more effective alternative to traditional adjudication, i.e., that those who

accept their services are less likely to recidivate, more likely to improve their self-image, secure gainful employment, or become self-sufficient than those who go through traditional court processes. A secondary benefit, proponents claim, is that they are cost-effective, i.e., that they provide a cheaper alternative than traditional adjudication in the courts. Finally, there is the claim that by diverting these cases, courts are free to spend more time dealing with the fewer serious cases.¹⁷

Developing just as LEAA began to insist on evaluations of projects it supported, many diversion programs have had substantial evaluation components, and by now there is a small library of research reports on diversion programs. There are even several book-length studies reviewing these reports, specifically:

Abt Associates of Boston has prepared major assessments of the original nine Department of Labor-funded diversion programs,¹⁸ and has also produced a major study assessing more than 100 evaluation reports of various diversion programs across the country.¹⁹ The National Science Foundation's Division of Social Systems and Human Resources in the Research Applied to National Needs (NSF-RANN) also supported a major review and assessment of evaluations of pretrial diversion programs, and the results are reported in Pretrial Intervention Strategies, by Roberta Rovner-Pieczenik.²⁰

Two questions concern us here: What do these surveys say about the strengths and weaknesses of the evaluation of diversion programs? And, in light of this, what can we say about the claims of diversion as an alternative to adjudication?

Both the Abt and RANN reports focused on many of the same diversion programs, and both of them reviewed the better-known, well-funded, well-staffed evaluation efforts attached to these programs. Thus their studies can reasonably be taken to be an examination of the best examples of evaluation research in this area. Despite this, both came to similar disillusioning conclusions. They found that on the whole diversion evaluations were not well-conceived or well-executed; and that as a consequence, they could not draw any firm conclusions about the effects of pretrial diversion. Posing the question, "Is Pretrial Intervention meeting its goal of reducing recidivism?" the Abt review of the reports on the nine diversion programs sponsored by the Department of Labor (DOL) concluded that the question "has not been satisfactorily explored."

Eight of the nine DOL projects reviewed had not even attempted to collect data from a control group.²¹ Nor, the report continued, were attempts to answer the question for other "experimental" projects in Washington, D.C., (Project Crossroads) and New York (the Manhattan Court Employment Project) successful. The report

then turned to two instances where some type of control group had been assembled (Franklin Zimring's reanalysis and retrospective construction of a control group for the Manhattan Court Employment Project (MCEP) and the one of the nine DOL projects which had made an effort to collect data on a "control" group), but still observed that "[I]n both instances, the number of participants represented is painfully small.... Drawing conclusions from the experience of so few people in only two projects is a difficult undertaking indeed."²² The report concluded: "Ultimately, measures of the impacts of pretrial diversion as opposed to traditional treatment can only be resolved by randomized control groups composed of subjects eligible in all respects for program participation."²³

Roberta Rovner-Piecznik's summary of the more recent RANN-supported review of diversion program evaluations reached similarly disheartening conclusions, and went on to offer the following explanation:²⁴

This author's assessment of program reports indicated major technical weaknesses in the design and implementation of evaluative research. First, evaluative research was not an integral part of early program planning. This resulted in ex post facto research designs beset by methodological and operational problems. Second, evaluation research was given a low priority in the allocation of program resources. This resulted in evaluations that asked and answered only a minimum number of questions. Third, programs were often asked to first demonstrate their feasibility and then their effectiveness. Program staff often interpreted this ordering of priorities to concentrate all their efforts on the former task. Fourth, and perhaps most important, evaluation research was seldom policy oriented. Rather than increasing a policy-maker's understanding of how, why, and under what conditions a program could be most effective, so that policy and procedures could be modified and refined, statistics were used to either "prove" or "disprove" effectiveness.

Both the Abt and the RANN studies are now several years old. What has happened to evaluation research on diversion since then? The Abt report referred to Franklin Zimring's reanalysis and reconstruction of retrospective control group of the Manhattan Court Employment Project, and his efforts are worth considering.²⁵ Confronting an evaluation which purported to show that the program's divertees had a substantially lower rearrest rate than those in a control group, Zimring argued that the findings were

altogether misleading because the two groups--the program and the "control" groups--were in fact not comparable in all respects, save for participation in the program. Zimring's complaint was that the study did not compare all the participants who had entered the program--"successes" and "failures" alike--with a control group comprising people with identical characteristics. He argued,²⁶

Those who succeed in treatment are a biased sample of all those who entered the project--their success in treatment shows that, all along, this was the group most likely to succeed. Assuming that the control group is a perfect duplicate of persons who entered the project, the correct comparison is that of rearrests among total project participants with the rearrest record of the control group [emphasis added].

To counter what he felt were biases in the two groups, Zimring retrospectively constructed a more comparable control group--comprising those who appeared to be eligible and likely to have entered the diversion program had they been asked--whose subsequent behavior he then contrasted with all the participants in the diversion program. Having done this, he found roughly similar rearrest rates for his retrospectively selected control and reconstituted experimental groups, and tentatively concluded that the diversion program was not markedly successful in achieving its goal of reducing recidivism.

The most important consequence of Zimring's work was that it led NILECJ to fund a major evaluation of MCEP, using a design which, for the first time, obtained people eligible and willing to be participants in the diversion program and then randomly assigned them to experimental and control groups. Preliminary findings of this evaluation are supportive of Zimring's skepticism about diversion. A still tentative interim study reports no substantial differences of rearrest and employment between the experimental and control groups.²⁷ As far as I know, this research--still in progress--is the first methodologically sound evaluation of a pretrial diversion project. It overcomes what I argued were the two major flaws in the evaluations of pretrial release programs: the lack of equivalence in the experimental and control groups, and the lack of reliance on objective nondiscretionary criteria for participation. As such, it should serve as a model for future evaluation studies in the courts area.

THE ADJUDICATIVE PROCESS

Plea Bargaining

Roughly 85 to 95 percent of all criminal cases are resolved through pleas of guilty rather than trial, but despite this there are pitifully few careful studies of the charging process and the decision to plead guilty. There are still fewer examinations of the consequences of pleading guilty. And there are only a small handful of studies assessing efforts to eliminate or restructure plea negotiations. Indeed, despite the fact that only a minute fraction of all cases are disposed of by trial, the jury is subject to much more empirically based research.

As a result of this paucity of research, plea bargaining and all its various forms are not well understood, and alternatives not given serious consideration. Because plea bargaining has come to explain so much, it is in danger of explaining nothing. Plausible assumptions substitute for empirically generated assessments, and "logic" replaces careful inquiry.

There are, however, signs that plea bargaining is beginning to receive the careful empirical (as opposed to only critical) attention it deserves. Within recent years there have been a spate of carefully documented studies on plea bargaining, the variety of forms it can take, and the consequences of each.²⁸ NILECJ recently funded a major study of plea bargaining in a number of jurisdictions in this country.²⁹ Albert Alschuler has published a series of seminal articles which, while admittedly "journalistic," have yielded important insights and stimulated lively debate.³⁰ LEAA recently supported a conference on plea bargaining which for the first time brought a number of scholars and practitioners together to report on the findings of studies of plea bargaining and explore their implications.³¹ And within recent years a number of jurisdictions have undertaken major plea bargaining "reforms" in their jurisdictions, and LEAA has had the wisdom to commit substantial sums of money to assess some of these efforts.

This lively and growing body of literature is providing a set of tentative answers for a set of problems appropriate for evaluation research. Some of them result in more sophisticated definitions of a complicated process of negotiation in the criminal courts, and others report on the consequences of trying to eliminate, reduce, or structure plea negotiations.

Perhaps the most important piece in this second set of studies is an article by Thomas Church which reported on an effort to eliminate plea bargaining in one Michigan court.³² Taking advantage of a "natural experiment," Church cataloged the changes brought about by a new prosecutor who was elected on a campaign

platform to "abolish plea bargaining." Church's findings point to a mixed assessment of this effort.

Trial rates initially increased, although in most cases prosecutors, judges, and defense attorneys quickly adapted to the new policy without significantly altering their old goals. There was an increase in the number of cases handled as "youthful offenders," an alternative which continued to provide for great flexibility in disposition. Judges began to take a more active role in "sentence bargaining," thereby guaranteeing much of what had previously been assured in charge bargains. Still, there were some marked changes brought about by the new policy. Some judges were reluctant to informally guarantee sentences and the higher trial rate continued in their courtrooms. There was widespread belief that the magnitude of reductions offered by the prosecutor declined.

Other, less ambitious efforts have been undertaken in Denver and Alaska, where the goal is to restructure and reduce plea bargaining. The Denver effort is too recent to have produced results, but the Alaska effort has been underway for several years and has, from the outset, been accompanied by a substantial evaluation component.³³

Another experiment in restructuring plea negotiations took place in Miami, Florida. Here, at the urging of a group of scholars from the University of Chicago Law School and with support from LEAA, the Miami court system established a pretrial settlement conference which involved not only the prosecutor and defense attorney, but, if they wished, the defendant, the victim, and the arresting officer. This meeting was presided over by a judge. In a tightly controlled experiment the team from the University of Chicago compared outcomes of cases handled in the regular manner, i.e. by bilateral negotiations between the prosecutor and defense attorney, and by this larger group.³⁴

In the project's view, the heart of the experiment was the presence of the judge at the conference. It was hypothesized that his or her presence would legitimize the process of negotiation and protect the dignity of the criminal process by assuring that prosecutors and defense attorneys were open about their reasons and based their decisions on legally relevant factors. Detractors of the idea argued that the presence of a judge and the defendant would curtail consideration of important factors, slow the process down to a snail's pace, and possibly disrupt the proceedings. Given this, perhaps the most important finding of the evaluation was that the experiment took place and that the conference worked. It did not create running backlogs, the defendants and victims were not disruptive, and the judge's presence was not perceived as coercive, nor did it appear to demean the dignity of the court. An additional result of some significance was that the

CONTINUED

2 OF 5

conference dramatically decreased the elapsed time from arrest to disposition (although not necessarily the total amount of time spent considering each case).³⁵ Furthermore, the rate of trials for cases which went to conference was roughly the same as for those in the control group which were handled in the routine manner.

Substantively, the results of this experiment were positive but modest. While those who participated in the experiment were generally satisfied with their treatment, so too were those who participated in the traditional procedures. The slightly higher rate of satisfaction among those victims and police in the experimental group who actually attended the conferences was encouraging. (While appreciating the opportunity, the vast majority of victims and police declined the invitation to attend.) But it appears that this increased satisfaction was due largely to the fact that participants learned of the outcome of the cases in which they were involved, a bit of information that just as easily could have been passed on by a post card.

Perhaps the greatest importance of this project is that it took place. The effort and the careful evaluation stand as a direct challenge to those who claim that it is impossible to innovate, experiment, randomly assign cases, and carefully monitor the behavior of a fragmented and disparate group of officials who traditionally have resented intrusion into their private domains. Although the results of the experiment were not startling, this and similar experiments are important reminders that true experimentation and first-rate evaluation can and do take place in the courts area.

None of the studies on plea bargaining, however, has systematically assessed what is of primary interest to most of those concerned about the process: What would have happened if the defendant had opted for trial rather than plead guilty? Critics of plea bargaining argue that two quite different types of results are likely to occur. Some argue that defendants who plead guilty would have been acquitted at trial; others maintain that defendants who are sentenced after trial are likely to receive substantially harsher penalties for no other reason than that they exercised their right to trial. The extent and magnitude of such behavior is hotly debated, but as with most debates about plea bargaining, "logic" and anecdotes rather than systematically obtained data substitute for hard evidence.

It is unlikely, however, that this void will be filled in the future. Even the most ardent proponents of experimentation are not likely to advocate experiments that would wreak havoc on the constitutional rights of the accused. And, given the gross imbalance and the "natural selection" between the handful of cases that go to trial and those which are settled by guilty pleas,

comparisons between the two groups are suspect. Sentences for the 85 to 95 percent of offenders who plead guilty cannot meaningfully be compared with the sentences of the 5 to 15 percent who are trial convicted. Experimental exploration of this aspect of plea bargaining might best be limited to simulation.

Expansion of Public Defense Systems

During the past two decades the Supreme Court has expanded dramatically the coverage of the sixth amendment's provision guaranteeing right to counsel. This right now covers virtually anyone accused of any criminal charge that involves a possibility of a jail or prison term.

The threshold question is, does presence of an attorney make any difference? Perhaps because the legal community and liberal sentiment are so overwhelmingly in favor of the expansion of the right to counsel, there is little systematic evidence on the effect of counsel in various types of criminal cases. There is, so far as I know, only one such effort to experiment with and evaluate the impact of presence of counsel, and it focused on juvenile courts, where, historically and philosophically, the presence of counsel has been more problematic. In their book, In Defense of Youth, Vaughn Stapleton and Lee Teitlebaum report on a project which provided attorneys at no cost to a group of randomly selected juveniles charged with delinquency in two different courts.³⁶ The outcomes in the cases with attorneys were compared with those in the control group, and these data were supplemented with in-court observations. The authors mount overwhelming evidence that presence of attorneys not only significantly altered the nature of the proceedings, but that it dramatically affected outcomes as well, the result being that those children represented by attorneys were far less likely to be adjudicated delinquent and sent to institutions.

While the authors carefully limit their findings to the two juvenile courts at which they looked, the book is nevertheless a model for sociolegal research and raises a host of questions about the adversary process in general. This book too stands as eloquent testimony to the fact that first-rate evaluation research can be undertaken in the court system and on a topic which is subject to heated debate and intense feelings.

If the presence or absence of attorneys is the threshold question, the more modest but also more policy-relevant question deals with the consequences of various types of representation, most notably the comparisons between private and public defense attorneys and among various ways of organizing public defense systems (e.g., public defender (PD) systems, appointed counsel, etc.). Here there is a substantial and interesting body of research, although there are--to my knowledge--no truly experimental evaluations which examine the benefits and costs of competing defense delivery systems.

In a recent major work sponsored by the Twentieth Century Fund, Robert Hermann, Eric Single, and John Boston compared public and private defense systems in three cities, Washington, D.C., New York City, and Los Angeles.³⁷ Amassing data on the participants' perceptions about the way cases were handled and outcomes of cases, (i.e., conviction rates and sentences), they concluded that, on the whole, public defenders compared favorably with private attorneys. Their findings are by and large consistent with other studies comparing public defenders with private attorneys in Denver and San Diego.³⁸ These findings run counter to conventional wisdom and must surely be a boost to advocates of public defense systems.

There is, however, a false precision in quantification of social processes, and the problem may be of sufficient magnitude here to obscure important differences. Recent descriptive studies of criminal courts show that courtroom work groups adopt a routine way of handling most cases, and that this informal standardization of tasks tends to minimize differences among types of attorneys.³⁹ Most attorneys handle typical cases in a routine manner. If, as it appears to be in many courts, this is true, then the telling differences among attorneys may occur only once in a great while, too few times to make themselves felt in light of the large number of "routine" cases. While such "special" cases occur only occasionally, they are important, for they may mean the difference between prison and freedom. It is these cases that might best reveal differences among types of attorneys, but because they may occur only occasionally such differences are most likely to be found through qualitative rather than quantitative analysis.⁴⁰ Looking beyond outcome to process, Jonathan Casper interviewed convicted offenders about their attorneys and has drawn less favorable conclusions about PD's.⁴¹ He consistently found that clients of PD's strongly believed that their attorneys were less concerned with their cases and did not try as hard as private attorneys would have. But even here Casper's findings tend to find differences only with respect to satisfaction, not outcome. That is, while he found that PD clients were far less satisfied than clients of retained counsel with the ways they were treated, he did not find that they were treated more harshly by the court.

While the comparison between public and private attorneys is instructive and useful for setting standards against which public defender systems can be assessed, perhaps the more policy-relevant concerns deal with the alternative ways public defense systems can be organized. Here there are a myriad of questions about the ways this service can best be offered (e.g., essentially full-time court-appointed attorneys; occasionally appointed attorneys drawn from a cross section of the private bar; full-time public defender offices, etc.); where these offices are located (e.g., in neighborhood offices, intermingled with the offices of private attorneys, in offices in the courthouse, etc.); and how these offices are structured

(e.g., "man-to-man" versus "zone defense," misdemeanor/felony specialization, etc.). While there is a considerable literature debating these and other topics, there are few systematic studies charting and comparing the impact of particular alternatives and still fewer evaluations of truly experimental programs testing alternative delivery systems. Given the importance of defense attorneys, the multiplicity of programs that exist across the nation, and the recent rapid expansion of public defense services, it is disappointing that there is not more and better evaluation research in the area. Fostering an experimental attitude and helping design experimental demonstration programs ought to be a primary goal of the national defenders' organization.

Changes in Prosecutors' Offices

During recent years prosecutors' offices have also expanded in size and functions. While there is no single major change in these offices comparable to the explosion of public defense systems, nevertheless they have undergone a major transformation. Many of them have grown from small collections of relatively autonomous individuals sharing an office to become much larger and highly organized teams, with the chief prosecutor increasingly assuming the role of "manager." One important innovation in prosecutors' offices which symbolizes this transformation to professionalism is the implementation of automated information systems, a change which, on the surface, does little more than put on computer tape and make available at the push of a button information that previously was recorded on paper and placed in file drawers. But under the creative efforts of Charles Work and William Hamilton, the prosecutor and management specialist team which conceived of and implemented the first such system (known as PROMIS, for Prosecutor's Management Information System) in Washington, D.C., this system has been used to stimulate a wide variety of improvements in courthouses across the country.

The ready availability of information on pending and completed cases facilitates streamlined management and increased accountability. In Washington, researchers have been able to monitor the workload and records of individual prosecutors to better allocate time and to spot problems. The ability to canvass the entire backlog of cases has allowed chief prosecutors to establish priorities and allocate resources to the more serious and older cases. Automated lists of defendants, victims, and witnesses facilitate scheduling and permit automatic notification of court appearance dates.

In Los Angeles, the chief prosecutor adopted the PROMIS system after a study revealed substantial inconsistencies in prosecutorial practices in his several branch offices.⁴²

As more offices adopt the PROMIS system or other similar information systems, the amount of data from different cities will

increase and thereby facilitate more sophisticated studies comparing different practices across cities and over time. It is already clear that PROMIS has made a difference, and during the next few years it should be possible to determine just how important this difference is.

The criminal law covers a broad spectrum of behavior and traditionally prosecutors have been expected to be experts in all areas of the criminal law and all steps of the court process. This attitude has changed in recent years. Increasingly prosecutors are specializing, handling only certain types of cases or only a limited number of duties. Although there are now a wide variety of types of specialized units within prosecutors' offices in major cities, I will restrict my comments to a review of only one such program, the creation of specialized units to prosecute "career criminals."

During the past few years, LEAA has supported efforts to increase the certainty and severity of criminal conviction against violent and serious offenders. In many offices this has been translated into projects to increase the level of effort in cases involving serious charges, or where the accused has a history of committing serious offenses. By concentrating resources on these cases, it is expected that they will be disposed of earlier and result in less plea bargaining, a higher rate of conviction, and longer sentences, all of which should lead to a reduction in crime. At present, LEAA has, or is supporting, 18 such career criminal projects, and no doubt there are numerous others also in operation.⁴³

Although defender organizations have voiced concern over the selective attention given to the prosecution of a group of cases, prosecutors and police strongly endorse the idea. One project that has been particularly well-received and was designated an "Exemplary Project" by LEAA, is the Major Offense Bureau (MOB) in the Bronx County District Attorney's Office.

The MOB project consists of several components, with a selected group of specially trained prosecutors backed up by a well-trained staff. The project created a special trial court whose primary function is to handle MOB cases on an expedited basis. It provided additional resources to facilitate early case preparation. MOB also maintains a policy of full disclosure and limited plea bargaining. Reports prepared by the District Attorney's office and the National Institute of Law Enforcement and Criminal Justice⁴⁴ stress that this effort has substantially shortened the time to disposition. MOB cases take a median of only 97 days from arrest to disposition in contrast to a median of over 400 days for non-MOB cases in the office.⁴⁵ Similar positive results were also found for conviction rates and sentences. These reports conclude that by increasing resources to prosecute selected cases, the result will be more, faster, and "better" convictions.

No one seriously doubts that this effort has had an effect. However, there remain serious questions: How much difference? Would not some additional effort "naturally" be mounted in serious cases anyway? What precisely accounts for the difference--the special prosecutors? the special court? or a generalized "get tough" attitude? Additionally, what alternatives were sacrificed as a result of a concentration of limited resources on these cases? Were the results worth the cost? Despite its designation as an "Exemplary Project," this effort was not tested against a control group. Rather, MOB cases were contrasted to a group of other non-MOB-eligible cases. Thus there is no real test of alternative rival hypotheses, and no convincing answers for the questions posed above.

While the LEAA publication describing MOB as an exemplary project does an excellent job in describing the project origins, functions, and operations, the chapter on evaluation does not actually evaluate the unit. Rather, it discusses how such an evaluation might be conducted.⁴⁶ Before LEAA supports too many additional career criminal units, it is hoped that it will pause and take the advice of its research division, so that when it publicizes the benefits of pioneering "experiments" and urges widespread adoption, they are in fact true experiments supported by hard data.

This same advice applies to a great many other "experimental" projects currently underway in courtrooms across the country. There are an increasing number of reports on the benefits of special victim/witness assistance units, early case assessment bureaus, specialized narcotics prosecutors, white collar crime units and the like, but a perusal of these materials reveals few empirical studies of any sort and a virtual absence of acceptable evaluation work. It may seem foolish to call for research to validate the impact of what appear to be such obvious benefits flowing from increased resources, but the "obvious" is not always so obvious, and evaluation research offers much more than simply demonstrating that changes have occurred. It should document the magnitude of changes and offer some basis for estimating the costs and benefits of particular changes. In a world of scarce resources, the question for policymakers is not only will more resources make a difference, but how can scarce resources best be used? It is in this context that detailed evaluations--which facilitate consideration of alternatives--of even the most obviously successful projects can be of greatest benefit.

THE POST-TRIAL PROCESSES

Sentencing

We are currently witnessing a major reassessment of the philosophy of the criminal law and a rethinking of the use and efficacy of the criminal sanction. Some argue that the criminal sanction is overextended and too severe, while others argue for more extensive coverage and harsher penalties. Still others call for greater consistency in sentencing and advocate procedures designed to reduce disparity. This concern has already made itself felt in a number of States and many more new sentencing schemes are expected in the next few years. While many of these new policies are incremental adjustments to old ideas, some of them are strikingly different from past practices, and of these, a few are being carefully watched and assessed.

Two major revisions of sentencing practices, in Massachusetts and New York, have been fully implemented and are being closely watched. They sprang from a desire to make penalties against selected offenders both harsher and more consistent. In 1975 the Massachusetts Legislature amended its existing gun control laws to require that persons convicted of unlawfully carrying a firearm be sentenced to no less than a year in jail or a house of correction.⁴⁷ Proponents of the Bartley-Fox Law--as it came to be known--hoped that the stiffer and mandatory sentences would serve as an effective deterrent to crime, while opponents argued that its inflexibility and mandatory nature would fail to adequately distinguish the unwitting and otherwise blameless citizen from the seasoned criminal, and that it would create a new and heavy burden on an already overburdened court system.

Recognizing its significance, plans were made early to assess its impact on criminal activity and the criminal justice system. To date, research efforts tracing the law's impact have resulted in one major Harvard-based study comparing the law's immediate impact on the courts and the availability of guns,⁴⁸ and another, Boston University-based study, still in progress as of this writing, that extends the analysis over a longer period.⁴⁹

Proponents of the Bartley-Fox Law claimed that one of its major benefits would be a reduction in the use of firearms, while its opponents claimed several defects: that it would catch naive and inappropriate people in its net, lead to increased trials which, in turn, would create havoc in the courts, and that police and prosecutors would change their procedures and evade the consequences of the law. Preliminary research suggests that none of these predictions has been borne out, although, on the whole, proponents seem to have been closer to the mark than opponents.

The Harvard study which compared reported gun-related charges for a 6-month period immediately preceding and following implementation of the law found a drop in arrests on charges affected by the law, and to the extent that it was possible to tell, the study concluded that this drop was attributable to the law and not to altered behavior of police or prosecutors (who on the whole had opposed the law). Furthermore, there was no substantial evidence to indicate that judges were thwarting the intent of the law by encouraging substitute charges. Other indicators also suggested that the law was producing its intended effect. Although there was no substantial drop in the number of armed robberies in which guns were used, there was a drop in the incidence of "firearm homicides" and armed assaults in which a gun was used. (There was, however, a corresponding increase in the use of other, but presumably less lethal weapons).

These inferences cannot be dismissed lightly, nor can they be accepted easily. There are at least two other related and confounding factors that might account for some of these observed changes. First, the new law was accompanied with a wave of publicity against illegal carrying and use of firearms generally, and it may have been the publicity campaign rather than the new and stringent sentences that produced this effect. Second, the police department in Boston--the jurisdiction for which the lion's share of the data were collected--was undergoing a major and well-publicized upgrading at the time the new law went into effect, during which special new units were created and new policies were announced. It may have been these new policies and not the law alone that accounted for the observed changes. We may hope some of these concerns will be addressed in the final report of the Boston University study.

As for the courts, some of the predictions of the skeptics were borne out. Massachusetts has an antiquated double tier system of courts which allows for trial de novo on "appeal" from the lowest court level, and the new gun cases created a significant disruption in the way these courts operated. With the advent of the law, these cases were contested and appealed to the higher level trial court more frequently, went to trial more frequently, and went on to appeal more frequently. They also were dismissed more frequently. Finally, people charged with gun offenses fled the jurisdiction more frequently than before. One critical observer of the impact of this law noted that its "practical effect...is not to penalize the major offenders--they usually receive a prison sentence regardless of the mandatory sentence--but rather, it affects minor offenders who would have received probation or been placed in a diversion program."⁵⁰

The Boston University interim report examined the impact of the law on the courts over a longer period of time and came up with more positive conclusions. Both interview and case file data showed that while the Bartley-Fox cases took disproportionately

more court time and resources, they did not throw the courts into anything approaching a crisis. Two factors seemed to account for this: a steady decline in the number of cases involving guns (although the study does not attribute this to the Gun Law itself) and a general decline in all types of cases. Thus, to some extent, the increased work imposed by the Bartley-Fox cases was offset by a decline in other work, which led the authors of the interim report to conclude that "there is no evidence that the Massachusetts Gun Law aggravated the Massachusetts' court system's case load problems."⁵¹

Changes in New York State's criminal code also imposed mandatory sentences and raised many of the same issues, but on a much larger scale.⁵² Adopted in 1973, the "Rockefeller Drug Laws," as they came to be known, provided for stiff mandatory minimum sentences ranging from 1 to 15 years coupled with lifetime probation upon release for those convicted of possession of specified types and amounts of drugs. Proponents of the measures, universally conceded to be the nation's "toughest drug laws," argued that the harsh penalties and the accompanying limitations on plea bargaining would stem the flow of heroin onto New York's streets. Opponents--and they included most judges, prosecutors, police, and defense attorneys--argued that the laws would have little or no effect on drug use and availability, and at best would be applied against low level pusher-users rather than major dealers. They also feared that the mandatory nature of the sentences would reduce the ability of the police to obtain informants, who traditionally seek leniency in exchange for information about sources. Finally, they argued that the increase in trials and motions which would surely come would overwhelm the courts. To counter this last objection, accompanying legislation provided for the creation of 51 new judicial parts statewide.

Concerned about the impact of the law on both heroin use and the courts, the Association of the Bar of the City of New York, in collaboration with the Drug Abuse Council, obtained a sizable grant from LEAA to study and report on the law's impact on both heroin use and the courts. Like the Massachusetts experience, the Bar-sponsored study group sought to take advantage of a natural experiment and compared criminal activity and court processes before and after the adoption of the new laws. In order to determine whether changes in New York were due to the impact of the new law or widespread market fluctuations, they also compared the availability of heroin in New York after the passage of the law with its availability in several other cities.

The law's sentencing provisions and restrictions in plea bargaining were rigorous, and the study group sought to determine their impact on the courts, particularly in New York City where the great bulk of the more serious drug use and court cases were concentrated. The study found that the police did not measurably alter their drug arrest practices in response to the new law;

they continued to concentrate on the same types of offenders and arrest at more or less their previous rate. Since the Legislature created more than 50 new judicial parts in the State and allocated 31 of them to New York City, there was no clear way to measure precisely the disruptive impact of the laws on courts. But the study did find that these new courts were eventually kept busy with business resulting from the new laws, and that these laws contributed to a dramatic increase in trials, motions, and appeals, and substantially longer disposition times. Whether this dramatic increase in workload would have continued once new "going rates" had been "established" is impossible to answer since, in 1976, a key provision--one requiring a mandatory minimum of 1 year for possession of a small amount of heroin and allowing for no plea bargaining--was relaxed and the catalyst in a potentially explosive mixture was removed.

Given the substantial costs of maintaining the increased number of judges, the study group attempted to provide a basis for making a costbenefit assessment of the new law. The group found that the new law was relatively successful in increasing the certainty and severity of sentence for those convicted, and that this represented a net increase of about 600 new prison sentences per year over previous drug felony rates--a substantial increase. Nevertheless, the authors concluded:⁵³

Implementation of the 1973 drug laws had not resulted in a measurable increase in the likelihood of punishment....The result is not surprising because even if implementation had been more successful, the potential for increased deterrence may be small because the laws focus on the sentencing state of the criminal justice process, and few crimes reach this very last stage in the adjudication process.

While the report itself did not draw any costbenefit conclusions, it did note that the cost of sending these additional offenders to prison translated into roughly \$40,000 per prison sentence. It also pointed out that these new prison sentences tended to be concentrated on first or minor offenders, since the more serious offenders were already likely to have received substantial prison sentences under the old laws.

Both the studies of the Massachusetts Gun Law and the Rockefeller Drug Laws are models of impact analysis. Both were conceived and designed before the laws went into effect; both were well-funded; both involved multidisciplinary teams sensitive to subtleties in the law, knowledgeable of the informal practices of police and courts, and sophisticated in research design and methodology. And both were conducted by groups with no axes to grind, independent groups whose primary interest was in tracing the consequences of the laws and supplying information to policymakers

in their own and other States who were contemplating similar legislation. As a result, to the best of my knowledge, the credibility of these studies has not been seriously challenged by anyone who has taken a careful look at them. We may hope they will serve as important sources of information in other jurisdictions which currently are contemplating mandatory sentences.⁵⁴

There are, of course, a host of other sentencing schemes which have received some assessment but little systematic evaluation. Many of these studies comment on plans to equalize sentences. These schemes are: appellate review of sentences, multijudge sentencing councils, and sentence review boards.⁵⁵ While there is a good deal written about these and other alternatives, there is very little empirical investigation and still less evaluation research assessing their effects. Two studies which make important steps in this direction deal with sentencing councils and review of sentences. Sherri Diamond and Hans Zeisel carefully examined the operations and results of two multijudge sentencing councils in the Federal courts and came to generally pessimistic conclusions about their capacities.⁵⁶ Pamela Samuelson conducted an assessment of the Sentence Review Division in Connecticut (a panel of judges whose function is to receive petitions from sentenced offenders to review and lighten their sentences) and concluded that it too was not very effective.⁵⁷ Samuelson found that while the division occasionally did reduce a sentence, it did so only under extreme and compelling circumstances, the most typical situation involving gross differences in sentences between two equally responsible defendants involved in the same incident. Samuelson found that the division was reluctant to move beyond this easy comparison and review individual sentences for general consistency in light of typical sentences in a class of all similar cases. She posits several reasons for this minimalist position: institutional disincentives make sentence review service a low priority for judges (it is in addition to their other duties); there is a lack of systematic information on sentencing patterns throughout the State; there is a strong belief that the Division is not the appropriate place to initiate widespread changes in sentencing policies; and there is a disinclination for judges to challenge the decisions of their fellow judges.⁵⁸ In itself Samuelson's study is not an evaluation, but it provides important data and would be extremely useful to anyone assessing the alternative methods of reducing disparity in sentencing. In this regard it stands as a model of sociolegal research.

ADDITIONAL OBSERVATIONS AND CONCLUSIONS

Some Additional Problems for Evaluation Research on Courts

This review of evaluation research studies on courts is by no means comprehensive; I have left out much and skipped lightly over a good deal. There are, however, a handful of other topics that deserve brief mention, either because they illustrate the problems of undertaking evaluative research in the courts area or because they illustrate what can be achieved in this area despite the many obstacles.

The Federal Speedy Trial Act of 1974⁵⁹ provided for the eventual adoption of limits on the length of time a criminal case can take in the Federal courts. The device to induce compliance with the time limits provided for in the act is the threat of dismissal. Because lengthy delays are so pervasive in many of the district courts, the act provided for a 5-year period of gradual implementation, during which time the district courts could develop plans to meet the act's requirements. Although funds and personnel were allocated to facilitate this planning process, surveys of the planning activity in the district courts reveal a low level of activity. It appears that few courts have engaged in a sustained process of self-evaluation. For the most part, the planning documents are products of a single author, reflecting neither the collection of important relevant data nor careful analysis. As a result, over 4 years after the act was passed and shortly before it is to take full effect, most districts are ill-prepared for it, not even knowing how they stand vis-a-vis the interim time limits specified in the act.

One of the reasons for this lax response appears to be a widespread belief among Federal judges--who on the whole opposed passage of the act--that it will be significantly modified by either expanding the exceptions to the strict time limits, postponing the date at which they are to go into effect, or extending the time limits so that fewer cases will pose a problem. It remains to be seen which, if any, of these alternatives will be adopted.

In light of what could have been, the history of planning under the Speedy Trial Act is disappointing. More than most pieces of legislation, it was meant to be experimental and to go into full effect only after careful analysis, adjustment of resources, feedback, and additional analysis over a period of several years. Yet, with but few exceptions, the result to date has been one of minimal or perfunctory efforts. Such a history, I think, serves to point out the weaknesses in the frequently made argument that those involved in implementing and adjusting to new policies are best equipped to evaluate and assess them.⁶⁰

What has been said about the failures to evaluate the interim stages of the Federal Speedy Trial Act also applies to the lack of analysis of State speedy trial rules. Although there is a large literature commenting on their likely impact and implications, there are--to my knowledge--virtually no careful empirical evaluations of their impact.

If the Federal judiciary has been reluctant to engage in serious planning for the Speedy Trial Act, this reluctance has not extended to all other areas nor to all parts of the judiciary.⁶¹ One of the best examples of true experimentation and evaluation in the courts has taken place in the U.S. Court of Appeals in the Second Circuit. Concerned about the rising number of appeals, the Chief Circuit Court of Appeals Judge, Irving F. Kaufman, worked with political scientist Jerry Goldman, then of the Federal Judicial Center, to conduct an experiment using a preappeals settlement conference to determine whether this device could reduce the burden on the appeals court by fostering settlements, thereby reducing the number of fully adjudicated appeals, proportion of cases requiring lesser judicial effort, and number of full-length briefs and oral arguments, and bringing improvement in the quality of presentation of those appeals that remained.⁶² A host of measures were used to compare these and related indicators of improvement in the experimental and control groups. While many of the indicators were positive, suggesting that the experiment was having its desired effect, most of the differences between the two groups were so small that they had no statistical significance. On the basis of this, Goldman concluded that while the experiment could not be labeled a failure, it could in no way be taken as a success.⁶³

Reaction to Goldman's study was interesting. Lawyers are skeptical of empirical research and are skeptical of conclusions drawn from numbers in tables. Perhaps most important, the adversary process fosters a determined advocate's perspective, and it was this, perhaps, that caused Judge Kaufman to persist in propounding the benefits of his preappeals innovation in the face of evidence to the contrary. Thus, while Goldman's study stands as a model of first-rate experimental evaluation research in the courts, it also serves as a reminder that conducting research is one thing and acting on it another.

Conclusions

The evaluation literature on criminal courts is much too diffuse to allow for an overview within the space of a few pages, so here I commented only on the evaluation efforts accompanying a few of the especially important innovations in the courts in the past few years. What is striking after nearly a decade of active promotion of evaluation research in the courts area is how much we still do not know. There has been pitifully

little evaluation research of real quality and of real use to policymakers. Some studies are suspect because they have been prepared by strong proponents of innovations, people with moral commitments to their programs. Many more studies are of limited value because they lack adequate research designs and comparative perspectives. Still, I have argued, even these partial studies can yield substantial insights.

Earlier I quoted Roberta Rovner-Piecznik, who suggested a number of reasons for the failure in the design and execution of evaluation research on pretrial diversion.⁶⁴ It is, I think possible to extend her arguments and apply them to evaluation research on courts generally. To summarize, she argued that: (1) evaluation research is usually not a part of early program planning; (2) it is given low priority in the allocation of program resources; (3) it emphasizes the feasibility rather than the effectiveness of projects; and (4) it does not address itself adequately to the needs of policymakers. No doubt this list could be expanded and refined, but I think that it does capture the primary reasons for the generally low quality of so much evaluation research on the courts. Because these conditions tend to remain constant, the same errors in research design and execution are repeated time and time again.

One need not conclude with such a pessimistic and deterministic view of the process. Throughout this review I have tended to dwell on the exceptions, examples of the better evaluation studies, rather than the large numbers of less successful efforts. While I have not identified all the better studies, nevertheless I have identified and examined enough of them to feel prepared to offer a few generalizations about the conditions that give rise to successful research efforts.

Above all, their very existence is, I think, testimony to the fact that first-rate and useful evaluation studies can be completed, a fact that all too many practitioners and researchers tend to ignore in their haste to settle for less-than-the-best. While second-best is as a rule better than nothing, this position is often embraced prematurely as an excuse, not a reason. Franklin Zimring has argued that it is nonsense to claim that a project is "too experimental" to be "experimentally evaluated," and I am in complete agreement with him on this point.

The existence of this small but substantial body of first-rate research suggests other lessons as well.

For the most part these studies appear to have avoided the pitfalls alluded to by Rovner-Piecznik. They usually have been planned simultaneously with the development of the program itself, not as afterthoughts tacked on after the program has begun operation. They have been accorded high priority, often because they

have remained separate and distinct from program implementation itself. That is, with separate budgets, separate staffs, and a separate identity, they have resisted the tendency to get caught up in program administration at the expense of program evaluation. Often the successful evaluation efforts have been conducted by people altogether divorced from the project management and sponsorship. While no doubt this distance hinders access, the benefits of independence appear to outweigh the hindrance. And because they have been independent, these evaluation efforts appear to have been more likely to maintain a focus on program effectiveness and impact, rather than getting caught up in examination of the host of interim goals and activities that invariably creep up to divert attention and slow down the pace of a project.

If these conclusions are more or less correct, and I think they are, they suggest some obvious and important lessons on how to structure evaluation research in a way that might promote (but by no means guarantee) higher quality and more useful research.

NOTES

1 U.S. Department of Commerce, Bureau of the Census, Historical Statistics of the United States: Colonial Times to 1970, Vol. I. (Washington, D.C.: U.S. Government Printing Office, 1975), pp. 415-416; U.S. Department of Justice, Law Enforcement Assistance Administration, National Criminal Justice Information and Statistics Service, Criminal Justice Research Center, Sourcebook of Criminal Justice Statistics--1977, ed. Michael R. Gottfredson, Michael J. Hindelang, and Nicolette Parisi (Washington, D.C.: U.S. Government Printing Office, 1978), pp. 44-46.

2 Gideon v. Wainwright, 372 U.S. 335 (1963).

3 Argersinger v. Hamlin, 407 U.S. 25 (1972).

4 Anne Rankin, "Pretrial Detention and Ultimate Freedom: A Statistical Study." New York University Law Review, vol. 39 (1964) pp. 631-636. See also Eric Single's appendix in Plaintiff's Memorandum in Johnny Roballo, et al. v. The Judges and Justices of New York City 74 Civ. 2113-MEL (1974).

5 Raymond Moley, Our Criminal Courts (New York: Minton, Balch and Co., 1930); idem, ed., The Missouri Crime Survey (New York: Macmillan, 1926); Roscoe Pound and Felix Frankfurter, ed., Criminal Justice in Cleveland (Cleveland: The Cleveland Foundation, 1922; reprinted Montclair, New Jersey: Patterson Smith, 1968); and Arthur Beeley, The Bail System in Chicago (Chicago: University of Chicago Press, 1927, 1966).

6 Caleb Foote, "Compelling Appearance in Court: Administration of Bail in Philadelphia," University of Pennsylvania Law Review, 102 (1958): 1031-1079.

7 Daniel J. Freed and Patricia M. Wald, Bail in the United States: 1964 (Washington, D.C.: U.S. Department of Justice and Vera Foundation, Inc., 1964).

8 For brief histories of the Vera Foundation see Charles E. Ares, Anne Rankin, and Herbert Sturz, "The Manhattan Bail Project: An Interim Report on the Use of Pre-Trial Parole," New York University Law Review, 38 (1963): 67-95; and Lee S. Friedman, "The Evolution of Bail Reform," Policy Sciences, 7 (1976): 281-313.

9 Lee S. Friedman, Evolutionary Economic Growth in a Public Sector (Berkeley, California: Working Paper presented at the Annual Meeting of the American Economic Association, 1977).

10 Wayne H. Thomas, Bail Reform in America (Berkeley and Los Angeles: University of California Press, 1976).

11 For an elaboration of these questions, see National Center for State Courts, An Evaluation of Policy Related Research on the Effectiveness of Pretrial Release Programs (Denver: National Center for State Courts, 1975), p. xvi.

12 For references to individual programs, see *ibid.*, *passim*.

13 The one very fine exception is Michael Gottfredson, "An Empirical Analysis of Pre-Trial Release Decisions," Journal of Criminal Justice, 2 (1974): 287-304.

14 For a related discussion, as applied to the problems of predicting "dangerousness" for purposes of preventive detention, see Note, "Preventive Detention: An Empirical Analysis," Harvard Civil Liberties and Civil Rights Law Review, 6 (1971): 291-396.

15 There is an increasingly sophisticated body of research developing in this area. See, e.g., Stevens H. Clarke, Jean L. Freeman, and Gary G. Koch, The Effectiveness of Bail Systems: An Analysis of Failure to Appear in Court and Rearrest While on Bail (Institute of Government: University of North Carolina, January 1976); and William Landes, "The Bail System: An Economic Approach," Journal of Legal Studies, Vol. 2 (1973), pp. 79-112. Most recently Christopher Aikin of the Department of Political Science at the University of California at Berkeley has been supported by the National Science Foundation in undertaking analysis of the problems involved in the development of predictive models of failure to appear and rearrest while on bail. These works should set the stage for more sophisticated future evaluations.

16 For useful surveys of these programs, see Raymond T. Nimmer, Diversion: The Search for Alternative Forms of Prosecution (Chicago: American Bar Foundation, 1974); Pre-Trial Intervention: A Program Evaluation of Nine Manpower-based Pre-Trial Intervention Projects Developed under the Manpower Administration, U.S. Department of Labor (Cambridge, Mass.: Abt Associates, Inc., 1974); and Pre-Trial Services: An Evaluation of Policy and Related Research (Cambridge, Mass.: Abt Associates Inc., 1974).

17 This is, of course, only a partial list. Each project is justified on slightly different grounds, although it is safe to say that to some degree all of them emphasize the points raised above. For a much longer inventory of objectives, see Roberta Rovner-Pieczenik, Pretrial Intervention Strategies (Lexington, Mass.: Lexington Books, 1976), p. 9.

18 Pre-Trial Intervention, see note 16 above.

19 Pre-Trial Services, see note 16 above.

20 Rovner-Pieczenik, Strategies.

21 Pre-Trial Intervention, p. 21.

22 Ibid., p. 21.

23 Ibid., p. 22.

24 Rovner-Pieczenik, Strategies, pp. 143-144.

25 Franklin E. Zimring, "Measuring the Impact of Pretrial Diversion from the Criminal Justice System," University of Chicago Law Review, 41 (1974):224-241.

26 Ibid., p. 228.

27 Research Design and Implementation: A Preliminary Report (Vera Institute of Justice: Court Employment Project Evaluation, December 1977).

28 See, generally, James Eisenstein and Herbert Jacob, Felony Justice: An Organizational Analysis of Trial Courts (Boston: Little, Brown, 1977); Milton Heumann, Plea Bargaining: The Experience of Prosecutors, Judges, and Defense Attorneys (Chicago and London: The University of Chicago Press, 1978); Pamela J. Utz, Settling the Facts (Lexington, Mass.: Lexington Books, 1978); and Lynn M. Mather, "Some Determinants of the Method of Case Disposition: Decision-Making by Public Defenders in Los Angeles," Law and Society Review, 8 (1973): 187-216.

29 William McDonald, "Plea Bargaining and the Criminal Justice Process: An Organizational Perspective," (Chicago: A paper prepared for presentation at the Annual Meeting of the American Political Science Association, September 1976).

30 See Albert Alschuler, "The Defense Attorney's Role in Plea Bargaining," University of Chicago Law Review, 84 (1975):1179-1314; "The Prosecutor's Role in Plea Bargaining," University of Chicago Law Review 36 (1968):40-112; and "The Trial Judge's Role in Plea Bargaining, Part I," Columbia Law Review, 76 (1976):1059-1154.

31 These proceedings will be published in a special issue of the Law and Society Review in Winter, 1979.

32 Thomas Church, Jr., "Plea Bargains, Concessions, and the Courts: Analysis of a Quasi-Experiment," Law and Society Review, 10 (1976):377-401.

33 A brief report on this effort is included in the forthcoming special issue of the Law and Society Review. See also "Methodologies to be used in the Evaluation of the Elimination of Plea Bargaining in Alaska" (Anchorage, Alaska: Alaska Judicial Council, Project Design Report, September 1976).

34 For a report on the project, see Wayne A. Kerstetter and Anne M. Heinz, "Pretrial Settlement Conference: An Evaluation" (French Lick, Indiana: Conference Discussion Paper for Special National Workshop on Plea Bargaining, June, 1978), and Law and Society Review, forthcoming.

35 Kerstetter and Heinz, note above, p. 33.

36 W. Vaughn Stapleton and Lee E. Teitelbaum, In Defense of Youth (New York: Russell Sage Foundation, 1972).

37 Robert Hermann, Counsel for the Poor: Criminal Defense in Urban America (Lexington, Mass.: Lexington Books, 1977).

38 Jean G. Taylor et al., "An Analysis of Defense Counsel in the Processing of Felony Defendants in Denver, Colorado," Denver Law Journal, 50 (1973):9-43. Jackson B. Battle, "Comparison of Public Defenders' and Private Attorneys' Relationship with the Prosecution in the City of Denver," Denver Law Journal, 50 (1973):101-136.

39 See Eisenstein and Jacob, note 28 above, pp. 24-39.

40 See David Sudnow, "Normal Crimes: Sociological Features of the Penal Code in a Public Defender's Office," Social Problems, 12 (1965): 255-276. Sudnow implicitly compares public defenders with private attorneys in this now classic article. In addition, see Malcolm M. Feeley, The Process is the Punishment (New York: Russell Sage Foundation, 1979, forthcoming).

41 Jonathan Casper, American Criminal Justice (Englewood Cliffs, N.J.: Prentice-Hall, 1972). See also his Criminal Courts: The Defendant's Perspective (Washington, D.C.: National Institute of Law Enforcement and Criminal Justice, 1978).

42 The study prompting this action is Peter Greenwood et al., Prosecution of Adult Felony Defendants in Los Angeles County: A Policy Perspective (Washington, D.C.: U.S. Government Printing Office, October, 1973).

43 U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice, Office of Technology Transfer, An Exemplary Project: Major Offense Bureau; Bronx County District Attorney's Office, New York, by Daniel McGillis (Washington, D.C.: U.S. Government Printing Office, 1977), pp. 131-136.

44 Ibid.

45 Ibid., p. 53.

46 Ibid.

47 Massachusetts General Laws, Chapter 269: 10(a).

48 James Beha II, "And Nobody Can Get You Out: The Impact of a Mandatory Prison Sentence for the Illegal Carrying of a Firearm on the Use of Firearms and on the Administration of Criminal Justice in Boston," Boston University Law Review, 57 (1977):96-146, 289-333.

49 A report describing some of the results of this study is Paul B. Froyd, "The Bartley-Fox Act: Impacts for Courts of the Massachusetts Gun Law" (Woods Hole, Mass.: Paper presented to the Panel on the Impact of Legislation on Courts, National Academy of Sciences, August, 1978).

50 James Dimm, "Sentencing Reform: The Control Rather than the Elimination of Judicial Discretion" (Madison, Wisconsin: Department of Political Science, University of Wisconsin, September, 1977).

51 Froyd, note 49 above, p. 85.

52 For a detailed study of this law see Anthony Japha, The Effects of the 1973 Drug Laws on the New York State Courts (New York: The Association of the Bar of the City of New York Drug Abuse Council, Inc., 1976).

53 Ibid., p. 14.

54 Brief mention should be made of a project just now getting under way in California to trace the impact of that State's new determinate sentencing law. This project is also likely to produce a careful and helpful study since it is well-funded, began operations several months before the new law became effective, is comparative, and focuses on a less strict sentencing scheme--determinate--which is more likely to have widespread appeal than mandatory minimum sentences. Thus, when completed, it would serve as an even more useful source of information than either of the two studies discussed here. The study is being conducted by Sheldon Messinger of the Center for Study of Law and Society, University of California-Berkeley, and Andrew von Hirsch and Richard F. Sparks of the School of Criminal Justice, Rutgers University.

55 See Marvin E. Frankel's discussion of various alternatives in his excellent book, Criminal Sentences: Law Without Order (New York: Hill and Wang, 1973).

56 Sherri Diamond and Hans Zeisel, "Sentencing Councils: A Study of Sentence Disparity and Its Reduction," University of Chicago Law Review, 43 (1975): 109-149.

57 Pamela Samuelson, "Sentence Review and Sentence Disparity: A Case Study of the Connecticut Sentence Review Division," Connecticut Law Review, 10 (1977): 5-89.

58 Ibid., p. 73.

59 U.S.C. SS 161-174 (1974).

60 There are too many egregious examples of project evaluations as project promotion pieces to list them here, and to single only some of them out might appear unfair. Any evaluation specialist in any of the LEAA-mandated State Planning Agencies (SPA's) has drawers full of them. In addition, I should note that "outside" evaluators are not immune from this problem. These same SPA officials also have drawers full of evaluations by "outside" researchers who appear to have "sweetheart" arrangements with the projects they evaluate. My basic point is that in order to conduct a proper evaluation, distance and perspective is required, and this is more likely to be achieved if one does not have responsibilities for day-to-day management of a project, especially if it is new and unwieldy.

61 I should note that the failures of the planning processes under the Speedy Trial Act are not due to a general inability to undertake first-rate research in an obviously sensitive area. The continuous stream of first-rate research from the Federal Judicial Center suggests quite the contrary. Since it was created, the Center has developed a capacity and well-deserved reputation for excellent and useful research on matters pertaining to the Federal judiciary. Perhaps it should have been more actively involved in planning for implementation of the Speedy Trial Act.

62 Jerry Goldman, "The Preappeal Settlement Conference: An Experiment in Appellate Court Reform," (Evanston, Illinois: Department of Political Science, Northwestern University, February, 1977).

63 Ibid., p. 59, pp. 87-88.

64 See p. 177 and Note 24 above.

Police Field Studies

A Review of Evaluation Research

Michael F. Cahn is Vice President of Public Systems Evaluation, a Cambridge, Mass. research organization. Mr. Cahn is directing two large-scale evaluative research projects in police field services.

Edward H. Kaplan is a research associate at Public Systems Evaluation. He is also affiliated with the Department of Urban Studies and Planning and the Operations Research Center at Massachusetts Institute of Technology.

John G. Peters, Jr., a former police administrator, is a senior researcher at Public Systems Evaluation, Inc.

BACKGROUND AND INTRODUCTION

Examination of the results of police field services research evaluation was particularly appropriate in 1978 since we had just passed the 10-year mark after publication of the reports of the President's Commission on Law Enforcement and Administration of Justice (1967). The decade since the Crime Commission has been an active one, productive with numerous field service studies, evaluations, and experiments being conducted; new technologies being tested; and new management tools for police being developed.

This section first will examine the issues in police field services and then circumscribe them consistent with the scope, purpose, and audience of this volume; then, explain the perspective with which the state of knowledge is being viewed; and finally, develop a normative, or idealized, model for field services experimentation as a frame of reference for looking at real-world studies.

A Police Field Services Focus

Establishing the state of knowledge attributable to the conduct of a large body of evaluation research calls for a systematic approach to its review and assessment. One "cut" for structuring such an inquiry follows police organizational lines:

- Patrol encompasses both traditional and specialized patrol modes as well as alternative responses.
- Investigation includes criminal investigation methods, procedures, and organization.

This paper was prepared with partial support from the National Institute of Law Enforcement and Criminal Justice, LEAA, through Grant No. 78-NI-AX-006L.

Copyright (c) 1979 Public Systems Evaluation, Inc.

As long as one remains aware of the responsibilities shared between the two functions, this is a useful dichotomy. First, as one might expect, the pioneering evaluation research activity of the last decade has tended to structure itself along traditional organizational lines--primarily in patrol and secondarily in investigation. This is perhaps due, in part, to the tendency of evaluators and researchers to take their "first steps" along an inquiry path which questions traditional assumptions. Clearly, the traditional assumptions in police field services are embedded in the patrol and investigation functions. Second, the technologies which relate to police field services (e.g., resource allocation; command, control, and communications, with regard to which see paper beginning p. 295 this volume; information systems) have rarely been themselves subjects of evaluation or research studies. Rather, they are more likely tools which can, and occasionally do, aid in the conduct of patrol and investigation research and evaluation studies.

This is not to reject highly significant issues of a social or organizational nature (e.g., police interaction with the public, police roles, unions, training, administration, etc.). Rather, such "institutional" issues are either specifically addressed elsewhere in this volume or have not been subjected to a significant level of competent research or evaluation to warrant special attention. In either case, the purpose of this chapter is to familiarize the nonresearcher with "what is known" about police field services through the conduct of research and evaluation studies. An examination of patrol and investigation issues is sufficient to satisfy that requirement.

A Hypothesis-Based Perspective

To the extent possible, it is desirable to review and document research and evaluation studies in the context of a standardized framework. Such a strategy affords the maximum opportunity to compare, generalize, and merge categorical conclusions. The approach pursued by the authors in the patrol and investigation domain is generally hypothesis based and involves a sequence of analytical steps:

- (1) Establish through a historical analysis the "generally accepted" functions of patrol or investigation field operations.
- (2) Document the associated measures of patrol or investigation effectiveness with attention to the emphasis upon operational, technological, and attitudinal measures.
- (3) Based upon review of "major" research efforts, identify a set of hypotheses which have been explicitly

or implicitly presented as explaining relationships between activities and the effectiveness of field operations.

- (4) Pinpoint the extent to which the research endeavors have tested the stated hypotheses either: (a) adequately, (b) inadequately, or (c) not at all.
- (5) Compile a general synthesis of findings identifying knowledge weaknesses or gaps, explicating knowledge achievements, and pointing to future research activities aimed at improving research-based information.

Figure I depicts this "logic track" diagrammatically. It should be noted that the functions of field services, as perceived by the police practitioner and police researcher or evaluator, have given rise to numerous accepted theories; often these theories have provided the spark to conduct police research and evaluation. The authors have focused particular attention on the choice of experimental performance measures, problems of data collection and aggregation, choice of experimental methodology, interpretation of statistical results, and the presence or absence of experimental controls.

The studies chosen for examination herein are certainly representative of the body of significant police field services research and evaluation in patrol and investigation. And when put in hypothesis-specific terms, the results of these "major" studies reveal much of what has been learned about patrol and investigation, specifically as the result of such studies. In this context, one can formulate substantive inputs to future research strategies through the identification of "dangling hypotheses," those that have yet to be tested conclusively, or more simply not tested at all; yet which could have considerable decision implications if confirmed or disproved.

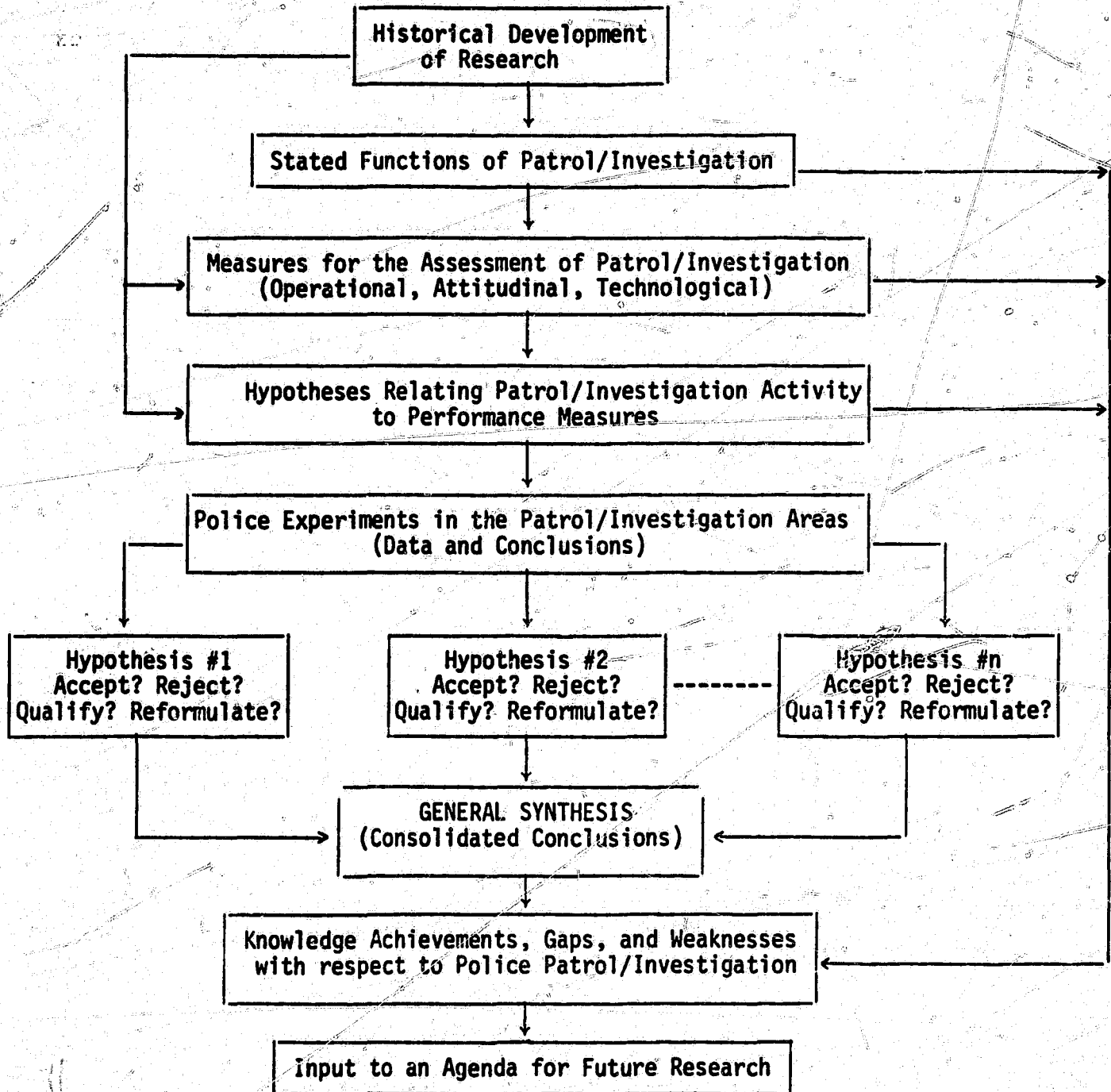
The following section examines the structure of "typical" police field services studies in terms of a normative model, since observed deviations from theoretical constructs frequently have strong implications for the assessment of such studies.

A Normative Model for Field Services Research/Evaluation

Most of the truly significant patrol studies conducted in the last decade are modeled after experimental designs, so it is useful to examine this analytical approach. In contrast, the major investigation studies have generally been conducted through the use of design strategies involving survey, case study, or both. Whereas explicitly stated hypotheses are common in patrol studies, they are virtually nonexistent in the body of in-

Figure I

Structural Analysis of Patrol and Investigative Studies



investigation research and evaluation and must be surfaced as implicitly. In essence, investigation has predominantly been subjected to a nonexperimental type of scrutiny. However, a hypothesis-based assessment of the total research body is both feasible and useful, particularly as a mechanism for identifying knowledge gaps and establishing agendas for future studies.

Given this hypothesis-based framework, a field service experiment may be thought of as consisting of five steps:

- Selection of hypotheses
- Selection of performance criteria by which to test the stated hypotheses
- Design of an experimental mechanism for testing purposes
- Execution of the experimental mechanism
- Evaluation of the experimentally derived results

Each of these steps will be examined in the following sections in the context of police field services evaluation research, with emphasis upon contrasting applications to patrol and investigation.

Selection of Hypotheses

The relevance of a particular field services research study is, to a large extent, determined by the hypotheses underlying the research. In selecting hypotheses for field services investigation, the researcher must keep in mind the potential implications that confirmation or denial of the postulated hypotheses hold for the police decisionmaker.

The majority of patrol research hypotheses have stemmed from a questioning of the stated functions of patrol; the 1967 Science and Technology Task Force Report¹ was clearly a catalyst in this regard.

The majority of criminal investigation hypotheses are predicated on the traditional functions of the police investigator. As David Farmer notes, "In police management and operations it is becoming increasingly clear that the old ways may not necessarily be the best ways."²

Selection of Performance Measures

The testing of specific hypotheses requires their redefinition in operational and measurable terms, whenever possible. Se-

lected attitudinal and operational measures should satisfy certain criteria:

- The measures should be able to gauge the desired phenomena under known conditions
- The measures should respond to environmental changes in a predictable fashion
- The measures should be based upon available or collectable data

The first of these points implies that the conditional values taken on by the measures are known, given different levels of program performance. The second point implies that the influence exerted by environmental factors other than the experimental program can be accounted for. The third point embodies the obvious --a performance measure is invalid if it cannot be observed and recorded.

It is with great difficulty that the first two points above can be satisfied. Thus, it is advantageous for a study to utilize a "family" of measures for any given study rather than to rely solely on one or two variables. Of equal importance is the a priori allocation of time to the collection of qualitative, process-oriented information. (This last point is discussed further in conjunction with experimental conduct and evaluation.)

The measures chosen for a given field services research study are, of course, dictated by the selected hypotheses. However, for any given class of studies, the same measures almost always appear to surface. Some examples are presented in Figure II.

Experimental Design

The overriding concern of patrol experimental design is the formulation of "experimental" and "control" groups (e.g., beats, squads, districts, etc.). Theoretically, both groups should be identical in character. However, while nothing changes from the usual mode of police operations in the control group, an experimental mode of police operation should be the only difference between the experimental and control group. The levels of pre-determined performance measures are monitored over the course of the experiment--typically 12 to 18 months--and at the conclusion, the control and experimental groups are subjected to statistical comparison.

In many experiments, it has not proved feasible to establish the controlled situations prescribed above. These typically rely on less "powerful" comparison groups, on before-and-

Figure II

Common Police Field Service Study Performance Measures

PATROL

Type of Study

Associated Measures

Crime Prevention

UCR Index
Local crime statistics
Victimization rates (survey)
Police visibility
Probability of crime interception
Citizen-perceived fear of crime
Citizen-perceived level of safety

Police Response Time

Travel time (with/without dispatch time)
Travel distance
Apprehension probability
Citizen satisfaction with response time

Patrol Productivity/
Manpower Allocation

Frequency of patrol passings
Patrol officer workload
Patrol officer safety (injuries)
Crime/victimization rates
Travel time
Citizen complaints
Officer complaints

INVESTIGATION

Type of Study

Associated Measures

Crime Prevention

UCR Index
Local crime statistics
Victimization rates
Citizen-perceived fear of crime
Citizen-perceived level of safety

Investigator Productivity/
Personnel Allocation

Investigator caseload
Investigator arrest rate
Investigator clearance rate
Investigator prosecution rate
Investigator follow-up rate
Citizen complaints
Officer complaints
Organizational structure
Evidence collection

Police Response Time

Travel time (with/without dispatch time)
Travel distance
Prioritization of calls
Apprehension probability
Citizen satisfaction with response rate

after observations, or both. Some quasi-experiments have managed to incorporate models of the relevant performance measures into experimental design. Utilizing this structural modeling strategy one can predict the levels of performance measures that would have occurred in the absence of experimental treatment; and hence, "replace" the control group. Clearly, the validity of such designs rests exclusively on the accuracy of the models involved.

Since investigation rarely lends itself to experimental or even quasi-experimental designs, other approaches are utilized. It should be noted that certain specialized field services strategies which incorporate investigation components--crime control teams or strike forces are examples--may be characterized by the performance measure comparison of control and experimental groups. More prevalent, however, are designs which could be categorized as: (1) the historical method, (2) the descriptive survey method, (3) the analytical survey method,³ and (4) the case study method.⁴ The historical method chronicles change in the investigative process over time as influenced by external factors (e.g., Supreme Court decisions) and internal factors (e.g., agency policy, staffing levels, etc.). The descriptive survey method is based primarily on observation as the principal data collection strategy, and employs questionnaires and interviews as its primary components. A more sophisticated survey approach is embodied in the analytical survey method which embraces more rigorous methodology including perhaps the statistically-based testing of hypotheses. Finally, the case study method usually encompasses a retrospective examination of a completed program designed to "learn from experience." Typically, this method does not involve major data collection or analysis activities.

Experimental Conduct

While the laboratories of the physicist or chemist allow desired conditions to be prolonged almost indefinitely, the environments of police field services experiments are only approximations of these laboratories. Hence, there is reason to be concerned about problems such as the contaminations of the control and experimental groups. The conduct of a successful experiment requires that the general conditions in the control and experimental groups remain the same throughout the course of the experiment.

To achieve this it is necessary to monitor the experiment continuously for the entirety of its duration. This activity may take the form of onsite observation, periodic interviews with involved personnel, routine statistical checks on various measures (e.g., number of cars in patrol areas, surveyed crime levels), or continuous time observation (e.g., use of automatic vehicle monitoring (AVM), tachographs).

The above comments are even more important when the program is of the quasi-experimental type. Without controls, one has an accountability problem with respect to the true determinants of program outcomes: Was it the program or some other environmental condition that caused the observed results? The collection of process information won't guarantee an answer, but such information may surely provide clues not otherwise available.

Continuous time observation is clearly the most powerful method for ensuring that the conditions demanded by the experimental design are maintained; in particular, the use of AVM technology is promising in the patrol area. Most patrol studies have utilized statistical checks combined with interviews, onsite observation, or both.

Evaluation of Experimental Results

The evaluation of experimental results in police field services generally, and particularly in patrol, has relied heavily upon the use of classical statistical procedures. The use of such procedures is predicated on the control group-experimental group design discussed earlier. If it is true that the sole difference between the experimental and control groups rests with the presence of an experimental treatment applied to the experimental group, then observed differences in the levels of performance measures between the two groups may be attributed to one of two sources:

- chance, or
- the experimental treatment

Statistical procedures along the lines of hypothesis-testing check to see if observed differences can be plausibly attributed to chance. If plausible attribution to chance cannot be established, then the experimental treatment is assumed to be responsible for the observed differences by a process of elimination.

Similarly, quasi-experimental procedures utilize statistical routines to compare program performance in the light of comparison groups, "before-and-after" periods, or a reasonable model as discussed earlier. However, as the design of a test program deviates from that of the classic experiment, the rationale behind the use of statistical evaluation devices is weakened. In such cases (which constitute the majority of police patrol research efforts), the collection of process data is extremely important, as such information aids in determining whether or not it was the experimental innovation or some other combination of factors which was responsible for observed outcomes.

A REVIEW OF KEY PATROL STUDIES

The primary functions of police patrol have not changed drastically over time. Sir Robert Peel stated in 1829 that:

It should be understood, at the outset, that the principal object to be obtained is the Prevention of Crime.

To this great end, every effort of the Police is to be directed. The security of person and property, the preservation of the public tranquility and all other objects of a police establishment, would thus be better effected than by the detection and punishment of the offender after he has succeeded in committing the crime. . . . Officers and police constables should endeavor to distinguish themselves by such vigilance and activity as may render it extremely difficult for anyone to commit a crime within that portion of the town under their charge.⁵

Over 140 years later, O.W. Wilson, in his book Police Administration, wrote that "the elimination of the actual opportunity, or the belief in the opportunity, for successful misconduct is the basic purpose of patrol."⁶ Similar statements may be found in any reference which discusses the functions of police patrol.

The preventive goal of patrol (as stated by Peel and echoed by Wilson) has given rise to what are here referred to as "traditional" patrol practices. For example, Peel's notion that "officers and police constables should endeavor to distinguish themselves..." is directly reflected in patrol strategies which emphasize visibility as a major component. Strategies geared towards minimizing response times also tend to aid in the creation of an image of police omnipresence.

For many years, these traditional practices were assumed by police departments to effectively deter crime. However, it had not been proven that preventive patrol does in fact deter crime, or that rapid police response affects the outcome of incidents involving the police. This is somewhat curious when one considers that standards for the evaluation of patrol forces are as rooted in history as is the notion of patrol itself. Indeed, Peel claimed in 1829 that "when in any division offenses are frequently committed, there must be reason to suspect that the police is not in that division properly conducted. The absence of crime will be considered the best proof for the complete efficiency of the police."⁷

Thus, there existed a normative sense of what patrol was supposed to achieve (i.e., the prevention of crime), and an intuitive sense of how patrol could achieve this goal (i.e., strate-

gies emphasizing high visibility patrol and rapid police response) throughout the first half of this century. With the exception of one or two efforts by individuals such as August Vollmer (in the 1920's Vollmer attempted to redistribute manpower in order to equalize unit workloads; he was also the first police chief to institute a fully motorized police patrol),⁸ no serious attempts were made to question whether or not the intuitive strategies were in fact attaining the normative goal.

One of the earlier studies (1967) which did question traditional patrol practices is contained in the report of the Science and Technology Task Force of the President's Commission on Law Enforcement and Administration of Justice. Based upon an analysis of police operations in the City of Los Angeles, the findings of this study were sufficient to raise questions pertaining to the utility of accepted patrol practice. One activity directly questioned by the Task Force was preventive patrol:

Police on "preventive patrol" cruise the streets to look for crimes in progress. Presumably, this activity prevents crime because it poses a threat of detection and immediate apprehension. However, there is little evidence on how much crime is thereby prevented or on how much would be prevented with alternative tactics.⁹

The Task Force researchers estimated that the probability of detecting a crime, such as robbery, via preventive patrol was so low that an individual patrol officer could expect an opportunity to detect a robbery about once every 14 years.¹⁰

Limited evidence presented suggested that apprehension probabilities were inversely related to response time, hence response time minimization on calls for service was seen as a useful goal. However, the Task Force stated that the effects of response time on arrests needed to be studied in more detail.

Another conventional patrol practice questioned by the Task Force was the equal manning of patrol shifts in large police departments. Such time-independent manpower allocation schemes do not reflect the time-dependent nature of demands for police service, nor do they reflect relative spatial needs for police service. The Task Force outlined some crude yet indicative statistical procedures to deal with this problem.

Stepping back for a moment, it is evident that three areas of patrol practice were being questioned by the Task Force researchers, namely:

- (1) Issues surrounding the merits of preventive patrol;

- (2) Issues surrounding the merits of rapid police response; and
- (3) Issues surrounding the merits of traditional manpower allocation schemes.

We have already discussed the history associated with the first two of these areas; interest in this third area arises from a questioning of organizational and deployment policies practiced by the police for decades.

These three notions of preventive patrol, rapid police response, and time/space-independent manpower allocation schemes are characteristic of what was assumed to be efficient and effective in police patrol. The Task Force report was one of the first studies to question these notions; most (if not all) of the patrol research that has been conducted since 1967 may also be viewed as reacting to and/or expanding upon the presumed usefulness of preventive patrol, rapid police response, and traditional manpower allocation schemes.

What follows is a review of selected studies which are quite representative of state-of-the-art patrol research. The intent here is to familiarize the reader with some of the substantive hypotheses which occur frequently in patrol research and the conclusions which have been reached by this research. While the studies reviewed do not exhaust the list of studies performed since 1967, the findings presented are indicative of what has been learned from patrol research.

The Crime Control Team Experiment¹¹

Beginning in July 1968, an experiment involving an organizational restructuring of the police department was conducted for a 1-year period in Syracuse, N.Y., a joint effort of General Electric's Syracuse Electronics Laboratory and the Syracuse Police Department. It utilized patrol strategies such as:

- the use of one-officer patrol cars;
- the use of time-dependent manpower allocation schemes; and
- the use of a mathematical model in an attempt to increase the detection probability of patrol.

The Crime Control Team Experiment was evaluated, and the results of the research included several interesting findings:

- The substitution of two one-officer units for one two-officer unit was shown to increase the apprehension capability of the patrol force.

- The use of systematic patrol procedures increased the crime interception rate to about six times the pre-experimental level.
- The new patrol strategies implemented were viewed favorably by citizens of the local community.

Although the Crime Control Team Experiment was not able to show conclusively that the program strategies reduced crime rates, it did demonstrate that organizational structures other than those normally associated with traditional police departments could effectively control crime in an urban area.

The Kansas City Preventive Patrol Experiment¹²

The results of the Crime Control Team Experiment were available in book form in 1971. While this book was generating interest among police researchers and practitioners, the groundwork was being laid for the most elaborate and well known police study to date. The intent of the Police Foundation's Kansas City Preventive Patrol Experiment was to determine the effect of varying levels of preventive patrol on outcome measures such as the crime rate and citizen satisfaction with the police. Beginning in October, 1972, fifteen Kansas City police beats were divided into three groups of five beats. Each group of five beats was to receive one of the following three levels of patrol activity for a 1-year period:

- (1) Reactive Beats--no preventive patrol was to be performed in these areas
- (2) Control Beats--preventive patrol was to be carried out as usual
- (3) Proactive Beats--two to three times the normal level of patrol was to be implemented

The general finding of this study was that variations in the level of preventive patrol had no effect on the relevant outcome measures. Stated differently, the crime rates and levels of citizen satisfaction found in reactive, control, and proactive beats were not significantly different from each other at the end of the 1-year experimental period.

The implication of these results could be far-reaching. Indeed, if it really is true that preventive patrol has little influence on the incidence of crime (this was suggested by the Science and Technology Task Force Report discussed earlier), then perhaps the amount of resources allocated to the patrol function should be seriously questioned. While some took the Kansas City

results as an indication that patrol forces could be vastly reduced in number without a concurrent degradation in service, other researchers questioned the experimental methodology and hence the validity of the stated results.¹³

While the emphasis of the Kansas City Preventive Patrol Experiment was on the relationship between patrol and the crime rate, some positive "side effects" in the area of manpower allocation have emerged. For example, R.C. Larson presents a case for "fluid patrol":

...if conditions warrant a change in the spatial deployment of units within a confined region (say a "precinct," "district," or "division"), then if procedures are followed such as those used in Kansas City, such redeployments can be made without suffering marked degradations in either actual or perceived service in the depleted regions.¹⁴

Without doubt, the Kansas City Preventive Patrol Experiment remains one of the most significant pieces of patrol research performed to date. While the experiment did not prove conclusively that preventive patrol does not influence crime rates, the experiment did demonstrate that the relationships between patrol and crime may be much weaker than has been assumed. Further investigation of the relationship between preventive patrol and crime is a major concern for future research.

Police Response Time: Its Determinants and Effects¹⁵

The Kansas City Preventive Patrol Experiment generated large amounts of data useful for the testing of hypotheses other than those central to the relationship between patrol and crime. In particular, data on response time and related outcome measures such as arrest rates and citizen satisfaction were collected. A detailed analysis of these data was released by the Police Foundation, in 1976, in a study entitled Police Response Time: Its Determinants and Effects.

While the small samples involved in this study render its findings somewhat weak, three separate surveys indicated that there was no relationship between response time and arrest rates (contrary to the drift of the 1967 Science and Technology Task Force results). Similarly, when examining the impact of rapid police response on citizen satisfaction with the police, researchers found that citizen satisfaction with response time remained at a (high) constant level over a large range of response times, thus demonstrating that citizen satisfaction does not depend on rapid police response.

Instead, the researchers found that citizen satisfaction with the police depended upon the difference between observed and expected response times, a difference that was previously not given much thought. If the police were able to respond quicker than expected, the citizen involved was more likely to be satisfied than if the police responded slower than expected.

Kansas City Response Time Analysis¹⁶

Another major study which examined the merits of rapid police response was also undertaken in Kansas City when, in 1973, LEAA awarded a grant to the Kansas City Police Department. Published in 1974, the report Response Time Analysis examined several hypotheses similar to those scrutinized by the Police Foundation's response time study just discussed. With respect to arrests, response time was shown to be inversely related to apprehension probability, but only marginally so. Also, surveys confirmed the earlier result that most of the population was satisfied with response time regardless of the actual response time involved, though the difference between observed and expected response time was again shown to be a determinant of citizen satisfaction with the police.

This study is important for another reason; it was the first study which attempted to determine the length of the delay associated with citizen's reporting of crimes to the police. This delay was shown at times to be larger than typical response times (including dispatch delays). Based on this finding and the results discussed earlier, it was concluded that the minimization of response time is not an empirically justifiable goal.

The St. Louis AVM Experiment¹⁷

While studies examining the effects of response time on various outcome variables were being undertaken in Kansas City, other researchers were focusing on the application of modern technology to police patrol using response time as a performance measure. In July, 1974, Public Systems Evaluation, Inc., began an 18-month study of an automatic vehicle monitoring (AVM) system in conjunction with the St. Louis Police Department. The potential gain in efficiency achievable via AVM was expected to arise from the practice of closest car dispatching, a practice not possible without exact car location information.

A somewhat surprising result of the initial 16-beat Phase I study was the finding that the use of an AVM system did not reduce response time, despite the expected benefits of closest car dispatching. This finding was confirmed in a city-wide Phase II effort involving all 135 beat cars.

However, as in the Kansas City Preventive Patrol Experiment, a positive "side effect" was noticed, again in the area of manpower allocation. It was suggested that appreciable improvements in patrol productivity may be obtainable by exercising the potential for improved supervision of the patrol force using AVM. Such a supervisory capability could greatly enhance the possibility of establishing a "fluid patrol" like that suggested by Larson in his review of the Kansas City Preventive Patrol Experiment presented earlier.

The Worcester Crime Impact Program¹⁸

Fluid patrol may be viewed as one alternative to traditional deployment and allocation schemes. A quite different alternative could entail a change in the role played by police officers with respect to the call-for-service response function. In particular, the use of civilians to respond to noncrime calls for service offers the opportunity for both a departure from traditional patrol staffing and an improvement in the productivity of the police.

In March, 1974, Public Systems Evaluation, Inc., undertook a 6-month after-the-fact evaluation of the Worcester Crime Impact Program, whose main innovative feature was the use of some 41 Police Service Aides. This study demonstrated that for many police calls, a trained civilian can handle the work required at a lower cost without compromising the quality of service provided. In fact, the fraction of calls handled by civilians can approach the fraction of calls handled by uniformed police officers.

As part of the evaluation effort, researchers were interested in whether or not the response times of Police Service Aides were acceptable to the public. As in the two Kansas City studies discussed earlier, citizen satisfaction with response time was consistently high despite wide variations in response time. While this indicates that Worcester residents were happy with the attention they received, it also indicates that citizen satisfaction is not too dependent on response time.

The Wilmington Split-Force Experiment¹⁹

Yet another approach to the utilization of police resources was tested in Wilmington, Del. The concept of split-force, which involves separating the preventive patrol and call-for-service response functions of the patrol force, was first tried in St. Louis in 1966 and in Chicago in 1971. An LEAA-funded test of the split-force approach was conducted in Wilmington in 1975, and again, Public Systems Evaluation, Inc., performed the evaluation of this program.

The results of this study showed that split-force patrol does increase the efficiency of the patrol force in both the call-for-service response function and the preventive patrol function. This improvement was achieved without decreased effectiveness on the part of the patrol force. Thus, this program was able to demonstrate the feasibility of split-force patrol as a costeffective alternative to traditional patrol strategies.

Another interesting result from this study was in the area of response time, where it was again found that response time had no effect on citizen satisfaction. In addition to this, a limited number of citizens were formally informed that a response delay would occur, and 45 percent of these citizens responded that they "couldn't care less". Hence, the Wilmington study supports the contention that police response time is not an important factor contributing to citizen satisfaction with the police.

Patrol Staffing in San Diego: One- or Two-Officer Units²⁰

When considering the allocation and deployment of police responses, a seemingly simple decision such as whether to staff a patrol unit with one or two officers can give rise to rather complex implications. Traditionally, many police departments have staffed their cars with two officers, primarily for reasons of safety. Those departments which have used one-officer patrol (such as the Crime Control Team in Syracuse) have usually exchanged two one-officer units for each two-officer unit.

In October, 1975, the Police Foundation began an experiment in San Diego where only one one-officer unit was substituted for each two-officer unit. It was found that, in terms of response time, onscene arrests, officer injuries, and other performance measures, one-officer units performed at a satisfactory level of efficiency and effectiveness. The cost implications of this drastic reduction in manpower are clear; if one wished to retain equal-cost staffing options, the potential for increased productivity through the use of one-officer patrol is very great.²¹ Thus, a simple switch in staffing policy may have significant implications towards patrol productivity.

To summarize, the major hypotheses tested in the last ten years of patrol research may be grouped into three major categories:

- (1) Hypotheses which postulate the merit of preventive patrol;
- (2) Hypotheses which postulate the merit of rapid police response;

- (3) Hypotheses which postulate the merit of alternative manpower allocation schemes and other productivity issues.

While research into some of these hypotheses has proved to be somewhat inconclusive (e.g., it has yet to be shown in a convincing manner that preventive patrol does or does not deter crime), there has been a consensus of research findings in certain areas of inquiry. Most notable is the finding that citizen satisfaction with the police is not affected by police response time; this result was reached by four studies which employed different methodologies and were designed for different purposes. It is also apparent that alternatives to traditional patrol practices (as illustrated in the Syracuse, Worcester, and Wilmington studies) can be implemented both efficiently and effectively in modern police departments. We are hopeful that with further research in the patrol area, a consensus of results governing the majority of patrol hypotheses will begin to surface.

A REVIEW OF KEY INVESTIGATION STUDIES

Unlike patrol, the criminal investigation function has not historically been viewed as a police responsibility. It has been said that Marcus Lucinius Crassus of early Rome used citizens to gather evidence for use in law courts,²² and it was not until well into the 19th century that criminal investigations formally entered the domain of police activities.²³

With police assumption of investigative responsibilities a myth quickly developed: that crime could only be fought effectively by former criminals.²⁴ Consequently, many of the early detectives were, in fact, former criminals. The belief was so strong that Parisian officials, in an innovative attempt to cope with a rising crime rate at the turn of the 18th century, appointed Eugene Vidocq (a former criminal) to head their new security organization.

Coupled with this belief was another which suggested that investigators ought to mingle with criminals in order to learn their habits and obtain information which would aid in the solution of crimes. Unfortunately, this association with known criminals led to the demise of many an investigative unit. For example, the London Bow Street Runners, or "Thief Takers," severely tarnished its reputation by a willingness to associate with known criminals. In some instances, these civilian investigators even committed criminal acts.

" Recognizing the deficiency of investigative units like the Bow Street Runners, Sir Robert Peel introduced a proposal in Parliament in 1829 calling for a formal police service. His pro-

posal accepted, Peel was appointed to head the newly created force. Although London's first police force had been formed, it did not become apparent until 1842 that uniformed policemen could not adequately perform both the patrol and investigative functions. Hence, 12 officers were assigned to a plainclothes unit and housed at Whitehall Place--also known as Scotland Yard--and specialization had, in essence, occurred for the first time in a formal law enforcement organization.

The use of such a specialized unit in the United States commenced 7 years later when the Boston Police Department assigned 3 of its 33-man force as detectives. Since these men were assigned to the police headquarters, the first detective function in the U.S. was a centralized one. (Still, the theory prevailed that detectives must be closely associated with known criminals in order to perform their duties adequately.)

As detective units expanded, so too did the number of scholarly attempts to examine various criminal investigation topics. Examples include a dissertation by Hans Gross on the impact of psychology on crime in his book Criminal Psychology; Leonhaid Felix Fuld's discussion of the administrative functions of police and specific units within police departments in Police Administration; Elmer Graper's examination of detective service in his book American Police Administration; and Charles O'Hara's classic text, Criminal Investigation, which analyzes criminal investigation techniques.

Many of these writings only perpetuated the myths associated with investigative work. One such erroneous assumption presumed that the solution of a case was a function of a detective's qualifications. Even today many unconfirmed suppositions are espoused as gospel by "experts" in the criminal investigation area. For example, a widely used contemporary text claims that the percentages of case clearances by arrest and the percentage and amount of property recoveries, reflect closely the ability and effort of the investigating officers.

Scientific Criminal Investigation

Interwoven with the development of the field of criminal investigation is the use of scientific techniques to assist in the solution of crimes. Simpson (1969) describes the advancement of scientific techniques applied to criminal investigation as evolving in three overlapping stages:²⁵

- (1) the Medico-Legal Era
- (2) the Criminological Era
- (3) the Scientific Era

The Medico-Legal Era saw an increase in the use of expert medical advice to determine the cause of death or injury in criminal cases and the progressive development of forensic medicine. The Criminological Era produced a series of scales for the identification and description of evidence, but the major accomplishments lie in the development of techniques for classification and identification of physical evidence.²⁴ Examples include Bertillonage--identifying criminals by comparing their physical measurements--"Rouge's" portrait gallery, and the development of fingerprints. Finally, the Scientific Era was characterized by the application of scientific techniques of measurement and analysis to physical evidence.

Until recently little attention has been given to the conduct of experiments or quasi-experiments in the area of criminal investigation field services. As a result, the old myths and unfounded assumptions about investigative practices perpetuated literally for centuries. The first significant attempt to dispel the aura surrounding the investigative function was made in 1967 by the President's Commission on Law Enforcement and Administration of Justice. The Task Force examined the interface between patrol and investigation, suggesting, specifically, that "a faster response time leads to more arrests." Coupled with a reduced response time was the recommendation that patrol officers conduct more thorough preliminary investigations at the scene in order to obtain more information for follow-up investigations by detectives. The researchers noted that few followups were made by detectives in cases where information was lacking.²⁶

The Task Force report went on to suggest that due to scarce patrol resources, civilians be used to form a preliminary investigative squad; that computer-aided search capability be developed and implemented to facilitate the searching of latent fingerprints; that detective forces should consider employing more field resources for dealing with burglaries; and that police departments modify their criteria for emergency dispatching, i.e., prioritize calls.

Perhaps the most important long-term contribution of the Task Force report was not its questioning of investigative procedures, but rather its catalytic effect on future investigation research. Of particular interest to the researchers was the improvement of police effectiveness. To this end, the Task Force recommendations were to:

- Develop a fundamental research and analysis program over a multi-year time period in one or more large metropolitan police department in order to explore in more detail the characteristics of crime, arrests, and field and investigative practices
- Develop computer-aided search systems

- Develop special data collection methods in order to obtain information on field practices²⁷
- Undertake experiments to improve statistical procedures for manpower allocation²⁸

As in the patrol area the Task Force report called for controlled experimentation. The Task Force researchers commented on the type of study needed:

The study would have to be of an iterative (step-by-step) nature. It would begin with a series of hypotheses based on the results of this study, and for which data would be specifically collected. As analysis is undertaken, new hypotheses would be developed and would also have to be tested and special data collected for those purposes. It is clear that studies of this kind in various departments are essential if the police are to gain their factually supported understanding of ways to improve effectiveness.²⁹

As in patrol, during the last decade researchers have initiated several major studies in an attempt to clarify the issues raised by the Crime Commission. However, unlike patrol, the area of criminal investigation has not been explored with great vigor. This has been due, in part, to the availability of more readily quantifiable patrol performance measures, and, perhaps, to a common perception of patrol as the core of police field service operations. When investigation issues are explored, survey research has been the most prevalent examination technique.

What follows is a review of selected research projects performed since the 1967 Crime Commission. While these reports are but a sample of all research efforts in the investigative area, their findings are representative of what has been learned.

Municipal Detective Systems--a Quantitative Approach³⁰

In 1971, Joseph Folk, a graduate student at the Massachusetts Institute of Technology, published his findings concerning alternatives in case assignment policies. Folk's research employed a computer simulation model which incorporated a "minimum service-time" criterion, "interrupt" matrices, and case priorities.

Folk reported:

- (1) that eyewitnesses were the most frequent and valuable sources of evidence;

- (2) that Boston detectives had caseloads too heavy to allow for detailed case-solving procedures;
- (3) that Boston district detectives were being assigned to many special details, thus reducing the investigative manpower;
- (4) that the supervisor's estimate of minimum investigation times was unrealistic;
- (5) that mug shots, eyewitnesses, or victim's friends and relatives can be helpful in the investigation of certain Section A crimes (classified by Folk as homicide, rape, robbery) and in less than 20 percent of Section B crimes (classified by Folk as burglary, larceny, auto theft, etc.);
- (6) that there was a correlation between case closing and seriousness of crime; and
- (7) that physical evidence was seldom found in Boston.

Folk acknowledges that his effort is a first-cut and by definition also parochial in nature. Despite these admitted shortcomings, Folk's effort was, and still remains, one of the very few mathematically-based investigative research efforts.

**The Investigative Function:
Criminal Investigation in the United States³¹**

While Folk was conducting his research in Boston, Richard Ward was preparing his doctoral dissertation (at the University of California at Berkeley) which was completed in 1971. Unlike Folk, who reviewed operations in only one police department, Ward expanded his research to include several. Comparing detective operations in the New York City and San Francisco Police Departments, he set out to determine:

- (1) if there were any differences in investigative effectiveness between these two departments, and
- (2) whether or not investigative effectiveness was dependent upon detective organization (generalist vs. specialist).

His sample for the latter comparison consisted of 21 police departments which employ one or the other style of organization. Ward, a former New York detective, found that upon examination of burglary and robbery clearance rates there were no significant differences between New York and San Francisco generalist and

specialist detectives. However, in his comparison of the larger grouping of agencies, he concluded that "generalist or specialist detectives have no impact upon robbery arrests. . .but those departments using generalists had greater success in the burglary arrest category than those using burglary specialists."³²

Like Folk, Ward also found detectives to be overburdened with cases, and a correlation between the seriousness of the crime and the likelihood of case-closing. In addition, he cited a high physical evidence recovery rate and suggests that the processing limitations of the crime laboratory limit its value. Raising serious questions about the efficacy of arrest and clearance rates as a measure of detective productivity and efficiency, Ward cites a lack of uniform standards among police departments regarding these rates. Some departments clear a case based upon suspect-related information, others clear it upon arrest, and still others only upon conviction for the offense.

Because his work is frequently cited in the investigation literature, one must look beyond Ward's own caution to the reader and examine his methods carefully. (Ward cites many methodological problems which include: lack of adequate records, agency non-cooperation; lack of relevant data; absence of related studies; use of a comparative systems approach; size of the police department; and the year of the study.) Under close scrutiny there appear to be some questions about the validity of his conclusions. However, the data needed to verify his computations are not available in text.

Enhancement of the Investigative Function³³

In 1973, Bernard Greenbert et al. of the Stanford Research Institute completed a major research effort and produced what is now known as the SRI solvability model. In essence, this model helps police officers identify those burglary cases which are most likely to be solved through further investigation.

Focusing upon the type of information obtained by the police officer after arriving at the scene, it was found that during the officer's preliminary investigation certain types of information about the crime itself (e.g., a vehicle license number) was found to be of greater value than a description of the stolen item.

Upon close examination, the SRI researchers were able to develop a burglary investigative checklist, which aided officers and police managers to decide which cases would most likely be solved. Under their system each piece of information is assigned a numerical value, based upon its importance. For example, usable fingerprints are assigned a higher value than a description of stolen property. After totaling the values assigned to all the information obtained at the crime scene, a case with a number

exceeding a prescribed threshold is pursued; a case with a number below the threshold is not considered likely to be solved.

The results of the study have been adopted by many police departments in an effort to increase the productivity and efficiency of the investigative unit. In most cases the implementing departments have been successful in focusing greater attention upon the cases which have a higher probability of solution. Other researchers have replicated this study and further replications are currently underway to verify, refine, and add precision to the model.

The Criminal Investigation Process³⁴

The Rand Corporation conducted the most comprehensive investigation research to date. Under an LEAA grant the research team carried out a 2-year study during which they surveyed approximately 150 law enforcement agencies, and conducted interviews and observations in 29 departments.

Publishing its report in 1975, Rand confirmed the results of other evaluative research and also broke new ground. Selected findings reported:

- (1) that in over one-half of cleared cases, the identification of the offender was supplied to the reporting officer by the victim (supporting a Task Force finding);
- (2) that most cases receive cursory attention (if there are few leads, little followup will occur);
- (3) that investigators choose cases using the strength of leads and seriousness of the crimes as selection criteria;
- (4) that more evidence--especially latent fingerprints--was being collected than could be effectively used;
- (5) that the structure in which police investigators are organized--specialist vs. generalist--had no correlation with crime, arrest, or clearance rates (parallels Ward's finding);
- (6) that investigative strike forces possess the potential to increase arrest rates for a few target crimes; and
- (7) that differences in investigative training, staffing, workload, and procedures have little effect on crime, arrest, or clearance rates.

Comprehensive in scope, the Rand study had many outcomes. One, in particular, was reduction in the "mythology" which hitherto had surrounded the detective. For example, the finding that in more than half of the cleared cases the identification of the offender was supplied to the police all but erases the image of the detective as a "super sleuth." The inevitable controversy surrounding the report's findings resulted in a large volume of criticism--some competent and some irresponsible in character. Legitimate criticism addressed study design problems, largely attributable to the enormous scale of the study. Of equal concern were some results which appeared to be based upon limited data.

To their credit the Rand team listened to the constructive criticism of its report. When appropriate they defended their results; in other cases, simply acknowledged their peers. The outcome, however, was a text, published in 1977, restating their findings. It addressed the earlier criticisms and in most cases clarified questionable results.

Detective Work³⁵

In 1977, sociologist William Sanders published a text based on his study of a California Sheriff's Department. In most cases, his work supported the findings of others.

For example, Sanders study confirmed that information supplied to the police officer about the crime is significant. In fact, he notes two rules which seemed operative when it came to screening cases. First, the "important case rule" stipulates that only cases deemed important by investigators were investigated. Second, the "workable case rule" implied that where no leads appeared to exist, cases would not be pursued.

Sanders also found a correlation between case closing and seriousness of the crime; that little could be done with physical evidence because of processing limitations; that various specialized investigative units produce higher clearance rates on target crimes (e.g., homicide, burglary, robbery); and that certain types of investigations require unique methods of investigation (e.g., arson), hence specialized training is required.

Varieties of Police Behavior³⁶

James Q. Wilson began his research in 1965, first published his results in 1968, and his book is now in its seventh printing. Always an influential figure in shaping criminal justice policy, Wilson's study characterized police departments according to their operational modalities.

The typology is comprised of three descriptors:

- (1) The watchman style which emphasizes maintenance of public order;
- (2) The legalistic style which promotes the law enforcement aspect of the police role;
- (3) The service style of policing in which police respond to the community as if it were the "market" and they the producers of a "product," which meets the demand of the community.

In Wilson's view, a police department's style will affect its procedures, hence oftentimes influencing its arrest or clearance rates. The notion of different styles of policing has important consequences for the involvement of patrolmen in the investigative process. It seems likely that in a legalistic police department arrests and clearance rates may be more important than in a watchman-style department. As Wilson notes, "a watchman-like department is interested in avoiding trouble and in minding its own business.... The pre-disposition to avoid involvement--to control (not eliminate) public disorder rather than to enforce the law--depends not on corrupt motives, but on the inclinations of the men recruited into police work and the norms of the organization to which they belong."³⁷

The legalistic department, according to Wilson, ". . . tries to get the men to work harder, confident in some areas (for example, vice suppression) and hopeful in others (for example, preventive patrol, field interrogation, juvenile citations, traffic summons) that will achieve a desired objective."³⁸ As Wilson notes, in this style, there has been an attempt made by some police chiefs to erase the long-standing separation between the patrol forces and the detective force, so that men could be assigned from one area to another.

Patterns of Metropolitan Policing³⁹

In 1977 Elinor Ostrom et al. released a study of the organization of metropolitan police departments in 80 Standard Metropolitan Statistical Areas (SMSA's) across the country. While most of their report does not impact greatly upon the investigative function, portions of it do support other investigation evaluative research studies. Closely related to Wilson's perception of a service-style of policing, Ostrom et al. view police departments in an industrial framework--a concept of producers (police) and consumers (users of police services).

The researchers noted, in general, that detectives were overworked; observed a high recovery rate of physical evidence that

was seldom used due to the inability of crime labs to process it; perceived that investigators are assigned differently depending upon geographical areas, and that such assignments may impact upon productivity and costeffectiveness (one- vs. two-man detective car); and that "more than 80 percent of the 'producers' of burglary investigation in the 80 SMSA's (Standard Metropolitan Statistical Areas) assign police generalists to investigate residential burglary."⁴⁰

Hypotheses Tested

In presenting a representative review of investigation research and evaluation in the past decade, the intent has been to acquaint the reader with the recurring issues addressed in the conduct of the studies; to convey a sense of the historical evolution of the research and evaluation; and to familiarize the reader with the kinds of conclusions drawn as a result. To summarize in hypothesis-specific terms, the last 10⁶ years have examined:

- (1) Hypotheses which postulate the merit of case-closing and -clearance methods;
- (2) Hypotheses which postulate the merit of the collection and use of evidence;
- (3) Hypotheses which postulate the merit of alternative manpower allocation schemes and other productivity improvement schemes.

CONCLUSIONS AND FUTURE CONSIDERATIONS

While the state of the art of police field services research and evaluation is well represented in the foregoing sections, it remains to focus upon the collective findings of these studies in an attempt to discern what has been learned. Two significant issues are evident. First, using the experiment as a research tool has engendered large investments of time and money to examine relatively few hypotheses. Second, reasonably conclusive evidence has been generated for only one or two of these hypotheses. As an example, the patrol hypotheses that "response time is inversely related to citizen satisfaction"--one of the more popular issues examined--has pretty much been disproved. In its place several studies have suggested that the difference between observed and expected response time is the major satisfaction determinant. Yet, these "conclusive" results are virtually singular in that very few additional hypotheses have been adequately tested.

Part of the difficulty may lie in the paucity of replication activity. If field services evaluative research is known by the

hypotheses it scrutinizes, we have too many disjoint studies. It is inappropriate to experiment only with previous untested hypotheses. Rather than exhibiting planned overlap in the hypotheses considered for investigation, it seems almost coincidental when studies address common questions. For instance, the following is a list of 10 substantive hypotheses:

- Increasing patrol intensity increases the number of crimes intercepted in progress
- Response time is inversely related to citizen satisfaction
- Response time is inversely related to apprehension probability
- Response time is inversely related to the number of officers in patrol
- Automatic Vehicle Monitoring reduces response time
- Automatic Vehicle Monitoring increases officer safety
- One-officer patrol is more efficient than two-officer patrol
- One-officer patrol is characterized by a higher workload than two-officer patrol
- Two-officer patrol is safer than one-officer patrol
- "Specialized" patrol is more productive than "routine" patrol

On the average each hypothesis has been examined by only a few major studies--not enough to yield conclusive results. Most remain as "dangling hypotheses" yet to be tested.

The experience of other disciplines is valuable here. In fields such as education and psychology it is not uncommon to find a large volume of studies testing essentially identical hypotheses. Given recently developed rigorous procedures for collective analysis of the results of many such independent studies,⁴¹ it would be advantageous if more studies examining equivalent hypotheses were available. At the very least, this would permit a possible consensus of results to surface.

When compared with elaborate field experimentation, secondary analysis is a far less expensive approach. By secondary analysis is meant manipulation of original source data in order to test, develop, or refine hypotheses that were either not analyzed or

inaequately analyzed in the original study. One can identify and specify extant data bases which afford the best opportunity for secondary analyses leading to: deeper probing into the jurisdiction--specific or project-specific questions addressed in the original study; and, comparative analyses across jurisdictions and agencies. The authors have identified several such secondary analyses which reanalyzed prior evaluative research using different statistical methods to test the original hypothesis.⁴²

Having viewed evaluative research on investigation in a hypothesis-based framework, the authors perceive a need for the conduct of investigation studies which utilize an experimental approach. To date, most investigation studies have involved analysis of operations as they were practiced in the police department(s) at the time the study was conducted. One could reasonably alter traditional investigative practices in order to test one or more specific hypotheses. For example, the relationship between the quality of the investigative process and crime analysis or report screening/enhancement activities merits experimental examination. Such experimentation would also shed needed light on the patrol/investigation interface. Too little is known about the relationship between the two functions in terms of their shared responsibilities for essential field service tasks. A larger aggregate body of experimental results in investigation would result in a more complete systemic view of people field services and, perhaps, improved methods of criminal apprehension and crime prevention.

A number of trends attributable to the prevailing austere economic condition have been reflected in the body of field services evaluative research. One such trend is the rapidly increasing attention being given to field service resource allocation by police agencies, with the objective of providing adequate service with fewer personnel and less material--i.e., improving productivity. To assist in meeting this challenge, police decisionmakers are turning to recently developed resource allocation tools, including some sophisticated computer-based models. This may be a sign that an enthusiasm for creative uses of technology is supplanting the technological "negativism" of the past several years. Another cost-based trend is the departure from crime-reduction methods of accountability. In its place we find that police agencies consider, for example, that patrol operations are cost-effective so long as the dollars allocated to patrol and the local crime rate remain at or below some established threshold.

While organizational in nature, there exists a diversity of operational constraints which impact upon the conduct of police field service activities--and are discussed in field service studies. Among them are union-imposed constraints which can, for example, impinge upon patrol flexibility by requiring two-officer staffing. Similarly, legal constraints have, in the past, estab-

lished that a jurisdiction may not deviate from a statutory patrol schedule. Of course tradition--the severest constraint of all--often finds police departments employing the same field service strategies for decades.

The process of evaluation of field service innovations is itself undergoing discernable changes. In particular, the use of structural models (described above) is growing rapidly. This is largely attributable to the dimensions of generalizability and insight regarding causality which model-based evaluations add. In fact, recent evaluative research solicitations issued by the National Institute of Law Enforcement and Criminal Justice have incorporated modeling as a design requirement.⁴³

The current state of knowledge in police field services is significantly advanced from that of 1967, due in large part to what has been learned from the conduct of a number of highly significant research and evaluation studies. Despite the inevitable controversy surrounding study findings, occasional contradictory results or technical inaccessibility--causing some practitioner, evaluator, and researcher confusion--there can be little doubt that evaluation and research have found acceptability in and contributed to the police practitioner community. The task before us is to establish what is known, through a process of assessment and synthesis; enhance that current state of knowledge through replication, secondary analysis, and other research and evaluation activities; and extend the state of knowledge by establishing and implementing a well-thought-out agenda for future evaluative research.

NOTES

1 Institute for Defense Analyses, Task Force Report: Science and Technology. A Report to the President's Commission on Law Enforcement and Administration of Justice (Washington, 1967).

2 David J. Farmer, "Fact versus Fact: A Selective View of Police Research in the United States," The Police Journal, 49, 2 (April-June 1976):104.

3 Paul D. Leedy, Practical Research Planning and Design (New York: Macmillan, 1974).

4 Robert K. Yin and Karen A. Heald, "Using the Case Survey Method to Analyze Policy Studies," Administrative Science Quarterly, 20 (September 1975):371-381.

5 Samuel G. Chapman, ed., Police Patrol Readings (Springfield, Ill.: Charles C. Thomas, second edition, second printing, 1972) p. 4.

6 O. W. Wilson and Roy Clinton McClaren, Police Administration (New York: McGraw-Hill, 1972) p. 320.

7 Chapman, Readings; see also Suzanne Young, "Historical Development of Selected Police Performance Measures," in Richard C. Larson, ed., Police Accountability: Performance Measures and Unionism (Lexington, Mass.: Lexington Books, 1978).

8 Young, previous note, pp. 6-9.

9 Task Force Report, note 1 above, p. 12.

10 Ibid.

11 J. F. Elliott and Thomas J. Sardino, Crime Control Team: An Experiment in Municipal Police Department Management and Operations (Springfield, Ill.: Charles C. Thomas, 1971).

12 George L. Kelling et al., The Kansas City Preventive Patrol Experiment: A Technical Report (Washington: Police Foundation, 1974)

13 See Richard C. Larson, "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment," Journal of Criminal Justice, 3 (1976):267-297; Edward M. Davis and Lyle Knowles, "A Critique of the Report: An Evaluation of the Kansas City Preventive Patrol Experiment," Police Chief, 42, 6 (1975); Stephen E. Fienberg, Kinley Larntz, and Albert J. Reiss, Jr., "Redesigning the Kansas City Preventive Patrol Experiment," Evaluation, 3, 1-2 (1976):124-131; Franklin E. Zimring, "Field Experiments in General Deterrence: Preferring the Tortoise to the Hare," Evaluation, 3, 1-2 (1976): 132-135.

14 Larson, previous note, p. 291

15 Tony Pate et al., Police Response Time: Its Determinants and Effects (Washington: Police Foundation, 1976).

16 Kansas City Police Department, Response Time Analysis: Executive Summary: Volume I, Methodology; Volume II, Analysis (Kansas City, Mo.: 1977).

17 Richard C. Larson, Kent W. Colton, and Gilbert C. Larson, Evaluation of a Police Implemented AVM System: Phase I with Recommendations for Other Cities (Cambridge, Mass.: Public Systems Evaluation, Inc., 1977); Gilbert C. Larson and James W. Simon, Evaluation of a Police AVM System: A Phase II City-Wide Implementation (Cambridge, Mass.: Public Systems Evaluation, Inc., 1978).

18 James M. Tien et al., An Evaluation Report of the Worcester Crime Impact Program, Vols. I and II, (Cambridge, Mass.: Public Systems Evaluation, Inc., 1975.)

19 James M. Tien, James W. Simon, and Richard C. Larson, An Alternative Approach to Police Patrol: The Wilmington Split-Force Experiment (Cambridge, Mass.: Public Systems Evaluation, Inc., 1977).

20 John E. Boydston, Michael E. Sherry, and Nicholas P. Moelter, Patrol Staffing in San Diego: One- or Two-Officer Units (Washington: Police Foundation, 1977).

21 Edward H. Kaplan, "Evaluating the Effectiveness of One- versus Two-Officer Patrol Units," forthcoming in Journal of Criminal Justice.

22 Richard H. Ward, "The Investigative Function: Criminal Investigation in the United States," (diss., University of California at Berkeley, 1971) p. 24.

23 Antony E. Simpson, Criminal Investigation: A Selective Review of the Literature (New Haven, Conn.: South Central Criminal Justice Advisory Board, 1976) p. 24.

24 Ward, "Investigative Function," p. 4.

25 Simpson, Criminal Investigation, p. 24.

26 Task Force Report, p. 101.

27 Ibid.

28 Ibid, p. xiv.

29 Ibid, p. 101.

30 Joseph Frederick Folk, Municipal Detective Systems: A Quantitative Approach (Cambridge, Mass.: Massachusetts Institute of Technology, 1971).

31 Ward, "Investigative Function."

32 Ibid., p. 231.

33 Bernard Greenberg, Oliver S. Yu, and Karen I. Lang, Enhancement of the Investigative Function, Volume I: Analysis and Conclusions, Final Report--Phase I (Menlo Park, Calif.: Stanford Research Institute, 1973).

34 Peter W. Greenwood, Jan M. Chaiken, and Joan Persilia, The Criminal Investigation Process (Lexington, Mass.: D. C. Heath and Co., 1977).

35 William B. Sanders, Detective Work: A Study of Criminal Behavior (New York: The Free Press, 1977).

36 James Q. Wilson, Varieties of Police Behavior (Cambridge, Mass.: Harvard University Press, 1978).

37 Ibid., p. 148.

38 Ibid., p. 185.

39 Elinor Ostrom, Roger B. Parks, and Gordon P. Whitaker, Patterns of Metropolitan Policing (Cambridge, Mass.: Ballinger Publishing Co., 1978).

40 Ibid, p. 142.

41 See Richard J. Light and Paul V. Smith, "Accumulating Evidence: Procedures for Resolving Contradictions Among Different Research Studies," Harvard Educational Review, v. 41 (1971); Robert Rosenthal, "Combining Results of Independent Studies," Psychological Bulletin, 85, 1 (1978).

42 See, for example, Thomas R. Willemain, "Bayesian Analysis of Crime Rate Changes in Before-and-After Experiments" (Cambridge, Mass.: Massachusetts Institute of Technology Operations Research Working Paper OR 075-78, 1978).

43 See, for example, Criminal Justice Research Solicitation: Research on Incapacitation (Washington: National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1979).

Police Reorganization A Review of Its Evaluation*

Roger B. Parks is associate director of the Workshop in Political Theory and Policy Analysis at Bloomington, Indiana.

For many years reform-oriented government commissions, students of public administration, and members of good government groups have worried that there were far too many police agencies in the United States and, in particular, too many small police agencies. These small agencies have been charged with being ineffective in supplying service to their own jurisdictions and with contributing negatively to the effectiveness of policing across broader jurisdictions through their fragmenting effects on delivery of police services.¹

A variety of structural reforms have been proposed to alleviate this perceived defect in American policing. Total consolidation of all police agencies in a county, a metropolitan area, a region, or even statewide has often been recommended as the most desirable solution. Second-choice solutions for the "problem" include partial consolidations through smaller scale mergers (particularly of agencies with less than 10 officers), interlocal service agreements or service contracting, consolidated production of particular sub-services in policing such as dispatch, training, and criminal laboratory facilities, and enhanced coordination and cooperation through the use of formal mutual-aid pacts and other binding intergovernmental or interagency agreements.²

Those reform recommendations have been based on principles of "good government" and on management and organization theories rather than on fact-based evaluations of the associated costs and benefits. None of the major reform statements contains empirical data bearing on the effectiveness or the efficiency of police agencies of differing size, nor do they contain data that might be used to demonstrate improvement in services through organizational change. Fragmented metropolitan areas are damned through a count of the number of separate agencies operating, while consolidated or partially consolidated areas are claimed to be successes on the basis of statements by the administrators responsible for implementing the reforms.³

More recently there has been a growing concern that changes might be necessary in the organization of the very large police agencies found in major cities. Critics have argued that these large agencies have become overspecialized, overprofessionalized, and more and more

* This review was prepared with partial support from the National Science Foundation through Grant No. GI-43949. The opinions expressed are of course the author's own and do not necessarily reflect the views of the Foundation or of the Workshop in Political Theory and Policy Analysis.

Copyright (c) 1979 by Roger B. Parks

removed from the citizens in their jurisdictions. Recommendations have included decentralization through precinct and lower level (storefront) police stations, increased usage of generalist police officers rather than specialists for many duties, and various forms of geographic-based team policing, with small groups of officers assigned responsibilities for policing specific subareas of the jurisdiction.⁴

THE SETTING FOR POTENTIAL EVALUATION

In recent years there have been several large-scale consolidations of all or most police agencies in a given area, several instances of consolidation of two or more police agencies, and many instances of functional consolidations for the purpose of producing one or more services for otherwise separate police agencies. Examples of recent large-scale consolidations of agencies included those undertaken in Nashville-Davidson County, Tennessee; Jacksonville-Duval County, Florida; Lexington-Fayette County, Kentucky; and Las Vegas-Clark County, Nevada. Smaller scale consolidations of police agencies have occurred in Northern York County, Pennsylvania; Roseau County, Minnesota; and Ward County, North Dakota.

Functional consolidations can be found as joint dispatch centers, joint training facilities, joint detention facilities, multiagency crime labs, narcotics task forces, and "major case squads" in many parts of the country. These consolidations have been particularly prevalent in arrangements for auxiliary and staff services. In a recent study of policing in 80 Standard Metropolitan Statistical Areas (SMSA) across the United States, 68 percent of municipal police agencies were found to produce their own radio communications and 7 percent their own entry-level training. Those agencies not producing their own services received them under a variety of interorganizational arrangements often including contracting with a county or interjurisdictional communications center, or participation in a regional training academy.⁵

Cooperative arrangements also have proliferated. Formal mutual-aid pacts are common, and informal cooperative exchanges of services are even more so. Fifty-one percent of the municipal police agencies in the sample of SMSA's referred to above participated in a formal mutual-aid pact. More than 90 percent of the consolidations of two or more police agencies, and in many instances of municipal agencies, reported that they provided and received aid from surrounding police agencies at various times, whether members of formal pacts or not.⁶

Cooperative exchanges often follow establishment of joint auxiliary service production, particularly joint dispatch arrange-

ments that can facilitate backup by adjacent police agencies in times of peak demand. Contractual arrangements for service provision have also become common. These often involve the supply of full or partial services by a county sheriff or police agency to municipalities within the county, but also include city to city and even city to county contracting.

Organizational changes consistent with decentralizing recommendations have also occurred. Most common have been variants of team policing. Well-known examples include those of the Los Angeles Police Department, the Syracuse Crime Control Team, Albany's Neighborhood Policing, and the Cincinnati Community Security (COMSEC) experiment. Precinct, district, or sector decentralization (in the sense of emphasizing the importance of these subunits in a large organization) has occurred to some extent in New York City; St. Louis, Mo.; Rochester, N. Y.; and other large cities. The use of store-front offices further to decentralize police-citizen contact points has become common in many cities. Concurrent with team policing has been a move toward a view of the patrol officer as a police generalist and the deemphasizing of specialization in some police departments. One additional form of decentralization has been the continued formation of new police agencies in growing suburban and rural areas as new communities incorporate and small communities grow to demand more than the undifferentiated services of State or county law enforcement agencies.

ORGANIZATIONAL CHANGES AS EXPERIMENTS

It is tempting to a social scientist to think of these organizational changes as "natural experiments." They are not the sorts of clearly delineated manipulations of one or a few variables, with other variables held constant or randomized, that are characteristic of the "true experiments" available to our colleagues in some of the sciences. But, reasoning by analogy, we argue that they are similar in kind if not in detail. If we were able to sharpen our methodological skills and obtain sufficient funds, we could use these "reforms as experiments," much as chemists use bottles of reagents and laboratory glassware, holding many variables constant and examining the effects of a change in the experimental variable, organizational form. This analogizing constitutes the major support for evaluation research.

Additional elements of evaluation research are drawn from systems analytic approaches to problem-solving. Goals and objectives should be stated in unambiguous terms. Alternative programs for achieving those goals and objectives can then be advanced, and the more promising programs tested "in the field," in communities willing to adopt them. Evaluators can then collect measures of

effectiveness to determine how close actual results come to meeting stated goals and objectives, and determine the costs involved. As different communities articulate their goals and objectives and try new programs to achieve them, we accumulate a body of data to show what works and what doesn't in addressing social problems.

EVALUATION RESULTS

Authors of this volume were asked to provide a "state-of-the-art review of program evaluation results" in the following specific areas of criminal justice: "(1) analysis of program impacts, (2) details and critiques of methodologies, and (3) suggestions for future research directions." In light of the many organizational changes in policing, it might seem likely that one could report substantial results from evaluation research. Many natural experiments have been conducted. Some goals and objectives have been delineated, alternative programs have been explicated, and some implemented. Evaluators have been on the scene to monitor the effects. Many documents labeled evaluation reports have been produced. Yet, in response to the question "What do we know with respect to organizational and interorganizational arrangements as a result of evaluation research?" one is forced to reply, "Not much."

For our purposes, "evaluation" was defined as "the measurement of the effectiveness of criminal justice programs and projects. Evaluation studies attempt to determine whether a program has accomplished its objectives directly, in relation to the particular problem the program addresses, and relatively, in comparison with other criminal justice efforts dealing with the same problem. Evaluation attempts to establish a causal relationship between a program's activities and the perceived results of those activities." In review of the holdings of the National Criminal Justice Reference Service and other literature resources available to the author, it was not possible to identify any evaluation studies that succeeded in meeting these criteria. Before exploring causes for this consistent failure and the author's suggestions for an alternate approach that may have some promise, a review of evaluation efforts in areas of organizational and interorganizational arrangements will provide some illustrations. The review is limited to evaluations of consolidations, both total and those limited to specific functional areas.

LARGE-SCALE CONSOLIDATION OF POLICE AGENCIES

The best documented of recent consolidations is that of the Jacksonville Police Department and the Duval County, Florida, Sheriff's Department, which occurred 10 years ago. Through the efforts of Koepsell-Girard and Associates, Inc., an extensive set of reports entitled Consolidation of Police Services Case Study: Jacksonville, Florida was prepared.⁸ These documents provide a description of the problems in the Jacksonville area that were used as the bases for the consolidation effort, the tactics employed successfully to sell the consolidation to the legislators and voters (whose approval was needed), and the organizational implementation of consolidation in the newly-formed Office of the Sheriff. A number of statistical series for the first 4 years of operation by the consolidation department are also provided.

The findings that are potentially relevant to an evaluation of the consolidation are as follows:

- (A) The cost of policing in Jacksonville increased following consolidation--approximately 8 percent in absolute dollars (adjusted for inflation) and 6 percent in per capita terms (again adjusted for inflation) when compared to preconsolidation expenditures;
- (B) Crime and crime rates increased dramatically following consolidation, with some reduction from the increased rates occurring in the last year studied (1972);
- (C) The number of clearances rose 29 percent in the 4 years following consolidation; the clearance rate first declined and then increased slightly over this time;
- (D) Arrests per 1,000 Part I offenses rose 58 percent in the 4 years after consolidation; arrests per sworn officer first increased, then decreased, leaving a net increase of 15 percent;
- (E) Traffic-related activity indicators increased substantially, around 100 percent in total, and nearly 70 percent per sworn officer;
- (F) Data stated in per-dollar terms closely parallel those in per-officer terms, though they do show slightly smaller changes due to the increases in salaries in the department over this period.

These findings, unfortunately, do not allow evaluation of the effects of consolidation in Jacksonville. Although expenditures increased, we cannot say whether this increase was larger than what might have otherwise occurred. The Sheriff's department was understaffed prior to consolidation in comparison to other agencies of similar size. Pressure to increase staff (and thus expenditures) could have resulted without consolidation. Jacksonville Police Department officers received lower salaries at given ranks than did Sheriff's officers prior to consolidation. Even without consolidation, pressure to increase these salaries might have arisen. The changes in crime and crime rates shown in the report are dominated by a change in definition: larcenies of any amount are included for the years following consolidation, only those over \$50 in estimated value were included in Part I totals prior to consolidation. The same pattern of a substantial increase in crime, followed by a decrease in the last year for which data were available, can be seen when an estimate for the number of larcenies of \$50 or less is subtracted from the figures provided. The data, however, are less dramatic. There is no attempt to estimate what crime figures might have been had consolidation not occurred.

The data on clearances, arrests, and traffic activity are unusable for evaluation purposes because they are only available for the years after consolidation. Furthermore, no attempt is made to estimate what they might have been without consolidation. The traffic figures show such a large change that one assumes a change in policy was instituted within the consolidated department toward emphasizing traffic activity. Of course, such a policy could have been adopted by the two departments in the absence of consolidation. The changes in clearances and arrests over the time studied may be bound up in the inclusion of lower value larcenies in the definition of Part I crimes (thus including a large number of shoplifting arrests by store detectives and the like), but there are no data presented to examine this possibility.

The reports do indicate that salaries were increased and facilities improved in the consolidated department. This could account for the approval of the consolidation by the officers involved (the reports indicate 85 percent approval among 142 officers interviewed in 1973). But these changes, too, might have been accomplished without consolidation. There is an indication that citizens of the Jacksonville area perceived an improvement in services, with about one-third of 300 interviewed saying that services had improved, versus 8 percent who felt services had deteriorated since consolidation, but again no comparison is provided.

Taken in sum, the Koepsell-Girard reports provide a wealth of useful information regarding the Jacksonville consolidation.

They are particularly useful for any area attempting to sell, and subsequently implement, a consolidation of police services. The major shortcoming from the perspective of this review is that they do not tell whether consolidation led to more or less effective police service in Jacksonville, or whether services have been provided more or less efficiently following consolidation. Still, they represent a mountain of information in comparison with that found bearing on other large-scale consolidations.

The data on the Nashville-Davidson County, Tennessee, police consolidation are marked by their absence. The report "How Well Is Metro Doing?", prepared by Nashville-Davidson County's Department of Finance (with help from the Urban Institute), provides data for the year 1974 for use as a "benchmark" for comparison in later years.⁹

No attempt is made to compare pre-Metro performance. As the authors of the report note, "...it is difficult to determine, in many cases, what the data means [sic], since there are no standards of comparison." One comparative study of Metro policing in a small area of the County to policing provided by a small, adjacent, independent community found citizens in the independent community much better served than those in the Metro area.¹⁰ However, no data were available to address whether police consolidation through Metro was a causal factor in this finding.

No reports, evaluative or descriptive, were found with respect to recent police consolidations in Lexington, Ky., or Los Vegas, Nev.

SMALLER-SCALE CONSOLIDATIONS OF AGENCIES

One report was found that dealt with smaller-scale consolidation of police agencies. This report, "An Evaluation of the Northern York County Regional Police Department," provided data to compare the consolidated department to other police agencies serving cities with similar populations.¹¹ The department compared quite favorably in terms of crime rates and was about average with respect to clearance rates. Unfortunately, there was no attempt to show that the comparison groups were comparable on any variable other than size. The disparity in crime rates probably indicates that the consolidated department is serving a quieter, less crime-prone jurisdiction than many of the comparison cities rather than preventing crime more successfully. In essence, the report is a management study similar to many done by the report's authors and other consulting firms in the police field. Although useful to the department itself, it does not contribute to an understanding of the costs and benefits of consolidation.

FUNCTIONAL, SERVICE-SPECIFIC CONSOLIDATIONS

Of all reports reviewed for this chapter, the Exemplary Project report on the Muskegon (Mich.) Central Police Dispatch appeared, on first reading,¹² the closest to a useful evaluation report. It documents per service unit cost reductions obtained through the operation of a joint dispatch operation for eight police jurisdictions in Muskegon County, including all of the large police agencies. Using cost and volume figures from 5 years of operation, the report demonstrates that the joint center was able to keep costs per service unit low in the face of rising public and police demands for communications services. It appears to show that the primary objective of the program--"to afford a more efficient and economical means of transmitting information from and to law enforcement officers..."--was achieved.

Further reading reveals that, although the report is useful for documenting the adoption and implementation of what most observers agree is a successful functional consolidation (including a good discussion of implementation problems), it fails as an evaluation study. No baseline data are provided relating to costs and service volumes in other, compatible areas served by different communications arrangements. Although the authors conclude that the centralized system saved money for the cooperating jurisdictions, one could argue that the individual agencies involved might have switched from sworn to civilian operators and obtained better equipment while maintaining their own communications centers, thus capturing most of the economies documented. The present author agrees with the authors of the report that it is reasonable to attribute much of the cost-saving to the centralized operation, but this conclusion rests more on the technological logic of the reform than on any data presented.

Other reports reviewed in this area provided even less evaluative content.¹³ Several focused on joint drug enforcement squads, but none was able to show whether the squads affected the drug traffic in their areas. One set of authors, writing on their attempt at evaluating the Allegheny Regional Narcotic Task Force, noted several limits on evaluation efforts:

Evaluation efforts are uniformly limited by (1) the unavailability of baseline data, (2) the lack of or a poorly organized record system, (3) the absence of a management information system related to goals and objectives and their measurable criteria; (4) a subconscious, if not conscious, opposition of project managers to evaluation--partly due to a fear of being measured by often limited and not especially relevant quantitative data, and, lastly, (5) limitations in the funding allocated to evaluation.¹⁴

These same limitations were evident, if not always articulated, in other reports of narcotics task forces, and appear generally applicable to all evaluation reports available for this review.

This literature review is embarrassing in its brevity, yet reflects an interesting phenomenon. In spite of many recommendations for structural changes in the organizational and interorganizational arrangements for policing in America, there is no evidence based on careful evaluations of organizational changes that can be brought to bear on this issue. This statement is supported by the findings of the National Sheriff's Association extensive review of literature bearing on contract law enforcement and police consolidation:

In large measure, the reported assessments of the efficacy or the utility of a particular program were found to be anecdotal or impressionistic evaluations with few offers of valid measurement criteria or other information upon which sound evaluative judgements could be made.¹⁵

One of the prominent proponents of consolidation in the criminal justice field, Dan Skoler, has recently reached a similar conclusion:

...it can be fairly said that the nation is being urged to proceed with major structural reform based on professional perceptions of deficiencies and problems that need to be addressed rather than research-validated data on the new solutions proposed.¹⁶

As my colleague, Elinor Ostrom, has argued, it is possible for reformers operating with such a dearth of empirical evidence on the effects of their reforms to be very righteous, yet, simultaneously, dead wrong with respect to the presumed consequences of adopting their recommendations.¹⁷

COMPARATIVE RESEARCH WITH RESPECT TO POLICE AGENCY SIZE

By stretching the definition of evaluation provided to the authors, it is possible to include a very short review of recent works on the comparative effectiveness and efficiency of police agencies of differing sizes in delivering police services to residential portions of several metropolitan areas. The review is included to show that we do have some evidence on organizational arrangements and their relationships to performance, if not evidence derived from rigorous evaluation designs.

These studies shared a common research design, a similar systems approach.¹⁸ Neighborhoods served by police agencies of differing size were otherwise carefully matched with respect to the socioeconomic status of residents, including income, racial, housing tenure, and age characteristics. They were also matched on levels of commercial or industrial activity (generally quite low), traffic flows, and other police relevant service conditions. Data were then collected from one or more sources on the quantity and quality of police services delivered to the matched neighborhoods. The data collection method common to all these studies was a survey of citizens residing in the neighborhoods, with questions bearing on their experiences with crime and their local police response if called with respect to a victimization; their experiences with the local police in other contacts, such as assistances or stops; their perceptions of the level of police activity in their neighborhoods (frequency of patrol and/or rapidity of response when called); their perceptions of the demeanor and behavior of the officers assigned to their neighborhoods; and their overall evaluations of services received. Other data collection methods used in one or more of these studies included comparisons of agency recorded data on crimes and service demands, interviews with the officers serving in the neighborhoods, and use of records maintained by prosecutor and circuit attorney offices.

These studies, conducted by several different (but related) research groups in a number of diverse sites, had common, if surprising findings (surprising at least to advocates of eliminating small police agencies). In no case were large agencies (those in excess of 300 sworn officers up to and including departments with more than 12,000 officers) found to be outperforming small- and medium-sized agencies serving comparable neighborhoods. This finding held whether comparing agencies serving white, middle-class neighborhoods in the Indianapolis, Grand Rapids, and St. Louis metropolitan areas; bluecollar white neighborhoods in the Nashville area; or black, poor- or middle-class neighborhoods in the Chicago and St. Louis areas. In all of these studies, citizens residing in similar neighborhoods served by police agencies that were small to medium in size reported higher rates of assistance by their police, whether in instances of victimization or non-crime-related matters. They reported that their police responded more rapidly when called and took more, and more favorable, actions once on the scene. They perceived a higher level of police presence in their neighborhoods (a perception supported by analyses of police deployment patterns). They perceived local police officers as generally more fair, courteous, and honest than did citizens in the matched neighborhoods served by the larger departments. Moreover, they rated the police services they received overall more highly than did citizens served by larger departments. Although there were wide variations among these

studies in the levels of experiences, perceptions, and evaluations of the interviewed, those served by small- and medium-sized departments consistently reported performance that was as good, and often was better than, that reported by those residing in similar neighborhoods served by the larger agencies. In the studies where additional data were used to complement the interview data, similar trends were found with respect to agency-maintained data on crime and service calls, interviewed officers' perceptions of the quality of service they were delivering, and the success of agencies of varying size in obtaining warrant issuances from local prosecutors' offices.

A number of analytic efforts are currently underway based on these studies and extensions. These efforts are aimed at developing explanations for the consistent findings, demonstrating how the small- and medium-sized agencies were able to supply more effective services to comparable areas.¹⁹ One part of the explanation appears to lie in the production strategies chosen by agencies of differing size, with larger agencies generally tending toward a task-oriented strategy emphasizing the use of many specialized units, while smaller agencies employ a patrol-oriented strategy, with the bulk of their sworn personnel deployed for onstreet duty.²⁰ Data on personnel utilization in nearly 800 municipal police agencies show, for example, that this difference in choice of deployment results in a reversal of the relationship between size of agency and officers per 1,000 citizens served. Looking at the total number of officers employed, one sees that large departments have a larger number of officers in relation to population served than do smaller agencies. But, when the comparison is limited to those officers deployed for onstreet patrol duties, the smaller agencies have the advantage.²¹ This appears to explain, to a substantial degree, the findings of more rapid response by the small- and medium-sized agencies: They have a larger number of response units available in similar circumstances. Other linkages are undoubtedly operable in addition to that of production strategy, and current research efforts are aimed at finding and documenting these.

EVALUATION AND COMPARATIVE RESEARCH COMPARED

In evaluation research, one focuses on the implementation of a specific new program and asks whether or not it is successful in accomplishing a more or less articulated set of objectives. This forces one to openly entertain the possibility that the program may not be successful. Comparative research allows a different thrust. Here the focus is on a number of different, existing programs aimed at dealing with a similar set of problems. One can accept that all of the programs are

dealing with the problems at some level of effectiveness (evidenced by the fact that they have been in existence for a period of time), and examine the extent to which some operate more effectively or efficiently than others. This may appear to be a semantic difference only; in practice, however, it may explain why we have been able to engage in comparative studies of the performance of differently organized police agencies, but evaluations of major organizational changes have not occurred.

Why Evaluation Doesn't Happen

The prototypical evaluation model is the true experiment. Although social scientists often are tempted to view reforms as experiments, it is safe to assume that virtually no participants in organizational changes view them as such. The chemist does not need to worry about how his reagents feel toward an experiment, but the attitudes of participants in organization changes are crucial. Participants may act to prevent or drastically alter a given change program, or may act to prevent or predetermine the results of any evaluative effort. While reagents cannot be said to have motives that would lead them to take such actions, human participants clearly do.

Organizational changes involve changes in power relationships among participants. In a consolidation of police agencies, for example, some level of local control must be relinquished. It is likely that a consolidated agency will allocate resources differently than had been the case among the agencies entering into the consolidation. Although it is often presented as a technical matter subject to decisionmaking by rational police administrators, this allocation of "who gets what, when, and how" is the essence of politics.²²

The political nature of organizational changes is illustrated by the extensive political campaigns preceding successful consolidations. To quote Mayor Briley with respect to the Nashville consolidation:

...ours was a very heavy program...a matter of making hundreds of speeches by different people at civic clubs, the men's clubs, in the schools, at the PTAs, the League of Women Voters, every group we could get to...I think that to really sell consolidation you have got to have a really hard sell campaign.²³

Koepsell-Girard Associates found politics at work in the Jacksonville consolidation.

The democratic form of government has certain inherent characteristics which frustrated Jacksonville's consolidation architects and permitted the antagonists of reform to take steps to block the ultimate merger...though logic and tenets of good government undergirded the proposed charter, "practical politics" and its resultant trade-offs and compromises played an important role in Jacksonville's successful vote to consolidate.²⁴

They discuss with approval, however, the massive media campaign used to obtain voter acceptance of consolidation. Campbell spoke directly to this issue and its impact on the conduct of evaluative research:

In the present political climate, reformers and administrators achieve their precarious permission to innovate by overpromising the certain efficacy of their new programs. This traps them so that they cannot afford to risk learning that the programs were not effective.²⁵

Those who have recently engaged in a hard-sell campaign for an organizational change cannot then be expected to assist an objective evaluation of the effects of the change. We might wish for persons to fit an objective, scientific model of behavior, but our intuition tells us this is unlikely.

A second reason for the failure to evaluate organizational change is the difficulty in untangling what changes occurred so as to examine their effects. Thus, consolidations in Nashville and Jacksonville involved a large number of changes in the structure of the police agencies and massive media campaigns arguing the benefits of consolidation. Hawthorne-like effects seem most likely in these environments.²⁶ A very careful research program, based on a well-developed theory of police organization and performance, would be required to successfully disaggregate the multiple causes and multiple effects in such a situation. As Kornfeld and Riordan noted, the funds for such a careful effort are generally unavailable. Certainly we have nothing resembling the required theoretical understanding at present.

Why Comparison May Work

Comparative research does not pose a direct test of a particular program. The comparative researcher generally seeks

to compare alternate programs that have been in existence for sufficient time to have reached a "steady state" level of performance. He or she hopes that no innovative programs (or any other transient-producing inputs) will occur during the course of data collection. Comparative data collection often includes measures bearing on a number of different outputs, outcomes, or both, and the activities believed relevant to their production, rather than on a more narrowly defined set of objectives for a single program. In this way, comparative research appears more useful for studying "broad-aim" programs, such as policing, than experimental research designs focused on particular innovations and their immediate effects.

Comparative research may be able to avoid some of the deliberate attempts to affect (or obscure) findings commonly cited in evaluation studies. Since comparative studies focus on organizations that have been successful at surviving over time, organizational representatives can be more relaxed about the research. The findings are unlikely to directly affect their tenure or that of their agency. Thus, one might expect greater openness and cooperation in the production of necessary data for a comparative study of existing organizations than for an evaluative study of a new one.

Understanding Policing and Evaluating Police Programs

Evaluative research is fundamentally applied research. Applied research thrives in an environment where basic underlying theory has been developed to a large extent. Evaluating police programs as a form of applied research requires, then, that we develop a theory of policing. In the particular area of organizational arrangements for policing, we require a theory that explains how differences in police organization might be expected to effect differences in police performance.

In his 1960 Presidential Address to the American Statistical Association, Rensis Likert made a fundamental distinction between two functions of statistics that is relevant to this discussion.²⁷ Likert argued that we needed two kinds of information in order to address problems: information about the nature of the problem, and information about the state of the problem.

By information as to the state of the system let us mean the statistical measurements which reveal the current situation of the nation or the economy....By information as to the nature of the system let us mean the basic conceptual model or understanding which serves as a guide to tell what dimensions of the nation, or society, or economy should be measured and how these measurements should be interpreted in making decisions. This information as to the nature of the system in-

cludes, of course, both the conceptualizations themselves and the extensive, quantitative measurements which are required for valid conceptualizations.

Clearly, information about the nature of a system should precede useful information about the state of a system. Likert found, however, that "we are doing a far better job of collecting information about the state of our nation than we are of obtaining data dealing with the nature of our nation and developing valid generalizations and theories based on these data." Yet, without an understanding of the nature of a problem, what influences bear on it, and how it influences other aspects of a system, measures of the state of a problem (or measures of a change in the state of a problem) are most difficult to interpret.

Scholars in the field of evaluation research have noted this difficulty. Rutman, for example, argues that one of the preconditions for evaluation of a program is a "linking rationale," an explanation of why and how the program is expected to work to alter the state of the problem.²⁸ Roos refers to a similar concept when she speaks of the "degree of obtainable knowledge."²⁹ In essence she says, "Do we know enough about the nature of the system to be able to specify relevant variables and adequate operational measures, and to specify research hypotheses as to their relationships?" If linking rationales cannot be specified in some detail, research is clearly necessary as to the nature of the system, what Rutman refers to as formative research. A low level of obtainable knowledge is another way of pointing to the same difficulty and necessity.

A failure to recognize and act on Likert's distinction appears, to this author, to be a fundamental stumbling block to useful evaluations of criminal justice programs. Evaluation efforts may collect voluminous data bearing on the state of the problem or problems that a given program is intended to address without benefit of the essential underpinning of the nature of the problem and its dynamics with respect to the program. The level of obtainable knowledge is low with respect to policing. Conceptual models are only slowly being developed in the detail necessary for an understanding of the nature of police service delivery and how it impinges on citizen-consumers.³⁰ Until these models are further elaborated and tested in diverse contexts, it is presumptuous to assume that we know enough to conduct program evaluations. Even if it were possible to overcome the difficulties posed to evaluation by political concerns, and adequate fiscal resources and competent evaluators were available, the lack of a basic, theoretical understanding of the dynamics of policing would undermine the evaluation effort.

Clearly the time, money, and effort currently expended for evaluation ought to be redirected. What is needed is simultaneously more theorizing about the nature of policing and much careful observation of ongoing police programs to inform and test that theorizing. That is, we should shift our dollars from measuring the state of the system to the attempt to better understand the nature of the system.

A shift toward research into the basic nature of policing might be done in such a way as to contribute to ongoing action programs as well. Rather than studying an action program with the aim of evaluating whether it was or was not successful, one can study it with the aim of documenting what happened, thus contributing to a better understanding of the ongoing policing process.³¹ By removing the burden of rating the success of the program, this approach could help remove some of the obstacles to obtaining useful information.

Program participants might be less reluctant to allow a researcher access to the internal operations of the program if the researcher were there to observe and try to understand what was happening than if the researcher were there to pass judgment on the program's effectiveness (and longevity). The participation of the researcher in the implementation of a program might then be seen as beneficial, bringing an outside point of view to bear on the problem at hand and simultaneously bringing the researcher into closer contact with the reality of the program.

A key element in a shift toward research into the nature of policing would be a heavy focus on comparative studies. One way to view the many diverse arrangements for the delivery of police services in the United States is as a series of experiments in police organization. The availability of police agencies of widely differing sizes serving comparable areas should be looked on as a great resource for investigating the effects of size (as well as other organizational characteristics). Where evaluative research on the effects of changes in organizational structure may be too infrequent, or too bound by the inherent difficulties of evaluating complicated (political) programs, quiet comparisons of existing organizations has begun to, and may continue to, provide additional insight to our understanding of how policing works.

CONCLUSIONS

In spite of substantial rhetoric on the perceived need to restructure American police organization, and a number of organizational reforms along the lines recommended, no information deriving from evaluation research could be found with respect to the effectiveness or efficiency of the reforms. None of the consolidations of police agencies, officers, or both, or smaller

efforts jointly to produce particular services has been subjected to any effective evaluation. It is clear that advocates of major structural changes in American policing base their arguments on tenets of good government and public administration theory rather than empirically grounded research findings. At the same time, findings from comparative studies of existing police organization cause one seriously to question the efficacy of the reform recommendations.

Evaluation research, it is argued, is an inappropriate model for generating information with respect to the effects of alternative forms of organizing police. Given the difficulties encountered when attempting to evaluate new programs, particularly those that alter political power relationships, comparative research on existing organizational forms may be more productive. Without a developed theory of the linkages between variables characterizing police organization and those of police performance, evaluation efforts cannot succeed. Research efforts that monitor the effects of organizational changes, asking "what happened" rather than "were goals achieved," can contribute to the necessity understanding of the nature of policing.³³ Comparative studies of police organizations are essential to this task as well.

NOTES

1 National Commission on Law Observance and Enforcement (Wickersham Commission), Report on Police, (Washington: U.S. Government Printing Office, 1931); President's Commission on Law Enforcement and Administration of Justice, Task Force Report: The Police, (Washington: U.S. Government Printing Office, 1967); Advisory Commission on Intergovernmental Relations, State-Local Relations in The Criminal Justice System, (Washington: U.S. Government Printing Office, 1971); National Advisory Commission on Criminal Justice Standards and Goals, Report on Police, (Washington: U.S. Government Printing Office, 1973); Committee for Economic Development, Reducing Crime and Assuring Justice, (New York: Committee for Economic Development, 1972); Donal E. J. McNamara, "American Police Administration at Midcentury," Public Administration Review, 10:181-190 (Summer, 1952); Gordon E. Misner, "Recent Developments in Metropolitan Law Enforcement," Journal of Criminal Law, Criminology 50:497-508, 51:265-272 (January-February 1960 and July-August 1960); Daniel L. Skoler and June M. Hetler, "Governmental Restructuring and Criminal Administration: The Challenge of Consolidation," Georgetown Law Journal, 58:719-741 (1970).

2 In addition to recommendations contained in the reports listed in footnote 1, see David L. Norrgard, Regional Law Enforcement, (Chicago: Public Administration Service, 1967). For a recent statement from a State perspective, see New Jersey County and Municipal Government Study Commission, Aspects of Law Enforcement in New Jersey, (Trenton, N.J.: County and Municipal Government Study Commission, 1976).

3 See for example, President's Commission on Law Enforcement and Administration of Justice, (1967 note 1), pp. 98-99 and 94.

4 The President's Commission and the National Advisory Commission, supra, both speak to these concerns.

5 Elinor Ostrom, Roger B. Parks, and Gordon P. Whitaker, Patterns of Metropolitan Policing, (Cambridge, Mass.: Ballinger Publishing, 1978), chapters 10 and 12.

6 Ibid., chapter 15.

7 With apologies to Donald Campbell, who was certainly aware that this was a not always attainable objective. Donald T. Campbell, "Reforms as Experiments," The American Psychologist, 24:409-429 (April, 1969).

8 Koepsell-Girard and Associates, Inc., Consolidation of Police Services Case Study: Jacksonville, Florida, (Falls Church, Va.: Koepsell-Girard and Associates, Inc., 1973). These reports include an Executive Summary, a full report, and two volumes of Research Papers.

9 Metropolitan Government of Nashville-Davidson County, Department of Finance, How Well Is Metro Doing? (Nashville, Tenn.: Metropolitan Government of Nashville-Davidson County, n.d.).

10 Bruce D. Rogers and C. McCurdy Lipsey, "Metropolitan Reform: Citizen Evaluations of Performances in Nashville-Davidson County, Tennessee." Publius, 4:19-34 (Fall, 1974).

11 Public Administration Service, An Evaluation of the Northern York County Regional Police Department, (Chicago: Public Administration Service, 1974).

12 John J. McDonnell. An Exemplary Project: Central Police Dispatch. (Washington: U.S. Government Printing Office, n.d.).

13 Reports reviewed included James Lynch, Evaluation of the Statewide Enforcement Coordinating Committee, (Hartford, Conn.: Connecticut Planning Commission on Criminal Administration, n.d.); Allen Alexander, The Investigative Strike Force, (Wilmington, Del.: Delaware Agency to Reduce Crime, 1975); Jack P. Kornfeld and John D. Riordan, Evaluation Study of Allegheny Regional Narcotic Task Force, (Chicago, : IIT Research Institute, 1974); Abt Associates, Case Incident Regional Reporting System, (Cambridge, Mass.: Abt Associates, 1976).

14 Jack P. Kornfeld and John D. Riordan, op cit., p. 51.

15 Robert R. Delahunt, Richard D. Engler, and Susan B. Petinga, An Evaluation Study in the Area of Contract Law Enforcement: A Review of the Literature, (Washington: National Sheriff's Association, n.d.), p. 197.

16 Daniel L. Skoler, Organizing the Non-System, (Lexington, Mass.: Lexington Books, 1977), p. 80.

17 Elinor Ostrom. "On Righteousness, Evidence, and Reform: The Police Story." Urban Affairs Quarterly, 10:464-486.

18 A more extensive review is available in Gordon P. Whitaker, "Does Structure Make A Difference in Police Performance?", a paper presented at the 1978 Annual Meeting of the American Political Science Association, New York. Studies in this series include the following: Elinor Ostrom, et al., Community Organization and the Provision of Police Services, Administrative and Policy Series 03-001. (Beverly Hills, Calif.: Sage Publications, 1973); Elinor Ostrom, Roger B. Parks, and Gordon P. Whitaker, "Do We Really Want To Consolidate Urban Police Forces? A Reappraisal of Some Old Assertions," Public Administration Review, 33:432-433 (September/October, 1973); Elinor Ostrom and Gordon P. Whitaker, "Community Control and Governmental Responsiveness: The Case of Police in Black Communities," in David Rogers and Willis Hawley, ed., Improving the Quality of Urban Management (Beverly Hills, Calif.: Sage Publications, 1974); Roger B. Parks, "Complementary Measures of Police Performance," in Kenneth M. Dolbeare, ed., Public Policy Evaluation, (Beverly Hills, Calif.: Sage Publications, 1975); Samir T. IsHak, "Consumer's Perception of Police Performance: Consolidation versus Deconcentration: The Case of Grand Rapids, Michigan Metropolitan Area," Diss. (Bloomington, Ind.: Indiana University, 1972); Elinor Ostrom and Roger B. Parks, "Suburban Police Departments: Too Many and Too Small?", in Louis H. Masotti and Jeffrey K. Hadden, ed., The Urbanization of the Suburbs. (Beverly Hills, Calif.: Sage Publications, 1973); Bruce D. Rogers and C. McCurdy Lipsey, (note 10).

19 Early efforts include Gordon P. Whitaker, "The Advantages of Small Police Departments," a paper presented at the 1978 Annual Conference of the International Association of Chiefs of Police, New York; Roger B. Parks, "Linking Organization and Performance Empirically: The Case of Response Time," Police Services Study Technical Report T-56. (Bloomington, Ind.: Workshop in Political Theory and Policy Analysis, 1978); Elinor Ostrom, et al., "The Public Service Production Process: A Framework for Analyzing Police Services." Policy Studies Journal (forthcoming).

20 Ostrom, Parks, and Whitaker, "Do We Really Want To Consolidate . . ." note 18.

21 Ostrom, Parks, and Whitaker, Patterns of Metropolitan Policing, note 5.

22 Harold D. Lasswell, Politics: Who Gets What, When, How, ((New York: Meridian Press, 1958). See also David L. Norrgard, "Intergovernmental Police Service Programs: Some Myths," Minnesota Municipalities (May 1974), pp. 4-5.

23 Anon., "Seminar: City-County Consolidation," The American County, 37:10-18, 59 (February 1972).

24 Koepsell-Girard and Associates, Inc., Consolidation of Police Services Case Study, note 8, Executive Summary, 6.

25 Donald T. Campbell, "Considering the Case Against Experimental Evaluations of Social Innovations," Administrative Science Quarterly, 15:110-113 (March, 1970), 111.

26 "Hawthorne effect" refers to phenomena noticed in the course of the pathbreaking experiments with respect to work situations and worker morale and efficiency conducted at the Western Electric Hawthorne plant from 1927 through 1932. Among other interesting findings, the experimenters noted that merely being the focus of some attention in an experimental situation had a tendency to alter behavior. Thus, workers in control groups, whose objective work conditions were unchanged, changed in their morale and efficiency by virtue of the experimenters' interest in them as controls. The experimenters reported that it was impossible to "hold all else equal" in order to determine the independent effects of the experimental work condition variables. With these findings in a careful, low-key series of experiments in a sheltered, industrial setting, one should be prepared to find much larger "Hawthorne effects" in widely advertised reorganizations of police agencies. For a description of the Hawthorne studies, see F. J. Roethlisberger and William J. Dickson, Management and the Worker (Cambridge, Mass.: Harvard University Press, 1943).

27 Rensis Likert, "The Dual Function of Statistics," Journal of the American Statistical Association, 55:1-7.

28 Leonard Rutman, "Planning an Evaluation Study," in Leonard Rutman, ed. Evaluation Research Methods: A Basic Guide (Beverly Hills, Calif.: Sage Publications, 1977), pp. 13-38.

29 Noralous P. Roos, "Evaluation, Quasi-Experimentation, and Public Policy," in James A. Capraso and Leslie L. Ross, Jr., ed., Quasi-Experimental Approaches. (Evanston, Ill.: Northwestern University Press, 1973), pp. 281-304. See also Donald R. Weidman et al., Intensive Evaluation for Criminal Justice Planning Agencies, (Washington, 1975).

30 See Ostrom et al., note 19, for some early steps.

31 A useful model is the "monitoring methodology" developed by the Brookings Institution for its revenue-sharing studies. See Richard P. Nathan et al., Monitoring Revenue Sharing (Washington, D.C.: The Brookings Institution, 1975).

Police Personnel and Training Evaluation

William C. Cunningham is president of Hallcrest Systems, Inc. a research and consulting firm in law enforcement personnel evaluation and training programs. He served earlier as Director of Personnel for the St. Louis Police.

Todd H. Taylor is Director of Planning and Research for the Oregon State Police. He served in two other law enforcement agencies and also worked as a police program consultant and evaluator.

Perhaps the most consistent finding of management studies conducted for law enforcement agencies is that issues of human resources management are not sufficiently emphasized. Law enforcement agencies, historically, have been labor-intensive organizations, allocating as much as 90 percent of their operating budgets to personnel-related expenditures. Personnel represents the most important resource in police organizations, with all other resources directed toward supporting police personnel on an individual and an organizational level--most notably the activities of patrol officers and investigators in the field. Law enforcement administrators, however, do not generally view the process of personnel administration and training as an integrated series of management techniques for maximization of human resource effectiveness, efficiency, and potential. In the authors' opinion, this is due, in part, to the fragmentary approach to police human resources evaluation.

The police evaluator and manager must assimilate numerous technical and nontechnical evaluations on a wide variety of topics from journals, magazines, books, and project reports, acquired through a haphazard system of collection and dissemination. The police practitioner must assess whether a particular program, technique, or approach can be transferred or adapted to a particular department with its unique environment and operating characteristics. Police administrators should be aware that implementation of a certain improvement or change in one aspect of personnel management and training can affect a number of other areas. They also need to know about the research upon which program and project implementations are based and how research findings can be translated into a workable and viable improvement in the operational environment. The authors' review of numerous evaluations discloses that most evaluations do not pursue the relevance of findings to other related areas; nor do they note, within the findings, material that is new or that contradicts previous findings.

In short, we do not find that evaluation content and dissemination are responsive to the needs of police researchers or managers nor to needs of funding agencies in choosing how to allocate resources. We have chosen to comment on the quality, quantity, and types of evaluations from available sources, and

Copyright (c) 1979 William C. Cunningham and Todd H. Taylor

to limit substantive discussion to those evaluations (and trends in evaluation procedures) that have obvious value for law enforcement researchers and administrators.

SEARCH FOR EVALUATIVE INFORMATION

The primary reference source for the authors was the evaluation data base assembled by the Evaluation Clearinghouse of the National Criminal Justice Reference Service (NCJRS). Of the more than 700 abstracts pertaining to police personnel and training reviewed, approximately 170 documents were selected for formal analysis because they appeared to actually report on evaluations of police human resource programs. This screening process was extremely difficult since the quality and content of the abstracts varied so greatly. As an example of the difficulty in screening the abstracts, Figure I lists the various terms used to describe the evaluation process.

In addition to the uneven quality of the NCJRS abstracts, the authors found that the extant literature on police personnel and training evaluation has, by no means, been completely catalogued by the NCJRS. This conclusion is based on a search of the authors' corporate and personal libraries, their participation in and supervision of at least 20 police personnel and training grant evaluations during the past 8 years, and their knowledge of other relevant personnel and training research sponsored by Law Enforcement Assistance Administration (LEAA) funds. The emphasis placed by LEAA and the State Planning Agencies (SPA) since 1970 on project grant evaluation has, in States like Virginia, California, Wisconsin, Michigan, and Florida, produced hundreds of evaluations. The NCJRS collection contains only a few of these evaluations, and it is difficult to determine why some projects are included in the Evaluation Clearinghouse and others are not.

The NCJRS, nevertheless, appears to be the most comprehensive source of evaluation documents available to practitioners, researchers, and evaluators. The commentary contained in this chapter has been gleaned largely from the sources in the NCJRS Evaluation Clearinghouse. Thus, while we draw on the most representative material available in a single source, we cannot make unequivocal statements that evaluation material does not exist in each of the topical areas to be reviewed under police personnel and training. In addition to emphasizing materials of obvious value to law enforcement researchers and policymakers, the authors provide substantive commentary in those topic areas where a preponderance of evaluation exists.

Figure I
EVALUATION TERMINOLOGY

- Evaluate
- Assess
- Review
- Study
- Test
- Measure
- Validate
- Explore
- Critique
- Examine
- Reexamine
- Analyze
- Observe
- Survey
- Rate
- Appraise
- Research
- Synthesize

QUALITY VS. QUANTITY

In the last 10 years tremendous growth has occurred in the amount and quality of police training and of research into police personnel administration. This is due largely to the efforts of the President's Commission on Law Enforcement and the Administration of Justice (1967), the Police Task Force of the National Advisory Commission on Criminal Justice Standards and Goals (1973), the availability of Federal funding through the LEAA, and the establishment of State-level commissions on Police Officer Standards and Training (POST).

The President's Commission urged the upgrading of police selection standards and training if law enforcement agencies desired to achieve true "professional" status, and the National Advisory Commission then recommended specific standards and goals to be attained. While in 1960 only the States of New York and California had minimum levels of State-mandated training for newly appointed law enforcement officers, by 1970 State commissions on police standards and training had been established in 33 States. A 1978 survey by the National Association of State Directors of Law Enforcement Training listed 47 States with such commissions, with 41 States mandating minimum personnel standards for entry-level training.

Evaluations of police personnel and training programs have had three distinct phases in relation to the impetus provided by these National and State commissions. The first phase received its impetus through use by State Planning Agencies of LEAA block grant funds for state-of-the-art assessments of personnel standards and delivery mechanisms for training programs. As the result of feasibility studies and needs assessments, training academies and programs, many of the POST's were established. These analyses are indicative of the majority of personnel and training "evaluations" conducted from 1969 through 1972. For the most part these efforts occurred at the State and local levels and were progress reports on efforts to attain recognized National and State-adopted personnel and training standards. While many of these standards had been almost universally adopted, there had not been any validation or evaluation of these standards when they were initially promulgated.

The second phase of evaluation applied formal evaluation techniques and designs in two major areas: (1) personnel recruitment, testing, selection, performance rating, promotion and career development, and (2) specialized training programs such as police-community relations, crisis intervention, drug abuse, and criminal investigations. Both areas received considerable national attention from LEAA's research arm, the National Institute of Law Enforcement and Criminal Justice (NILECJ).

The third phase, concurrent with the second, evolved with the requirement by many SPA's for a formal evaluation component as part of State and local action grant projects. The bulk of "true" evaluation occurred in these latter phases since they were concerned with the formulation of specific research hypotheses and/or the establishment of quantified relationships to standards, goals, and objectives of police improvement projects and innovations.

During each of these three phases of evaluation, a relatively large number of reports and journal articles (some of which were reviewed in the screening process) have appeared that could best be described as development projects. These projects frequently use evaluation techniques and statistical analyses, but essentially represent only initial research into a subject area. The projects usually lack definitive conclusions to guide police planners and managers in the feasibility and mechanics of implementation. There also exist numerous journal articles and project reports explaining how an agency implemented a particular program, but lacking an adequate summary of research findings and a quantifiable and objective assessment of impact or goal attainment. While much of this material we have classified as developmental, some of it is concise, well-written, and highly useful reference material for researchers and practitioners alike. It includes excellent statements of trends, issues, developments; summaries of research findings; and state of the art reviews. For example, in the personnel area, reports such as the following, while not "true" evaluations, are excellent contributions to the literature of the field:

- Performance Appraisal in Police Departments, by Frank J. Landy for the Police Foundation, 1977.
- Police Selection and Career Assessment, by Marvin Dunnette and Stephan Motowidlo, for NILECJ, Law Enforcement Assistance Administration, 1976.
- Psychological Testing for Police Selection, by Eugene Sullivan and Robert Cloward for Rhode Island Governor's Justice Commission, 1976.
- Police Personnel Administration, by the Police Foundation, 1974.
- Guidelines for Police Performance Appraisal, Promotion, and Placement Procedures, by Sidney Epstein and Richard Layman, NILECJ, Law Enforcement Assistance Administration, 1973.
- Development of Improved Examination Procedures for the Promotion of Police Officers in New York City, by Educational Testing Service for the City of New York, 1972.

Personnel and training should be viewed as integral parts in the management of police human resources. In Figure II, an outline of major topics is listed in a progressive sequence of components comprising human resources management. These topics reflect areas of specialization and concern for administrators, researchers, and evaluators. There are twice as many topics in personnel administration as in training, yet the preponderance of evaluation material available for screening was in the area of training. Both personnel and training evaluations are heavily skewed toward a few topic areas. For example, in this review, the number of personnel evaluations of screening techniques (psychological testing, physical agility examinations, etc.) was three times higher than those of the next highest categories: performance appraisal, affirmative action (employment of women and minorities), selection criteria, job analysis, and career development. In the other personnel categories, evaluations either do not exist in the NCJRS Evaluation Clearinghouse or are so few as to preclude meaningful commentary other than to note the voids.

Training evaluations are clustered around recruit training and assessments of training academy operations and training standards compliance. The next highest frequency of evaluations is found in specialized training and higher education. As noted above, this emphasis seems closely related to the evolution of LEAA funding, the National Advisory Commission, and the establishment of statewide POST commissions.

Figures III and IV are graphic representations of the relative frequency of evaluation techniques found in our review of 170 evaluation documents screened from the NCJRS. The size of each box corresponds to the relative frequency of evaluation techniques. While there is some overlap of techniques, generally it can be stated that different approaches to evaluation are used in personnel and training subjects--primarily because of the nature of the topics.

While Figures III and IV depict relative frequency of technique, it is significant to note that personnel evaluations of screening or selection processes tended to use more sophisticated statistical techniques (multiple regression analysis) and had more formal evaluation designs. Many evaluations in training, on the other hand, relied on measures of central tendency and dispersion and more basic measures of association (e.g. rank-order correlations). The following sections in this chapter will discuss study findings and, where appropriate, the evaluation techniques and methodologies utilized.

Figure II
COMPONENTS OF POLICE HUMAN RESOURCE MANAGEMENT

PERSONNEL

- Role Determination and Job Analysis
- Position Classification
- Compensation
- Human Resource Needs Forecasting
- Recruitment
- Affirmative Action
- Selection Criteria
- Screening Techniques
- Assignment of Personnel
- Performance Appraisal
- Promotion and Career Development
- Personnel Support Services
- Conduct and Discipline
- Labor Relations and Collective Bargaining
- Attrition
- Personnel Management Systems

TRAINING

- Training Needs Assessment
- Training Techniques and Delivery Mechanisms
- Recruit
- In-Service
- Supervisory and Management
- Specialized
- Education
- Training and Skills Inventory

**Figure III
PERSONNEL EVALUATION TECHNIQUES**

LITERATURE REVIEW	SURVEY QUESTIONNAIRE	VALIDATION STUDY
		ATTITUDE SURVEY
FIELD OBSERVATION	STRUCTURED INTERVIEW	PILOT TEST
		JOB ANALYSIS
	RECORDS EXAMINATION	PSYCHOLOGICAL TESTS
		PROFESSIONAL JUDGMENT

**Figure IV
TRAINING EVALUATION TECHNIQUES**

ASSESSMENTS						
ACADEMY OPERATIONS			TRAINING STANDARDS COMPLIANCE			
POST TEST	PARTICIPANT SURVEYS				STATISTICAL ANALYSIS	
	INSTRUCTOR RATINGS	COURSE CONTENT		JOB APPLICATION		
		ACCEPTANCE				
PRE-TEST POST-TEST						
OFFICER SELF REPORTING	PERFORMANCE RATINGS			CITIZEN SURVEYS	CONTROL GROUPS	LONGITUDINAL STUDIES
	FIELD OBSER- VATION	SUPER- VISOR RATING	AGENCY RECORDS EXAMINA- TION			

COMMENTARY ON PERSONNEL EVALUATIONS

Assignment of Personnel

State, county, and city law enforcement agencies employed more than 600,000 personnel in 1975; 80 percent of these were sworn employees, and 90 percent of them were classified as "primary operating positions" including patrol and investigation officers and line supervisors and managers.¹ In law enforcement agencies, sworn personnel are also assigned to a variety of direct and indirect support positions such as communications, booking, jail transport, forensics, records, identification, data processing, and other technical or administrative duties. As police agencies have increased in size in response to rising population and crime rates, the number of tasks beyond field operations requiring specialized knowledge and skills has correspondingly increased. These specialist positions have been filled by using both sworn and nonsworn (civilian) personnel.

The National Advisory Commission and several studies have addressed the issue of increased use of civilians in support positions. This use of civilian personnel reportedly lowers compensation costs, releases sworn personnel for operational tasks for which they are primarily hired and trained, and meets increased need for specialized skills not normally possessed by sworn police officers. The National Manpower Survey (NMS) reported that two-thirds of responding police chiefs and sheriffs anticipated increased use of civilian personnel, and projected that the number of civilian employees will increase annually at a rate of 4.0 percent compared with a 2.2 percent increase for sworn personnel between 1974 and 1985.² The survey also estimates that growth in support positions will increase during this period by 52 percent, contrasted with an overall growth rate of 33 percent for all personnel. Civilian positions will then account for 25 percent of all personnel employed in police agencies. Despite the emergence of this trend, the authors' review of evaluation materials disclosed only one evaluation of the role of civilians in police work (limited to communications, detention, identification, and community service affairs), and no evaluations concerned with the assignment of personnel in the police organization.

Forecasting, Recruitment, and Retention

With the imminent arrival of taxpayer revolts, evidenced by "Proposition 13"-type restrictions on statewide property taxes and the resultant public expenditure limitations, the police administrator will be faced with either maximizing the use of existing personnel or operating with fewer employees. Human resources management will become an important area of concern, yet no evaluations or substantive reference material were found that dealt with forecasting human resource needs in law enforcement agencies.

The NMS data indicated that about one-fourth of police departments serving populations of more than 17,000 reported an inadequate supply of qualified applicants and significant problems of voluntary resignation. No evaluation materials were uncovered in our search concerning attrition and turnover of personnel; only one evaluation (12 years ago) took a comprehensive look at recruitment and retention of police personnel.³

Affirmative Action

Exercising a profound impact on law enforcement in recent years has been affirmative action--requirements of the 1972 Equal Employment Opportunity (EEO) Act for full employment opportunities regardless of age, sex, and race considerations. In law enforcement this has meant increasing (sometimes under court order) the proportion of women and minority group members employed as sworn officers. EEO guidelines, the frequently cited Griggs v. Duke Power Company case, and other court decisions have required law enforcement administrators to remove race, sex, and cultural bias from selection and screening techniques. Agencies must also demonstrate the relationship of background characteristics, physical standards, and aptitude testing to requirements for successful job performance. In the evaluation material reviewed, the subject of minority recruitment is a large void, with only a handful of journal articles describing developmental programs, even though most large cities with substantial minority group populations have initiated program efforts. The few existing evaluations do not yield universal findings which can be readily adopted by police administrators into operational programs, nor do the evaluations lend themselves to adequate comparative analysis.

Until about 1970 female police officers had traditionally been relegated to support positions, to juvenile and matron duties, and occasionally to investigative functions such as sex crimes. Since that time, women have been hired in greater numbers and assigned to operational positions in patrol and investigations. Evaluations of the performance of women police on patrol have been sparse in number though exceedingly thorough in scope. While measurable differences have been found between men and women in certain aspects of patrol duties, evaluations in metropolitan police departments and in two State law enforcement agencies have clearly demonstrated that women can perform effectively on patrol.⁴

Much of the literature and evaluation pertaining to minorities and women in police work alludes to the enhanced quality of personnel and police service that results from a department's being responsive to and representative of demographic characteristics of the community served. Other benefits have been cited as well: one recent study suggests that increased proportions of minorities and women in law enforcement agencies can improve department efficiency and effectiveness. Correlation coefficients were obtained for higher rates of major crime apprehension in

police departments that include "larger proportions of females and minorities among its various ranks and positions."⁵ This finding is sure to evoke controversy among police administrators and certainly warrants further study and independent validation.

Selection Criteria

As noted above, affirmative action has required most police agencies to reevaluate both selection criteria and screening techniques for police officer applicants. Essentially, three dimensions are involved in both processes: physical characteristics, personality attributes, and background of the applicant. Reviewing the research findings of evaluations on the validity and reliability of these applicant dimensions as relevant qualifications for police service and as predictions of police officer performance would require an exhaustive account of methodologies and compatibility among findings. Therefore, we have elected to discuss the implication of evaluative findings for the police researchers and administrators concerned with validating and improving current selection criteria and screening techniques in their departments. The bulk of these studies address physical characteristics.

Most studies of physical qualifications have examined minimum height and maximum weights of applicants. Many law enforcement agencies have used minimum height standards of 68 and 69 inches with weight in some degree of proportion to height and frame. Some studies have reported higher rates of assaults, injuries, and accidents to officers between 68 and 70 inches, but no evaluations were found that empirically study the relationship of height to performance. Attempts to examine the relationship between heights lower than 68 inches and performance are often frustrated due to lack of sufficient officers of that height. The great variety of maximum weights among law enforcement agencies in itself raises doubts on the validity of a rigid maximum weight requirement.

Physical fitness specialists suggest that ratio of body fat is more relevant than height and weight alone, but the larger question is whether any of these criteria is job-related. Considerable attention has been focused on physical fitness or agility tests which measure physical skills utilized by police officers on the job, such as lifting and carrying injured persons, scaling barriers, chasing suspects, and so forth. Performance criteria for these tests are generally validated by using currently employed police officers. The employment of women and Hispanic minority group members has in large part caused some agencies to remove minimum height and weight requirements.

Screening Techniques

Virtually all law enforcement agencies now use some form of written examination, ranging from basic intelligence tests to development of personality profiles, to screen applicants. The following is a representative list of psychometric tests which have been subjected to evaluation or validation:

- Minnesota Multiphasic Personality Inventory (MMPI)
- Edwards Personal Preference Schedule (EPPS)
- Guilford-Zimmerman Temperament Survey
- OTIS Mental Ability Test
- Strong Vocational Interest Test
- Rorschach Psychodiagnostic
- Kuder Preference Record
- Army General Classification Test (AGCT)
- Wechsler Adult Intelligence Scale (WAIS)

These tests include such items as personality, vocational interests, cognitive abilities, and situational judgment. The most significant finding in these evaluations is that it is difficult to pinpoint precisely the relative contribution of selected personal attributes to recruit training or later field performance. Police officers are not a homogenous community, and the findings of one test validation which produces a predictive equation from multiple regression analysis cannot be universally adopted. Specific population norms must therefore be established and local validation conducted prior to implementing psychometric testing for police applicants.⁶

The authors conclude from the evaluations they studied that a battery of tests is preferable to any single test. Validation studies first must determine concurrent validity of the attributes being measured in existing field personnel and relate these desirable attributes to several measures of job performance. The adequacy of performance appraisal systems used in the validation process will greatly affect the reliability of resulting correlations or predictive questions obtained. In a pilot study of the predictive validity of a test battery and a subsequent longitudinal study on relationship to performance levels, it was found that "race-specific" equations will generally be more effective predictions because of differences related to race.⁷

Evaluations have demonstrated that entrance tests are both valid and invalid depending on the test batteries employed, quality of performance measures used, and cultural or racial bias present in the testing. The selection process can be viewed as a funnel which operates as a "redundant, cumulative measurement. The more stages it includes, the greater the attrition it is capable of producing."⁸

Predictive tests should be considered along with background investigations, training academy performance, and probationary field performance as a total selection and screening process. While more costly and time-consuming, this approach to personnel selection is "simply the price to pay for superior police personnel and public service."⁹

Performance Appraisal and Job Analysis

Some form of supervisory rating is used in nearly all law enforcement agencies to evaluate the performance of police officers. The following rating areas were commonly used in the performance appraisal system evaluations screened by the authors:

- Appearance
- Work quantity
- Work quality
- Dependability
- Attitude
- Attendance
- Disciplinary actions
- Interpersonal relations
- Job knowledge
- Discretionary judgment
- Departmental awards
- Dealing with the public

Personnel rating systems usually rely on some form of quantitative scale to assess the individual police officer in each of the performance areas. The most recent research has developed

rating areas based on functional job analysis techniques which examine components of the job, specific tasks performed, or critical incidents performed on the job.

Rating scales are subject to several supervisory rating errors or biases which must be reduced to minimal levels in order to assume validity of the rating: leniency, central tendency, and halo. Leniency errors occur when there are differences in the rigidity of supervisors doing the rating; that is, one supervisor might be strict or another lenient in the personal standards applied in the rating process. Central tendency errors occur when a significant majority of all officers are rated at the same level, thus making it difficult to determine an average performance level. Halo errors occur when the supervisor's overall judgment of the individual influences rating in all areas rather than rating independently on the separate factors. Interrater reliability coefficients must be established to validate the use of rating scales, and sufficient latitude must be given the supervisor to provide a response to a clear statement of performance.

Evaluations of rating scales have led to several important developments in the past few years to counteract these errors. Comparison of police officers establishes a relative ranking by simply rank-ordering the overall ratings and comparing each officer with everyone else in the group being rated. These paired-comparison ratings determine which of any two police officers being compared performs better. Forced-choice rating scales provide several, usually few, statements from which the rater must select, while not in themselves assessing good or bad performance. The rater is forced to choose between statements which best describe the officer. Behaviorally anchored scales attach statements to the numerical ratings based on critical incidents describing the performance area. Officers are rated by selecting examples of performance ranging from good to bad within a given category. Two findings emerge from the evaluations of these different approaches to rating scales. First, varying degrees of resistance to rating errors will be demonstrated dependent upon the particular department and environment used. Secondly, similar to use of psychometric testing, agencies can use these scales effectively although department administrators may not be directly involved in development of the rating scales.¹⁰

Voids in Personnel Evaluation

No evaluations were found in the authors' screening of materials which addressed classification and compensation,¹¹ personnel support services (medical, counseling, educational assistance, retirement counseling, recreation, etc.),¹² and labor relations and collective bargaining.¹³ While as many as 80 percent of municipal police departments have some proportion of their sworn per-

sonnel subject to civil service or merit systems, only one recent study has attempted to assess their impact on police personnel administration.¹⁴ This study is an ambitious undertaking and adds to the small body of knowledge on this topic, but fails to contribute any significant findings on the impact of civil service on police effectiveness. One assessment of practices in police conduct and discipline led to the development of prototype rules of conduct and systems for citizen complaint review and disciplinary action.¹⁵ The prototype rules and procedures are currently undergoing a pilot implementation in several police agencies.

Little significant evaluation has been done on promotion procedures and testing in police agencies. The concept of assessment centers which use simulated situations to simultaneously evaluate a number of candidates for promotion has been advanced in the last few years as a more comprehensive and valid form of promotion and identification of candidates for career advancement. This review does not include evaluations of assessment center effectiveness in producing better qualified command personnel than traditional "paper-and-pencil" testing methods. However, excellent reference material on this promising concept is available.¹⁶

TRAINING

Entry-level Training and Training Delivery Systems

A 1968 survey of police recruit training by the International Association of Chiefs of Police (IACP) found a mean length of 72 training hours for those agencies providing entry-level training for their recruits. Subsequently, the National Association of State Directors of Law Enforcement Training (NASDLET) adopted a standard of 280 hours for entry-level or basic training. According to a 1978 NASDLET survey, 41 States had legislatively mandated training, with a mean length of 300 hours to be completed within usually 12 months of appointment as a law enforcement officer. A 1975 survey of 240 training academies by NASDLET listed 66 percent of them as having been established in a 10-year period since 1965. This survey further noted that 23 percent of these academies had received Omnibus Crime Control and Safe Streets Act funding directly from LEAA, while 17 percent reported obtaining Federal funds through their State Planning Agencies.

Agencies employing the largest number of law enforcement personnel and accounting for the majority of employed personnel rely primarily on their own training academies or programs. In the NMS surveys, about three-fourths of the agencies employing 500 or more personnel conducted their own training programs, yet the majority of small and medium-sized law enforcement agencies relied on State and regional training academies or educational institutions for recruit training. The 1978 NASDLET survey reported

that POST's in 22 States administered central academies and in 43 States were responsible for some form of certification of regional training academies.

Many early evaluations of training consistently note the limited amounts of training being provided by small- and medium-sized agencies and those in rural areas and States. Many of these agencies found it difficult to conduct more than rudimentary and procedurally oriented training due to limited numbers of field personnel and the need to fill position vacancies immediately to reach authorized police strength. Personnel and time constraints also precluded the establishment of fulltime training staffs or instructors in these police agencies.

Thus, it is evident that the tremendous growth in recruit training programs and length of training in the past 10 years is due largely to the efforts of POST and the availability of Federal funding from the LEAA.

The authors reviewed more than 50 State, regional, and local evaluations of recruit training, training academies, and the delivery systems for training programs. It was determined that only a very few could be legitimately classified as evaluation. Most of these evaluations simply compare the quality and quantity of training in relation to the standards and goals of the National Advisory Commission and other training academies. The majority of those that could be construed as evaluation generally utilize training participant survey instruments to determine student reactions to course content, the level and quality of instruction, and perceptions of training "usefulness" in performing their duties. Occasionally, participant supervisors and agency heads of the participants will be surveyed for their opinions as to any perceived impact of training on the participants upon return to their agency.

Survey instruments most frequently employ a Likert-type scale with responses ranging from "strongly disagree" to "strongly agree" in conjunction with open-ended questions and responses. Findings typically list frequency distributions or a rank-order of responses, selective reiteration of openended responses, and then conclude with a subjective analysis. The end products of the evaluations tend to be rearrangement of training courses or projections of demand for future training course offerings.

To the extent that such participant surveys or needs assessments provide a feedback mechanism to the training content and delivery systems, they could be categorized as evaluation. However, these nontechnical evaluations which rely solely on participant surveys do little more than indicate that police officers prefer certain types of training subjects, formats, and instructors. Most of these officer surveys do not measure any knowledge or skill transfer with the exception of occasional pre- and posttest-

ing on subject areas. Criteria for sound evaluation design would require, in addition to before and after testing for acquisition of learning, that dependent and independent variables be isolated and control groups established for comparison with the trainees.¹⁷ This usually is not done because of time and funding constraints and/or because evaluation is so frequently conducted by training personnel unskilled in research design and evaluation techniques. Such constraints notwithstanding, the consensus of most evaluators is that some measurement of the objectives stated or implicit in the training curriculum must be undertaken to be classified as evaluation. Based upon our review, a large number of evaluations do not attempt to measure changes in attitude, behavior, or job performance subsequent to the training.

We do not suggest that meaningful evaluation of basic training programs has not been done, but rather that the significant issue of training impact is largely not addressed. When training impact is of primary concern in the evaluation, findings and conclusions are often postulated without quantitative measurement of clearly stated performance objectives and behavioral outcomes. Thus, it is extremely difficult for us to identify training academies and programs that can be adopted as models or to determine what components and elements of the training were correlated with or contributory to the stated success of these programs.

The most significant evaluations of basic training have contained some element of functional research into police duties and into discrepancies between expectations, perceptions, and performance of roles and tasks. Officers are surveyed as to how often and whether they perform certain tasks, the relative amounts of time spent on each task, and whether specified skills or knowledge are relevant to their daily performance of tasks. Training content and objectives are then evaluated and reformulated into operational programs on the basis of this research. Conclusive statements of training content, format, and strategies cannot be made from these evaluations except to note that the design of training programs should utilize job analysis or critical incident techniques to identify deficient areas.

In addition to structuring basic training requirements on job content, it has been suggested that a complete job inventory be constructed which defines the total set of work requirements.¹⁸ This job inventory is an application of the Comprehensive Occupational Data Analysis Program (CODAP) developed by the Occupational Research Division of the U.S. Air Force Human Resources Laboratory. The percentage of police officers performing each task and the average amount of time spent by officers on each task are considered inadequate for establishing training content. Roberts is concerned also with the difficulty of learning the task, practical consequences of inadequately performing the task, and tolerance allowed in time to perform the task.

CODAP was used in a study undertaken by the Texas Commission on Law Enforcement Officer Standards and Education. More than 3,000 police officers throughout the State listed amounts of time spent on each of 395 tasks. A questionnaire containing the additional factors of task difficulty, task delay tolerance, and inadequate task performance consequences was distributed to their supervisors. The CODAP program was then applied to the data collected in the questionnaire to produce a prioritized list of tasks. The results indicated that the less difficult tasks were performed most often by officers with less experience, while more difficult tasks were generally performed by more experienced officers.¹⁹

In extending this design to probationary deputy sheriffs in Texas, supervisors established training priorities for the tasks performed by the probationary deputies.²⁰ Supervisors completed the questionnaire, and interrater reliability coefficients were established for each rater on each task factor. Designating training priorities as the dependent variable, Goodgame reports that the percentage of officers performing each task and the time spent on each task were highly correlated with the means of training priority ratings.²¹ The data analysis also showed a high degree of association between ratings relating to tasks performed on the job and the perceived training priority of tasks.

A similar study found that training instructors and command officers disagreed on training priorities for field supervisors in police departments in Texas. Training priorities of instructors were not highly associated with the work of and performance requirements for supervisors. Goodgame thus concludes that training instructors are not an "accurate source of information" in training program design. Rather than negating the importance of instructor input to training program design, the findings simply confirm the consensus of evaluations reviewed by the authors that training programs should both be designed and evaluated on the basis of job requirements and performance expectations.²²

Perhaps the best example of the dearth of technical evaluations on recruit training content and academies is the number of evaluations focusing on the entry-level training standard of 400 hours for new police recruits recommended by the National Advisory Commission in 1973. The following six core areas were recommended with percentage distributions of time for each area (Standard 16.3):

- Introduction to the Criminal Justice System 8%
- Law 10%
- Human Values and Problems 22%
- Patrol and Investigation Procedures 33%

- Police Proficiency

18%

- Administration

9%

The NMS data for 1974-1975 listed only 32 percent of law enforcement agencies meeting or exceeding this standard, which accounted for an estimated 47 percent of all law enforcement personnel employed. Average course lengths were directly related to agency size. Ninety percent of police departments and 72 percent of sheriff's departments with 400 or more sworn employees provided 400 hours or more training. While course lengths in State or regional academies are generally lower, many departments provide additional training to supplement State-mandated training hours.

The 1975 NASDLET survey of over 200 training academies indicated that disproportionate emphasis was placed on the core areas of patrol and investigation procedures and police proficiency (67 percent). It further noted only 7 percent of curriculums addressing the Commission's recommended course time of 22 percent for training in "human values and problems." Although evaluations discussed in a later section of this chapter on "role learning" in fact support the higher level of emphasis in this area, the significant point is that the Commission's core content and percentage distributions of time were not supported by any research findings or evaluation when they were recommended. The authors' review of evaluation material for this chapter reveals that neither the training duration nor content standards have been subsequently validated. The Commission's report, in discussing the standard, noted that a New York City study showed wide divergence of thought on training curriculum emphasis among the 15 highest and lowest duration training programs of 60 surveyed police agencies.²³ Yet, the Commission developed its recommended core content and distribution of training hours primarily from the curriculums of larger police agencies such as Seattle, Los Angeles, Oakland, and Chicago.

As part of a process to develop A Master Plan for Law Enforcement Training in New York State, a National survey of recruit-level training was undertaken in 1977.²⁴ The results of this survey are similar to the 1975 NASDLET survey which showed variance with the National Advisory Commission percentage distribution recommended for core content areas. An average of 31 percent was devoted to the area of "Criminal Law, Evidence, and Investigation," and an average of 24 percent of the curriculums provided training in "Patrol and Traffic Training." Based on this survey and the authors' review of police personnel and training evaluations, it is estimated that about half of the States have conducted task analyses for the police. It is not clear, however, if all the analyses have addressed both sheriff and police positions, if they were conducted on a statewide basis, or if the job analyses were conducted for development of training

programs or job-related testing. Nevertheless, in our opinion this emphasis on course content is better supported by the States that have conducted job analyses.

Performance Objectives and Field Training

As is evident from the preceding discussion, the success of training programs cannot be adequately evaluated unless there are clear statements of objectives. The National Advisory Commission recommends that performance objectives be established which designate minimum levels of demonstrated proficiency against which training programs can be measured (Standard 16.2). In addition to measuring training programs, the standard of performance objectives was also recommended to measure whether trainees acquire these levels of proficiency and apply them in their job performance. A corollary standard recommends that 4 months of supervised recruit field training be provided to complement the recommended 10 weeks of classroom instruction. The objective of field training is to provide reinforcement of formal classroom training and an applied learning situation for the knowledge, attitudes, and skills imparted in the classroom setting.

In 1968 an IACP study reported an average of 72 hours' field training,²⁵ and Saunders (1970) stated that in cities of over 10,000 population,⁵⁸ percent of the police agencies provided no field training.²⁶ Several training content surveys have focused on the amount of recruit field training, but it is difficult to determine the relative number of police agencies throughout the country which conduct supervised field training as an integral part of the recruit training process. The surveys reveal that most police agencies and POST-affiliated academies have been placing increasing emphasis on field training, but the most effective way of accomplishing this is not clear.

One approach is to designate a period of supervised field training at some point in the formal basic training program. Another approach designates a specified period of time after completion of formal basic training for supervised field training under the direction of a trained coach. The latter approach appears to develop a more comprehensive measure of job performance by using a training record which rates the trainees and charts their progress in mastering a number of tasks considered essential to field performance. Most programs reviewed by the authors were not developed in relation to overall objectives of both the formal and field training experiences. This situation is particularly true in States where field training is conducted by the individual agency independent of the formal training provided by a State or regional academy.

Instructional Staff

The instructional staff of most police academies is part-time, even though some proportion of academy staff are full-time employees in such areas as training coordination, materials and curriculum development, and academy operations. The 1975 NASDLET survey of 200 police academies found that 80 percent of the faculties were part-time. While there is no evidence from training evaluation to support the need for full-time instructors, it is clear that part-time instructors need specialized training and certification in order to provide the same quality of instruction as full-time academy staffs. Thirty-two States in the 1978 NASDLET survey certified instructors, and a number of these States provide 40 hours of "training of trainers." Evaluations of some training academies with full-time staffs indicated that heavy course loads, large class sizes, and specialized instructional fields tended to detract time from course preparation, revision, research, and opportunities for professional staff development. Training academies which most frequently received favorable ratings on instructional staff by outside evaluators have a nucleus of full-time training staff cross-trained in core subject areas supplemented with outside or part-time instructors certified in specialized areas.

Law enforcement agencies have historically utilized both sworn personnel and supervisory or command personnel as academy instructors. The selection of instructors appears to be based on tradition in the agency or on a "stress environment" where discipline and paramilitary concepts are emphasized. Neither training literature nor evaluations reviewed by the authors supported the need for an instruction staff comprised exclusively of sworn or command personnel. The key variables or characteristics of an effective instructor are consistently listed as knowledge of the subject areas, teaching skills and experience, desire to teach (as opposed to an assignment to teach), and interpersonal communication skills.

Instructional Techniques

A recurring criticism found in training evaluations is the use of the formal lecture as the primary mode of instruction supplemented with films, reading and memorization assignments, and demonstration and practice of skills. In this mode, lectures are given from texts or reference handouts which are often embellished with anecdotes from the instructor's practical field experience. As a result, the recruit's knowledge of field situations to be encountered must be deduced from the anecdotal accounts. These assessments frequently conclude that the formal lecture must be supplemented with varied instructional methods such as audio-visual, reading assignments, demonstrations, and practice sessions to achieve a satisfactory learning experience.

Many training programs have attempted to place greater emphasis on group instructional techniques to highlight field situations and stimulate discussion of appropriate police officer responses. Some evaluations suggest that group-centered learning is more conducive to presentation of such topics as human relations, family crisis intervention, patrol techniques, and, in general, the proper orientation of police roles and tasks. The authors tend to agree with these findings although they are largely subjective, do not use control groups, and thus have difficulty demonstrating increased learning effectiveness from the group instructional techniques.

California's noted Project STAR, a comprehensive research effort aimed at criminal justice personnel roles, functions, and skill requirements, listed several categories of widely used group instructional techniques. These categories are: conferences--debate, discussion group, panel seminar; field experiences--field trip, interview, operational experience, operational observation; simulations--case study, game, role play.²⁷

A "contingency model" for police training is advanced by Finkelman and Reichman (1974) which states that instructional techniques should vary according to subject matter and the characteristics of the trainees.²⁸ Too often, in their opinion, trainers attempt to develop a single comprehensive program that can be used by all, and others willingly adopt it without regard for clear differences in trainee backgrounds and unique characteristics of State, county, and local agencies.

The largest single impact on instructional techniques has been the expanded use of closed-circuit television and video tape cassettes. The "stop action" and "instant replay" capabilities of video tape permit emphasis on and discussion of key program elements of the recording. As an instructional aid, video tape recordings have stimulated discussion on the proper knowledge and application of law and on the use of discretion and judgment in field situations. Role-playing exercises and simulation of street encounters improve the cognitive learning process, sharpen observational skills, and provide visual and audio cues of participant behavior. In addition to simulations, video tape is used to present a wide range of subject areas, department policies and procedures, preliminary demonstrations of skills (self-defense, first aid, firearms, etc.), and to bring remote locations (courtrooms, booking procedures, etc.) into the classroom setting.

Video tape has also been used as a form of programmed learning: The individual interacts with a programmed television script at his or her own pace. This has been recommended as particularly effective for remedial learning and for specialized training needs.

Evaluations of the role of video tape in the training programs have been limited primarily to participant reaction to its use

and to the impact of closed-circuit television as a delivery mechanism for training accessibility. Experiments in Suffolk County, New York, (1972) and in the State of South Carolina (1969)²⁹ have demonstrated the capability to provide training in remote and different locations simultaneously and to a large number of groups in a short period of time. This capability is based on the use of a library of cassettes replicated from a master tape or the use of a closed-circuit educational television network. Departments are able to rapidly disseminate, for example, changes in field operations which occur as a result of new policy, court decisions, or legislation. Video tape also facilitates communication between large and geographically decentralized State and county law enforcement agencies.

None of the evaluation material reviewed by the authors identified a specific curriculum as being especially well-suited to presentation by closed-circuit television, or attempted to measure achievement of specific learning objectives, or used control groups to specifically isolate the video tape or closed-circuit television as the causative factor in enhanced knowledge transfer, reduced learning time, or improved retention. The potential for computer-assisted instruction (CAI) to provide individual self-paced learning with continual student evaluation has been mentioned in several training assessments. The authors discovered no technical evaluations of computer-assisted instruction or programmed learning in an operational setting.

Role Learning

The heavy reliance on the lecture method of instruction is in part due to the paramilitary tradition of the police and the emphasis on "stress" in the training environment. Several evaluations have challenged the effectiveness of stress training, but the findings cannot be universally accepted because of differing meanings and degrees of "stress" in highly varied police agencies. Some training programs could be classified as "stress" yet offer a high degree of participant training (discussion, seminars, role-playing, etc.). The recruit training process forms an initial role orientation for the police officer that impacts the remainder of the officer's career. Considerable variance exists between role concepts in recruit school and the realities of performance in field situations as an experienced officer. Even some of the earliest training evaluations agree with these frequent findings.

On the basis of a longitudinal study of recruits and experienced officers in the New York City Police Department, John McNamara (1967) found that many of the "gains" achieved in recruit training were lost over time in the field due to variance of field conditions with training program content.³⁰ While his findings are not supported by hard data, McNamara argues that the semi-military model of the police department and training academy

CONTINUED

3 OF 5

graduates a "standardized product" who can adapt to the administrative structure and policies of the department. Yet, McNamara feels this police officer is ill-equipped to make "maximally effective decisions" involving discretion and judgment generally made in the absence of direct supervision. McManus noted an altered image 10 years after recruit training from predominantly a law enforcement officer to primarily an "agent of social control." The study suggested that curriculum development focus on presenting varied educational situations that reflected the "variety of continuing small crises encountered on the street."³¹

Project STAR is the only evaluation of training programs reviewed by the authors which specifically measured the ability of certain training approaches to develop the desired attitudes and behavior necessary for successful role performance.³² A job analysis on the basis of performance objectives identified 13 roles and 33 tasks. Operational behavior required for satisfactory performance of a particular task and the actual conditions under which it is performed in an operational setting were clearly stated. Role training programs were designed, tested, produced, and evaluated to examine the hypothesis that students actively involved in both individual and group learning experiences centered on real life training situations will more likely develop the desired attitudes and behavior for successful role performance.

A series of training modules was developed to focus both on police roles and tasks and on their relationship to other components in the criminal justice system. The training modules follow a series of learning steps for each role: (1) presentation of the role concept and its performance objective and their related tasks, situations, and examples of role behavior; (2) selection from four to five optional learning activities; and (3) review. The modules are estimated to involve 104 to 156 hours of instruction time and are designed to be integrated into existing training programs. Pre- and posttests of experimental and control groups suggested that the training approach of STAR more successfully reinforces and retains role concepts and develops desired values. Both groups had entered their respective training with similar knowledge and attitudes on the desired role concepts, but the STAR training group showed statistically significant differences in measure of achievement in understanding and accepting the role concepts. In light of previous findings that operational experience tends to erode desirable concepts in traditional police recruit training programs,³³ a longitudinal study should be conducted to determine if STAR training participants retain similar role concepts in job performance over time.

Inservice, Supervisory, and Management Training

The National Advisory Commission recommended that annual inservice training of 40 hours duration be provided to all sworn employees to maintain, improve, and update knowledge and skills (Standard 16.4), and that upon promotion, supervisory and management personnel should receive appropriate training for the new position. Technical evaluation of training programs in these areas is virtually nonexistent in terms of curriculum, emphasis, and instructional techniques. The authors' review of training assessment surveys and evaluation reveals that law enforcement administrators in principle agree with these recommendations, but in practice few agencies regularly conduct or provide opportunities for such training.

Inservice training is usually conducted for personnel assigned to specialist positions, and generalist training for all sworn employees is rarely provided. Generalized, inservice training most frequently focuses on criminal law, operational policies and procedures, and weapons, with lesser emphasis on order maintenance, peace-keeping, and service functions of the police. The 1975 NMS survey data revealed that less than one-fourth of all sworn personnel had attended a formal inservice training course (other than rollcall training) in 92 percent of the agencies with 400 or more employees. The major constraints faced by most agencies are reported to be fiscal and manpower--the inability to provide adequate numbers of replacement officers in the field for those attending inservice training programs. There were no evaluations screened which assessed levels of job performance and effectiveness as related to quantity and type of generalized inservice training. The use of briefing sheets, rollcall training, and video cassettes in a number of departments is deemed adequate by them to maintain, update, and improve necessary skills.

Training records and histories are maintained by many departments, but in most cases these are merely chronological listings of training, rather than a tool for evaluating training needs. Several researchers have advanced the concept of a training manager to assess present and future agency training requirements through use of a "training assessment matrix." This tool can be used for scheduling needed training and coordination with available outside training courses; it can also be used to assure adequate numbers of trained personnel for specialized assignments when needed, and to establish career development plans for management succession and placement.

Only recently have clear distinctions been made among training content and needs of supervisory, middle management, and executive training. Evaluations of these programs, similar to most basic training evaluations, tend to rely on participant surveys of the general usefulness of the training content to fulfillment of their job requirements. No task analyses specifically directed at these three levels were discerned by the authors in our review of training

evaluations. The differences among many of the course offerings appear to be simply the labels which are changed while the course content is offered at several levels. The varying educational and training levels of supervisory and command personnel are in large part responsible for the interchangeability of course content in many advanced training programs. In this manner one is assured of getting certain supervisory and managerial concepts at some point in time after promotion from field officer status.

The National Advisory Commission recommended that personnel should receive adequate training for the responsibilities of their new position upon promotion. The NMS surveys reported that 90 percent of police chiefs and 80 percent of sheriffs favored supervisory training, but few departments except the largest require the training, and less than one-fourth of supervisory personnel surveyed reported receiving formal training for their new position.

The promotion criteria for most first-line supervisory positions place emphasis primarily on officer-level knowledge and skills under the assumption that good field officers make the best supervisors. The NMS survey identified, for example, 12 major tasks performed by line supervisors which are not ordinarily performed by field officers, yet a majority of supervisors reported on-the-job experience as the means of acquiring knowledge and skills for their supervisory tasks. The majority of police officer standards and training commissions offer a first-line supervisory course, and seven States have mandated supervisory training according to the 1978 NASDLET survey.

In participant surveys, supervisors consistently cite the following areas needing emphasis: decisionmaking processes, motivation and performance appraisal of subordinates, assignment and scheduling of personnel.

Middle-management and executive training follows somewhat the same pattern as supervisory training--recognition of a need by police administrators with few actually having received management training. Styles of leadership and management, policy formulation, program development and planning, budgeting and fiscal management, labor relations, and overall department performance measurement are frequently cited as needed subject areas in participant surveys of police executives. Unless selected to attend a lengthy law-enforcement-oriented management training program, such as the FBI National Academy or the Southern Police Institute at the University of Louisville, police managers attend short courses in specialized workshops and seminars of 1-to-5 day duration, e.g., collective bargaining, management of criminal investigative function, police discipline, fleet management. Evaluation of management training has been difficult or rarely attempted due to the difficulty of measuring training impact on managerial effec-

tiveness. Management training programs in POST and regional academies and educational institutions have had to rely primarily on participant surveys of perceived job relevance of course offerings.

Specialized Training

Specialized training programs have generated the highest quality technical evaluations with formal evaluation designs. The authors reviewed a dozen program impact and process evaluations in this category. Both groups of evaluations are most often required as part of LEAA-funded action grant projects. Occasionally, the training program itself is the object of evaluation as an innovative subject area or a specific training project, but more frequently specialized training is evaluated as a component in a grant project. Training in human relations, community relations, crisis intervention, conflict management, team policing, and management of criminal investigations are examples of training in specialized subject areas that have generated both "hard and soft" evaluations.

Team policing and management of criminal investigations--programs developed with LEAA funds, from research findings and experimentation--have major impacts on the delivery of police services and policymaking. Training efforts in these areas center on knowledge transfer of concepts and new role orientations to patrol and investigative responsibilities. Evaluations are concerned more with the relationship of change to overall project success than the training program itself.

The highest quality training evaluations screened by the authors were those concerned with crisis intervention training.³⁴ In these evaluations, knowledge transfer, officer performance, and department impact are analyzed for crisis intervention programs. They stress social service agency and counseling referrals for officers responding to family disturbance calls. These evaluations point out a number of problems recurrent in evaluation of many training programs. Some training programs select those officers most receptive to the training content rather than those most in need of the training, thus inducing a bias into the training evaluation. Department members who are trained are not always widely dispersed throughout the department to maximize availability of skills when needed in field encounters. Some training programs are so dependent upon the expertise of program designers and instructors that transfer of technology to other departments is difficult.

In our review of selection criteria and testing validation we noted the influence of local factors which precluded universal adoption without local validation. Similarly, in a NILECJ evaluation of crisis intervention training programs

in six cities, a local factors variable (including department support) was found to be the most highly significant variable predictive of program success.³⁵

Education

The so-called "police reform literature" has advocated over the past 10 years increased levels of education in addition to training as a means to development of a "professional" model of law enforcement. The complexity of tasks and changing role of the modern patrol officer and the need to draw qualified supervisory and management personnel from the ranks are frequently cited reasons for educational attainment by police officers. The Law Enforcement Education Program (LEEP) was initiated by LEAA to provide institutional support for curriculum development and funds for criminal justice system personnel to undertake degree programs. In 1960, there were only 26 institutions in the country offering degree programs in criminal justice areas, but 519 institutions had applied for LEAA scholarship and loan funds for FY 1970.³⁶

A 1974 survey by the Bureau of Census indicates that nearly one-fourth of State and local law enforcement had received LEEP financial assistance by 1974, and police personnel received 80 percent of the LEEP funds distributed.³⁷ By the end of FY 1976 LEEP appropriations totaling \$234 million had been provided to approximately 1,200 educational institutions and 250,000 students. Yet, the authors, in screening evaluation materials, found neither comprehensive evaluation of the effectiveness of the LEEP program³⁸ nor unequivocal empirical evidence that college education or criminal justice degrees improve police performance.

The best attempt at assessment of the LEEP program was undertaken by the National Manpower Survey which analyzed course content and faculty, police officer and police executive perceptions of the job-related importance of education, and levels of educational attainment. The NMS surveys of police chiefs in cities of 17,000 or more population revealed that larger agencies tended to rate college-educated personnel more favorably. The survey also showed ratings to be correlated to educational level and age of the chiefs, with younger and college-educated chiefs expressing a strong preference for college-educated officers. While entrance-level requirements for education were less in the smallest agencies (less than 150 employees), a broad consensus was found in all agencies that higher education of existing sworn personnel should be continued.

In reviewing the LEEP program, the NMS analysis noted that 35 percent of all law enforcement courses in LEEP-associated educational institutions could be classified as "training type" courses, especially in the 2-year degree programs which are

the largest in number. Some evaluations of POST training curriculums noted an overlap between State-financed training in specialized and advanced areas and the course offerings of colleges. Stressing that the sine qua non of any educational program is its faculty, the NMS survey revealed that full-time faculty with master's degrees in criminal justice programs do not measure up to the standards for faculty in other programs at these LEEP-supported institutions. This is especially true for 2-year institutions, which most police personnel attend. The continued emphasis on higher education for police officers and the LEEP program was justified by LEAA primarily on the grounds of career development "to develop a broader pool of educationally qualified personnel for advancement to supervisory and managerial level positions."

Through application of the General Educational Development (GED) rating scale, it was determined by NMS that a significant number of tasks performed by criminal justice supervisors and managers required a college education.

A number of studies have sought to establish a correlation between police officer performance and educational attainment. The findings are inconclusive at best, yet deserve further study and validation. One of the earliest studies was a longitudinal analysis by Cohen and Chaiken (1972) which found that officers with 1 year of college were good performers and those with a degree even better, but a multiple regression analysis of 33 predictor variables and 12 performance measures did not show education as a factor.³⁹ Finnegan (1976) stated that police agents (required to hold a degree) performed significantly higher on performance evaluation reports than police officers with no college.⁴⁰ Wierman (1978) reported that State police troopers with a bachelor's degree consistently outscored those with less than an associate's degree on performance appraisal ratings, and that college-educated troopers scored higher than the mean rating of all troopers, but not impressively higher.⁴¹ While considerable emphasis has been placed on curriculum development of police science and criminal justice degree programs, the evaluation literature finds no measurable differences in performance of officers based on type of degree obtained.⁴²

Ostrom and Smith (1974) found only weak support for college education relationships to authoritarian attitudes concerning civil liberties, the Supreme Court, and police organizational characteristics.⁴³ Similarly, Weiner (1976), using a 2-year police science program, concluded that educational level did not affect attitudes toward blacks, other ethnic groups, law and order, or the police themselves.⁴⁴ Miller and Fry (1976) could not correlate a "professionalized" view of police work with educational level, although education has been consistently advanced as the path to professionalism in law enforcement.⁴⁵ The invalidation of traditional assumptions about college edu-

cation and the development of more professional attitudes toward police work and the community may be explained by the fact that police officers attending college possess certain predisposed attitudes which are not affected by college education.

Perhaps the most significant finding in several of the evaluations noted in this section is that college-educated officers are less likely to receive complaints against them or to be dismissed from police service, yet they have higher rates of voluntary resignation. This suggests that additional research is needed on job satisfaction and motivation of college-educated police officers if they are to be retained as the primary pool for selection of supervisors and managers.

SUMMARY CONCLUSIONS AND FUTURE NEEDS

Considering a national effort of almost a decade and several billions of taxpayers' dollars expended to improve both the criminal justice system and its impact on criminality, the state of evaluation data on the justice system's largest component--police personnel--is indeed sad. The good evaluative research that does exist is so scattered that police administrators, planners, and evaluators can find only bits and pieces. Further, the authors failed to find but a few isolated cases where evaluation results have been integrated in future operational planning. For the most part, evaluation reports appear to gather dust on bookshelves or are stuffed in the back of grant project files and soon forgotten.

The gaps, voids, and inadequacies in police human resource assessments are perhaps more significant than the existing evaluation material. Consider the importance of voids in a scarcity of evaluations found by the authors in the Evaluation Clearinghouse, a specialized collection within the broad NCJRS data base, such as:

- Human resources needs forecasting
- Recruitment and retention
- Pay and classification
- Promotion and career development
- Personnel support services
- Labor relations
- Personnel management systems
- Conduct and discipline

- Training needs assessment
- Inservice training
- Management training
- Training and skills inventory
- Education

From the 3,680 so-called evaluation documents in the NCJRS Evaluation Clearinghouse as of June 1978, a staff search produced about 700 abstracts on police personnel and training topics. From these abstracts, the authors selected 170 documents which appeared to be actual evaluation reports. After reviewing each of these 170 reports, we found only 95 (29 personnel and 66 training) "true" evaluation reports. The remainder were research and development project reports or journal articles reporting on State and local personnel and training programs.

This condition indicates a clear need for an effective clearinghouse which is able to collect, classify, and disseminate evaluation material at all levels of criminal justice. Some method must be developed to obtain evaluative material from the State and local units of government as well as from all program offices of LEAA and other police research organizations.

Assuming that evaluations from all levels of law enforcement can be collected and properly classified, a simple system of feedback is needed if evaluative results are to be used to achieve desirable change. Rather than disseminating an annotated bibliography of evaluation results, we suggest a "bottom-line" digest or newsletter--a document which highlights the information most valuable to the police administrator and planner; that is, in nontechnical terms, provide the police administrator and planner with significant evaluative research findings with suggestions for interpreting and applying the findings. Also, such a digest would give the research or evaluator information needed to locate and obtain the full report.

Although some LEAA and SPA officials have been strong proponents of process, program, and impact evaluations, no evidence was found by the authors that program or grant funding consistently and automatically included independent assessment of results. Further, impact evaluations and costeffectiveness measures of major police personnel and training programs are few. Future efforts should emphasize costeffectiveness and impact assessments. For example, in the authors' view, hundreds of millions of dollars should not have been spent on the LEEP program without strong evidence of improved service delivery or of the program's direct and indirect contribution to the prevention and reduction of crime.

Finally, we recommend the development of national evaluation standards and guidelines pertaining to police personnel and training. Included in such standards should be suggested evaluation methodologies for various types of personnel and training programs and a reporting, abstracting, and indexing mechanism for evaluation findings.

NOTES

1 National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, The National Manpower Survey of the Criminal Justice System, Vol. I, Summary Report, 1978.

2 Ibid.

3 Ralph J. Gutekunst, Recruitment and Retention Factors in the Metropolitan Police Department. Century Research Corporation, 1966.

4 See Policewomen on Patrol (1973); Women in Policing, A Manual (Milton, et al. 1974), Police Foundation; Deployment of Female Police Officers in the United States, International Association of Chiefs of Police, 1974; Women Traffic Officer Project, California Highway Patrol, 1976; Women on Patrol: A Pilot Study of Police Performance City (Sichel, et al.), National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1978.

5 George W. Greisinger, Joseph J. Molkup, and Jeffrey S. Slovak, Civil Service Systems: Their Impact on Police Administration. Public Administration Service, 1978.

6 See Ernest C. Furcon et al., Longitudinal Study of Psychological Test Predictions and Assessments of Patrolman Field Performance. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1971; Susan J. Saxe and Martin Reiser, "A Comparison of Three Police Applicant Groups Using the MMPI," Journal of Police Science and Administration, Vol. 4, 4:419-425; Marvin D. Dunnette and Stephan J. Motowildo, Police Selection and Career Assessment. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1976.

7 Furcon et al., note 6 above.

8 David Barry, Robert Goldstein, and Thomas Smith, The Screening and Selection of Police Candidates in Rochester, New York. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1975.

9 Terry Eisenberg and James M. Murray, "Selection" in Police Personnel Administration. Police Foundation, 1974. (This book is an excellent reference on contemporary police personnel administration.)

10 Furcon et al., note 6 above; also Frank J. Landy et al., "Behaviorally Anchored Scales for Rating the Performance of Police Officers," Journal of Applied Psychology. Vol. 61, 6:750-758 (1976).

11 See Carl F. Lutz and James P. Morgan, "Jobs and Rank" in Police Personnel Administration. Police Foundation, 1974. (An excellent review of classification and pay issues.)

12 Terry Eisenberg et al., Police Personnel Practices in State and Local Governments. Police Foundation, 1973.

13 See Chapter VI, "Police Unions in Police Personnel Administration," in Greisinger, cited note 5 above, for a quantitative assessment of impact of police unions.

14 Greisinger, op. cit.

15 International Association of Chiefs of Police, Managing for Effective Police Discipline, 1976.

16 See Roger Reinke, Selection Through Assessment Centers, A Tool for Police Departments. Police Foundation, 1977. Also Dunnette and Motowildo, cited note 6 above.

17 Gad J. Bensinger, "The Development and Use of Evaluation Designs in Criminal Justice Training Programs: The Cook County Model," in Evaluation of Education and Training Programs. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1978.

18 A.E. Roberts, A Career Development System for Texas Law Enforcement. Texas Commission on Law Enforcement Officer Standards and Education, 1975.

19 Ibid.

20 Douglas Goodgame, Analysis and Definition of Critical Training Requirements for Apprentice Sheriff's Deputies in Texas Via Job-Task Analysis and the General Linear Model. Occupational Research Program, Texas A&M University, 1978.

21 Ibid.

22 Ibid.

23 George P. McManus, Police Training and Performance Study. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1970.

24 Robert J. McCormack, John F. Brennan, and Bill Hamilton, A Master Plan for Law Enforcement Training in New York State. Vol. II, Criminal Justice, 1977.

25 International Association of Chiefs of Police, Police Recruit Training, 1968.

26 Charles B. Saunders, Upgrading the American Police. Brookings Institution, 1970.

27 C.P. Smith, D. Pehlpe, and C.D. Weller, Role Performance and the Criminal Justice System: Volume I: Summary of Project STAR. California Commission on Peace Officer Standards and Training. (Cincinnati, Ohio: Anderson-Davis, 1976.)

28 Jay M. Finkelman and Walter Reichman, "Police Training Strategies: A Contingency Model," Journal of Police Science and Administration. Vol. 2:423-8 (1974).

29 David H. Burns, Statewide Law Enforcement Education Through Television, South Carolina Law Enforcement Division, 1969; and Project Teletraining--Suffolk County, New York Police Department, Abt Associates, Law Enforcement Assistance Administration Exemplary Project Screening and Validation Reports, 1974.

30 John H. McNamara, Role Learning for Police Recruits: Some Problems in the Process of Preparation for the Uncertainties of Police Work, diss., University of California, Los Angeles, 1967.

31 McManus, note 23 above.

32 Smith et al., note 27 above.

33 See Arthur Niederhoffer Behind the Shield: The Police in Urban Society. (Garden City, New York: Doubleday, 1967.) (For a discussion of police cynicism.)

34 See Peter B. Wylie et al. An Approach to Evaluating a Police Program of Family Crisis Intervention in Six Demonstration Cities, Human Resources Research Organization (Alexandria, Va., 1976) and David R. Hensley, Evaluation of Crisis Intervention Training for Wisconsin Law Enforcement Officers. Center for Advanced Studies in Human Services, University of Wisconsin (Milwaukee), 1977.

35 Wylie et al., previous note.

36 Charles W. Tenney, Higher Education Programs in Law Enforcement and Criminal Justice. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1971.

37 U.S. Bureau of the Census, Criminal Justice Employee Characteristics Survey (1975).

38 Government Accounting Office's 1975 audit and evaluation report on the LEEP Program took a sample of only 550 students who had obtained a certificate or degree. See also Tenney, cited note 36.

39 Bernard Cohen and Jan Chaiken, Police Background Characteristics and Performance. National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, 1972.

40 James C. Finnegan, "A Study of the Relationships Between College Education and Performance in Baltimore, Maryland," Police Chief. Vol. 43, 8:60-62 (1976).

41 Charles L. Wierman, "Variances of Ability Measurement Scores Obtained by College and Noncollege Educated Troopers," Police Chief. Vol. 45, 8:30-33 (1978).

42 Wierman, *ibid.*; Finnegan, cited note 40; John D. Madell and Paul V. Washburn, "Which College Major is Best for the Street Cop," Police Chief. Vol. 45, 8 (1978).

43 Elinor Ostrom and Dennis C. Smith, The Effects of Training and Education on Police Attitudes and Performance: A Preliminary Analysis. Workshop in Political Theory and Policy Analysis, Indiana University, 1974.

44 Norman L. Weiner, "The Educated Policeman," Journal of Police Science and Administration. Vol. 4, 4:450-457 (1976).

45 Jon Miller and Lincoln J. Fry, "Some Evidence on the Impact of Higher Education for Law Enforcement Personnel," Police Chief. Vol. 45, 8:30-33, (1978).

Police Command, Control, and Communications

James M. Tien is an associate professor of electrical and systems engineering at Rensselaer Polytechnic Institute and vice president of Public Systems Evaluation, Inc.

Kent W. Colton is Professor of Public Management at Brigham Young University. Dr. Colton also is a vice president of Public Systems Evaluation, Inc., Cambridge, Mass.

ABBREVIATIONS

Police command, control, communications (PCCC) is discussed in these six areas:

911--The increasingly standardized telephone number for emergency calls to police, fire, ambulance, and similar services.

CAD--Computer-assisted dispatching.

AVM--Automatic vehicle monitoring.

MDC--Mobile digital communication.

MIS--Management information system.

RCS--Regional communication system.

The reader may want to mark this page with a paper clip or the like so as readily to refer back to it.

INTRODUCTION

This paper addresses the use and evaluation of computer based technology as it relates to police command, control, and communications (PCCC). The focus is on command and control; police communications is included only to the extent that it directly impacts the command and control objectives.

Unfortunately, there have been very few formal evaluations of PCCC innovations. The literature is mainly feasibility studies, trade journal articles, and subjective assessments by police administrators. Further, much communications literature can be overlooked for command and control purposes--it covers such issues as radio signal distortion, radio spectrum allocation, facsimile transmission, speech scrambling, and the switching and sharing of data between different jurisdictions. Nevertheless, the available PCCC information does provide a basis for identifying future needs in research and evaluation.

The paper is divided into three parts. The opening, background section of this paper discusses the active Federal role in PCCC development, the availability of computer technology, and the strong desire to improve the performance of PCCC systems.

The second section examines six specific PCCC applications: 911, computer-assisted dispatching, automatic vehicle monitoring, mobile digital communication, management information systems, and regional information systems. We will describe four different ways in which evaluators can measure them: input, process, outcome, and system. Studies by the President's Commission on Law Enforcement and Administration of Justice (1965-67) suggested that PCCC applications might achieve costeffective reductions in police response time, defined by other authors as "the interval between a citizen's request for service and the arrival of a field unit at the scene."¹ In the second section we will explore how well the Commission's hopes have been achieved.

The third section (p. 326) discusses future alternatives. It recommends that research be undertaken to develop PCCC systems which are better integrated, more proactive in assistance to police dispatchers and decisionmakers, and more flexible in terms of changing technology. Future evaluations with a common framework are urged, and the authors recommend that these more intensive evaluations extend beyond input and process and begin to examine outcome and systemic impacts.

Background Issues

In 1877 Albany, N.Y. had five telephones installed connecting the mayor's office with police precinct stations.² Thus began the development of modern police command and control procedures. The telephone, first simply an auxiliary to the telegraph, gradually allowed the public to gain voice access to the police department as well as to give the department better means of communication with and control of its personnel.

In 1923, the Pennsylvania State Police adopted teletype machines. In 1928, Detroit deployed police cruisers equipped with radio receivers. Two-way police radio came to Boston in 1934. Transistorized circuitry, in the 1960's, gave police smaller and more reliable radios. Progress in communications technology was even more dramatic in the following decade.

The Federal Role

Like so many other developments in criminal justice, most present involvement dates back to 1965 and the establishment of the President's Commission on Law Enforcement and Administration of Justice. The Commission identified science and technology as a promising area for research and innovation. In its final report,³ the Commission called for the use of and experimentation with new technologies both to aid in police operations and to store, transmit, and transform criminal justice information. PCCC was among several areas highlighted for further exploration. The Commission wrote:

The entire police command-and-control function should be subjected to a basic reexamination taking full account of the promising new technologies offered by computers and communications links...An experimental program to develop a computer-assisted command-and-control system should be established with Federal support.⁴

Given the labor-intensive nature of police work and the tradition of devoting only a small percentage of police budgets to research and development, policymakers of this period focused on improving police services by allocating dollars for equipment and technology. Creation of the Law Enforcement Assistance Administration (LEAA) in 1968 added large-scale Federal resources for such purchases. With the end of the Vietnam war, technology-intensive industries sought to increase their domestic markets, thus contributing to the expansion of computer-related innovations.

Introducing technological change, however, requires more than pressure from the top, in this case from the Crime Commission. Although much police literature promotes technology, surprisingly little covers evaluation of the process of implementing new technologies and the results achieved. This paper, then, explores what has transpired since the Crime Commission report and, most importantly, what difference the use of PCCC technology has made in promoting efficiency of control and command.

Computer Technology

Computers had been available commercially in some form or other for more than a decade before police began using them in the early 1960's.⁵ By that time, technology had begun to bring together several streams of development: information transmission, the development of data analysis and reduction systems, and computer modeling. This ingathering of developments set the stage for effective police use of computer technology.

A survey of 79 cities in 1970 found that law enforcement was the single most recurring municipal application of computer technology.⁶ By 1975, city computer use for police applications still ran a strong second to financial uses (payroll, accounting, and collection). In 1971, in response to an International City Management Association (ICMA) survey of 498 U.S. police departments in cities with populations over 50,000, 44 percent of the departments indicated they were using computers. By 1974, an ICMA survey indicated this figure had risen to 56 percent.⁷

Still, the large majority of computer applications by police remain "routine"--straightforward, repetitive information-processing activities such as traffic records or maintaining real-time patrol and inquiry files. Nonroutine use such as resource-allocation models or computer-assisted dispatching (CAD) systems, in which the machine begins to become a tool for decisionmaking and strategic planning, has been disappointing.

In fact, a significant number of nonroutine efforts have failed. In the 1971 ICMA survey, 61 departments indicated they would install a CAD system within the next 3 years. The followup survey in 1974, however, indicated only 15 departments with systems close to being operational. The number of departments with CAD systems has increased since 1974, but the early setbacks highlight some of the difficulties surrounding use of PCCC technology.

PCCC Performance

The Crime Commission identified the police communications center as the focal point for command and control. Improvements in its function, it was suggested, might improve apprehension rates, thus serving as a deterrent to crime. The Commission pointed out these problems limiting police effectiveness:⁸

- A great deal of information flows through a communications center, but little can be readily recalled to provide feedback on the results of actions.
- The location of patrol cars is only crudely known.
- The communications center has a significant impact on response time; technological improvements in processing consumer requests might reduce delays in response.
- In many police departments, radio congestion impairs communication between dispatchers and field officers.

Computer-assisted dispatching (CAD) brought together many new police tools, through the partial automation of the call-answering, processing, and dispatching activities. CAD automatically matches

the address of a call for service with the appropriate police patrol beat through a computerized geographic file. Instant recall of dispatch data is possible.

CAD, however, does not track the location of police vehicles, so automatic vehicle monitoring (AVM) systems also were suggested. As defined in this paper, AVM systems provide a dispatcher with real-time location estimates of each vehicle in a fleet. They provide additional vehicle status information such as "in pursuit," "enroute to scene," or "driver door open."

Mobile and portable digital terminals were suggested to permit officers to communicate by keyboard with the dispatch center. Also recommended were "911" emergency telephone systems and improved radio equipment.

Command and control not only pertains to the ability of the dispatcher to deploy vehicles but to the police administrator's control over patrol operations. Thus, though some innovations are routine (e.g., the automated transfer of information from the telephone operator to the dispatcher), some are definitely nonroutine: tracking and monitoring vehicle location; timing the length of time spent on a call for service and raising a "flag" if a call takes over, say, 30 minutes; and providing new information for a broader management information system.

Study Framework

As illustrated in Figure I the approach in this study is based on a three dimensional framework including four evaluation measures, six distinct applications which have been implemented in PCCC, and four basic functions of PCCC. Thus one could consider the PCCC assessment problem as comprising 96 separate assessments: 4 x 4 x 6. Unfortunately, the available information does not allow for such detailed assessment. Figure I, however, may portray the interrelationships of these various factors in a manner useful to future evaluators.

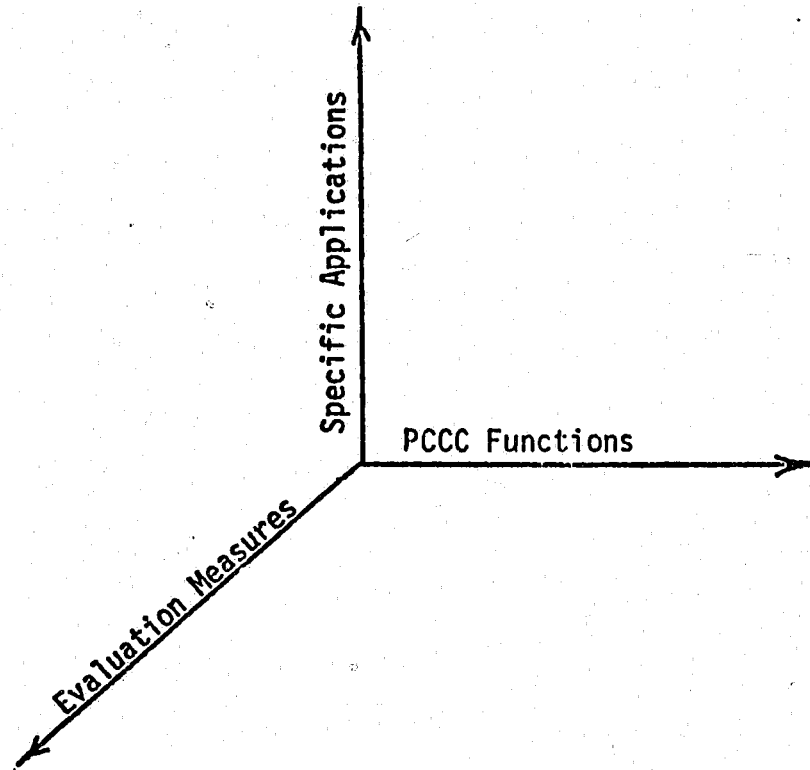
PCCC Functions

On an immediate, tactical basis, an effective PCCC system must be able to monitor available police resources so as best to respond to the needs identified. On the longer-term, strategic basis, PCCC must be able productively to manage the available resources. As Figure I identifies the four PCCC functions, Figure II further details them in terms of specific actions and time measures.

The functions can often overlap. An officer on patrol might witness a criminal incident in progress. This would result in identification of a need for police assistance, a change in the officer's pre-

Figure 1

PCCC Study Framework



PCCC Functions:

- Needs Identification
- Status Monitoring
- Response/Adjustment
- Resource Management

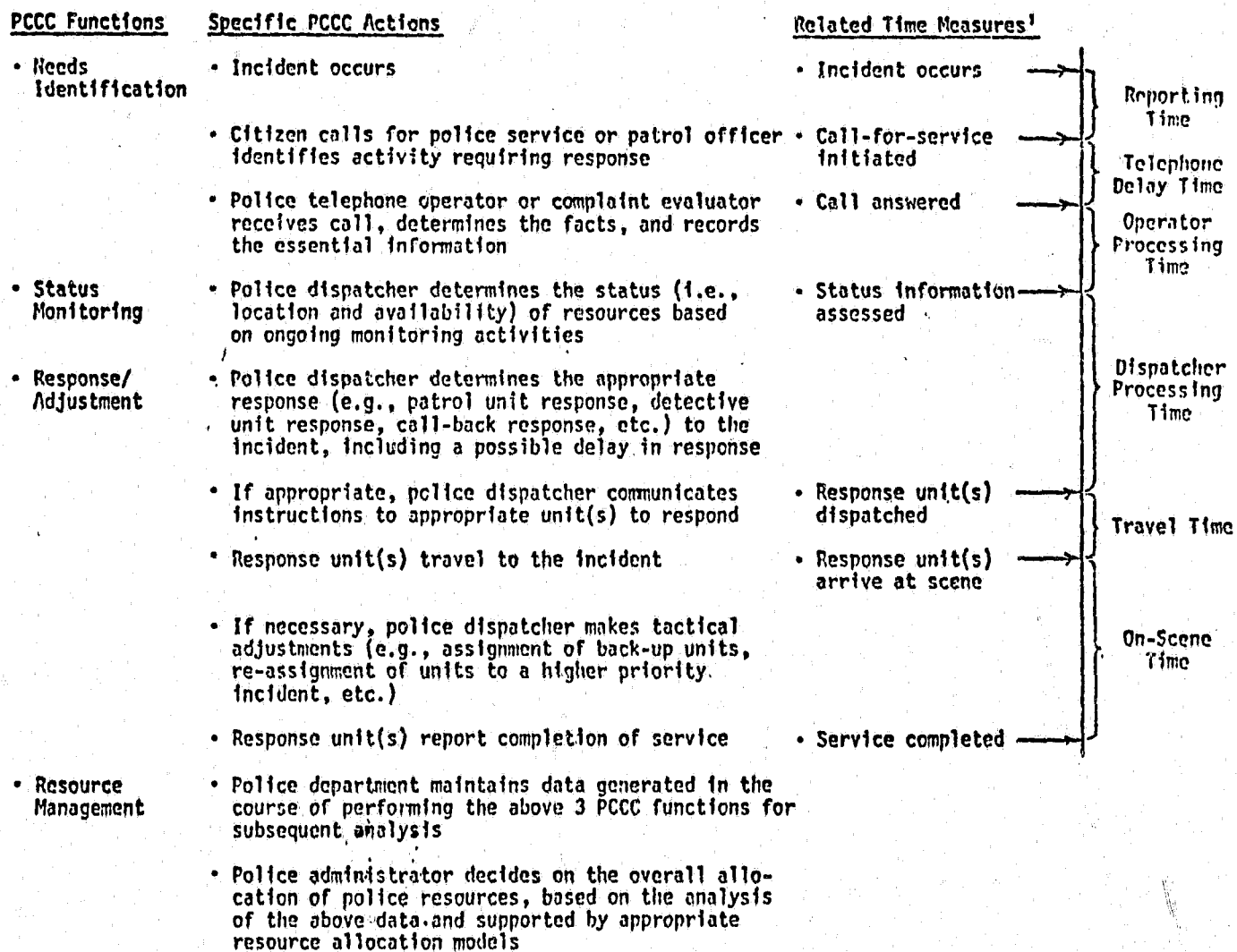
Evaluation Measures:

- Input
- Process
- Outcome
- Systemic

Specific Applications:

- Nine-One-One (911)
- Computer-Assisted Dispatching (CAD)
- Automatic Vehicle Monitoring (AVM)
- Mobile Digital Communication (MDC)
- Management Information System (MIS)
- Regional Communication System (RCS)

Figure II
PCCC Functions



¹The time measures are *not* drawn to scale. Other composite time measures include *dispatch delay time* (i.e., sum of the operator and dispatcher processing times), *response time* (i.e., sum of the dispatch delay and travel times), and *service time* (i.e., sum of the travel and on-scene times).

vious "nonbusy" status, an appropriate response by the officer, and backup assistance if needed.

The communications means by which each of the four functions is accomplished provides a way to distinguish between the functions. Viz.,

Needs identification:	Telephone primarily, but also intrusion alarms, call boxes, radio.
Status Monitoring and Response/Adjustment	Radio, digital, and computer-assisted communications, plus officer interactions both in the communications center and the field. The automatic vehicle monitoring system is primarily used for status monitoring although it relates to the other functions.
Resource management	A management information system fed by data generated in the course of performing the first three functions.

Evaluation Measures'

Figure III identifies in greater detail the four sets of evaluation measures sketched in Figure I. In practice, as might be expected, most available PCCC evaluations are fairly explicit about input measures, and somewhat fragmentary about outcome.

The fourth set, systemic measures, have been overlooked to a large extent in evaluation literature. A program's impact on the immediate organization and on other organizations should be assessed. For example, a management information system may change or be changed by the power structure of the organization.

Further, a long-range perspective is needed in systemic evaluation. The impact of a program can be assessed not only in view of an immediate "before" period but also on a longer horizon. The successful implementation of a computer-assisted dispatching system in a police department which has had no technological innovations during the past 10 years is more significant than an equally successful introduction of the system in a department which has recently implemented a number of other innovations. In examining response time, a 5- to 10-year trend line should be determined, not just the possibly misleading statistics of immediacy.

Cost benefits and productivity must also be assessed. Program results should be matched against those of similar programs. Replica-

Figure III

Program Evaluation Measures

INPUT

- . Program Rationale (Objectives, Assumptions, Hypotheses)
- . Program Responsibility (Principal Participants, Participant Roles)
- . Program Funding (Funding Level, Sources, Uses)
- . Program Constraints (Technological, Political, Environmental, Legal)
- . Program Plan (Performance Specifications, System Design, Implementation Schedule)

PROCESS

- . Program Implementation (Design Verification, Implementation Cost)
- . Program Operation (System Performance, System Maintenance, System Security, System Vulnerability, System Reliability, Operating Cost)
- . Concurrent Programs (Technological, Physical, Social)

OUTCOME

- . Attitude (Citizen (Fear), Criminal, Police)
- . Behavior (Citizen, Criminal, Police)
- . Crime (Detection, Deterrence, Apprehension, Displacement, Level)

SYSTEMIC

- . Organizational (Intra-Organizational, Inter-Organizational)
- . Longitudinal (Input, Process, Outcome)
- . Programmatic (Derived Performance Measures, Comparability, Transferability, Generalizability)

Source: James M. Tien and Vincent O'Donnell, A Guide to Single Project Evaluation Design (Cambridge, Mass.: Public Systems Evaluation, Inc., 1978).

bility of the program should be studied, and the evaluator should consider to what extent the program can be generalized. For example, computer-assisted dispatching probably is not beneficial to police departments below a certain size.

GENERAL ASSESSMENT

The fact remains that there still are only a handful of formal evaluations of the new technologies in PCCC. Let us look at the six applications we have listed and explore what evaluations are available.

Nine-one-one (911)

American Telephone and Telegraph Co. (AT&T) announced in January 1968 that the 3-digit number 911 was being reserved as the universal emergency telephone number. But in so announcing, AT&T effectively began to resign one of its time-honored functions: the forwarding of "0" operator emergency calls to the appropriate agency. In jurisdictions which adopted the "911" system, such calls were to be switched directly to public safety answering points (PSAP's) which were to be staffed by personnel from one or more local public safety agencies.

This brought two critical problems:

- Local government officials are now burdened by implementing and operating an integrated system for processing emergency calls, including requests for police, fire, ambulance, and other emergency service (suicide prevention, poison prevention, gas leakage, etc.). This requires close cooperation between the various local agencies, and typically this is not an insignificant political problem.¹⁰
- The telephone company's central-office exchange boundaries do not usually coincide with municipal or other jurisdictional boundaries.

These problems slowed the adoption of 911 and, today, only one person in four has access to the system. A steady growth in the system's use is predictable in the next decade as the political and technological problems are solved.

Indeed, there even are State laws (as in California) requiring that 911 be implemented. Roger Reinke suggested in 1968¹¹ that if it were possible to automatically identify the calling number, the call could be selectively routed to the appropriate jurisdiction's PSAP. Some advanced 911 systems thus include such options as automatic number identification (ANI) and jurisdictional selective routing (JSR).

Despite a decade's use of 911 systems, the authors have found no documented formal evaluation of either a basic or advanced system. There have been some studies of the merits of alternative 911 configurations and a few informal reports of 911 impacts. Figure IV summarizes the objectives of 911, demonstrating that the system supports only the needs identification function of PCCC. Objectives are divided between basic and advanced 911 systems.

The basic objectives emphasize easier and quicker access to emergency services. However, one widely cited study, Kansas City 1977,¹² found that it took citizens longer to report a crime than either to dispatch police or to travel to the location--and, in fact, reporting time was nearly as long as the combined time for dispatch and travel. Response time was found unrelated to the probability of making an arrest or locating a witness for most Part I crimes discovered after the crime occurred. For those crimes involving a victim or witness, reporting time--the citizen's role--was most important in locating a victim or witness. Another study found that citizen satisfaction with police service was more closely related to the citizen's expectations of response time than to actual response time.¹³

The advanced system involves another set of objectives based on the existence of technical options such as automatic number identification (ANI), automatic location identification (ALI), and jurisdictional selective routing (JSR).

Obviously, most 911 systems are basic systems. No formal evaluation appears to have been undertaken, but it can be safely said that the very act of implementing a 911 system by definition achieves most of its objectives. Of greatest interest are the advanced systems, one of the largest of which was recently implemented in Chicago. It possesses automatic number identification (ANI), automatic location identification (ALI), supplementary dispatch support data (SDSD), and internal selective routing (ISR). Alameda County, California, is to implement a major system whose ANI, ALI, and jurisdictional selective routing (JSR) features are to be evaluated as part of a 3-year, LEAA-funded study on "Dial 911 Systems Assessment and Transfer."¹⁴ Hovey's 1974 study¹⁵ as part of the Alameda County planning effort summarizes key findings on each of the advanced 911 options. Alameda County also discovered little public objection to giving government access to telephone subscriber records, a key privacy issue.

Figure IV

911: Stated Objectives.

BASIC 911 SYSTEM

- To have a simple to remember number for emergency purposes.
- To have a central reporting potential for all emergencies.
- To minimize the time required to dial and to reach an emergency operator.

ADVANCED 911 SYSTEM

- To reduce the number of false fire alarms, bomb threats, and other malicious calls--using the automatic number identification (ANI) option.
- To allow for call-back and address identification in case a distressed caller gives inadequate information--using the ANI and automatic location identification (ALI) options.
- To overcome the fact that the municipal or jurisdictional boundaries do not usually coincide with the telephone company's central office exchange boundaries--using the ALI and jurisdictional selective reporting (JSR) options.
- To overcome the fact that the identified zones or precincts of the emergency response units do not usually coincide with the telephone company's central office exchange--using the ANI, supplementary dispatch support data (SDSD), and internal selective routing (ISR) options.
- To enhance an emergency operator's area familiarity, awareness of resource availability, and identification of redundant calls--using the ALI, SDSD and ISR options.
- To minimize the level of transcription errors--using the ANI, ALI and automatic registered name identification (ARNI) options.

Chicago reports that its 911 system is operative 99.9 percent of the time and that it takes only 0.7 seconds to find the address of a calling number and route the call to the appropriate console.¹⁶ Mainly, however, only input measures have been studied in the analysis of 911. These have been most extensively studied with regard to Alameda County, and the Alameda County investigation should close some of the gaps in the literature with regard to process, output, and systemic measures.

But four important concluding points must be made:

- Nine-one-one cannot be the sole emergency telephone number. Another standard 7-digit number should be made available not only to overcome technical limitations but to permit citizens to make anonymous calls on incidents in which they do not wish to become "involved".
- Despite the Alameda County survey, doubtless a court must someday resolve the privacy questions raised by advanced 911 options. A legal study should be funded by the Federal Government in this area.
- Although the 911 concept is sound, it is perhaps an excellent example of a concept which should have undergone intensive technological assessment before implementation. Not only might advanced solutions to technical and political problems have been developed, but background and training requirements for 911 operators could have been identified. The Police Commissioner of New York City recently transferred 121 civilian operators who were "unqualified or unsuited to cope with the demands of the stress-filled job."¹⁷
- The fact that 911 serves not only the police but also fire, ambulance, and other emergency services could eventually lead to a combined public-safety command, control, and communications system. Since PCCC systems are technologically more advanced than those of other emergency service, future PCCC systems should be flexibly designed so as to interact with or accommodate the needs of other emergency systems.

Computer-Assisted Dispatching

CAD systems bring together many of the technological innovations in PCCC through the partial automation of the call-answering, processing, and dispatching activities of the communications center. CAD matches the address of the call for service and the police beat to a computerized geographic file, and instant recall of dispatch data is possible. Also, a CAD system might include an incident records file with a log of each incident from the first call to final disposition,

a file on the status and availability of field units, a geocoding address to corresponding beat number, a file which tracks personnel status and assignments, and a dangerous-situation file which records addresses with a "dangerous history" and lists dangerous persons or places.

The rate of CAD implementation has been slower than expected. The Colton survey of 1971¹⁸ indicated that 61 departments would install CAD within 3 years. A second survey in 1974¹⁹ revealed only 15 departments with systems nearly operational. Thus, we find CAD in an early stage of development, with departments finding it difficult to implement. A study by the Jet Propulsion Laboratory in mid-1975 revealed that only 10 percent of the 135 police departments in jurisdictions of more than 100,000 population had CAD, and many of these systems were new.²⁰ Recent years have brought a number of advancements and it now appears there are around 50 CAD systems in the country, although they differ significantly in hardware, software, and type of system.²¹

Most of the available literature is city-specific and written by a member of the city's police department rather than by an independent evaluator. Little of it compares CAD among cities on a regional or national basis: In essence, there have been no formal evaluation studies.

The stated objectives of CAD are summarized in Figure V and they reflect the range of functions encompassed by a CAD system. Most of the objectives focus primarily on the status monitoring and response/adjustment functions of PCCC. However, CAD also has the potential of contributing to resource management and as such several of the objectives relate to this function.

Many CAD activities are routine: the tasks the computer performs are straightforward and repetitive; the computer simply carries out tasks formerly done manually. Some argue that CAD has simply automated earlier manual processes without taking advantage of the machine's computational power.²² But when CAD begins to touch on resource management, adjusting or changing dispatch and interaction between field personnel, it may be nonroutine and as such may pertain to the police administrator's control and management.

Many of the CAD systems thus far installed have been in cities and counties of more than half a million people such as New York,²³ Indianapolis,²⁴ Phoenix,²⁵ and San Francisco.²⁶ But cities and counties of much smaller size have established CAD systems: Huntington Beach,²⁷ Virginia Beach,²⁸ Las Vegas²⁹ are among them. Smaller cities have also joined together to provide regional CAD systems.³⁰

Results have been uneven, it appears, although the number of systems continues to grow. Systems in San Diego³¹ and New York City³² seem quite successful, for example, but those in Boston³³ and Denver³⁴ have met major difficulties. However, the question

in most police departments no longer is whether there will be computer-assisted dispatching, but when.

Few evaluations have been done, and those available focus on input and process measures. Practically no literature is available on outcome and systemic impacts. The information available, however, helps establish a framework for future evaluation.

Most available literature is city-specific and offers little to further a nationwide evaluation literature. Sohn et al., however, present a detailed examination of the tradeoffs involved in selecting CAD system components, guidelines for choosing among these tradeoffs, and potential measures of effectiveness.³⁵ From the city-specific literature, however, it is unclear whether such guidelines have been followed. Each city seems to focus on its own view of what measures are appropriate and there is little consistency in evaluation criteria. A common evaluation framework is needed for future comparative analysis.

Most CAD documents are descriptive only, outlining the features of the system in a particular city and its anticipated objectives. Even for input measures, there is little of an evaluative or comparative nature. Only a few articles present cost data, and even then the data are not comparative. Better nationwide cost information is needed and a cost model developed so that the cost of various systems can be compared in terms of, for example, hardware, software, computer consultants, engineering services, and project management costs.

Few cities have detailed system performance specifications before implementing CAD. San Diego, an exception, outlined such detailed specifications that it was able to require satisfactory performance, even though it meant the vendor had to wait an extra 8 months from the time the system was in operation until payment was made.³⁶ In contrast, Denver has been working on CAD implementation since 1975. Because of inadequate planning and failure to outline contract specifications, that city's CAD efforts are still in doubt.³⁷ Boston has had similar problems.³⁸

As for process measures, CAD appears to have made better information available in the command and control process. This has been documented in a number of cities (e.g., Huntington Beach, Indianapolis, Las Vegas, New York City, Phoenix, San Diego, and Virginia Beach). In some places (e.g., New York City; Palm Beach County, Florida; and San Diego) citizen calls are automatically distributed among telephone operators, resulting in a more even distribution of workload.

However, although it seems clear that better information is available with CAD, there is little comparability between the findings in the literature. It also remains to be answered whether this information is actually used and whether CAD improves services delivered to the public.

Figure V

CAD: Stated Objectives

- To monitor and display better data regarding the PCCC system, including complaints received which are still outstanding, status of all patrol units under the dispatcher's control, and actions taken by patrol units.
- To assure the availability of better information to all personnel in the PCCC process, including telephone operators, complaint evaluators, dispatchers, and supervisors.
- To improve officer safety by more effectively monitoring the status of officers and responding more rapidly in case of emergency.
- To recommend potential patrol units for assignment to a call for service--this recommendation would be based on some algorithm, including, for example, the assignment of the patrol unit which is estimated to be closest to the scene of the accident.
- To improve the response time of the dispatch process, including the time to answer the telephone; the time for the operator to process the call; the time for the dispatcher to assign a unit to the call; and the time for the officer to travel to the scene of the incident.
- To improve the police service to the public and to improve police effectiveness in regard to criminal activities, including, for example, improving the apprehension rate by responding to the scene of a crime more rapidly.
- To provide automatic or easy access to remote files, including outstanding warrants, stolen property, State and National police patrol and inquiry systems.
- To improve the quality of the data maintained as a part of the dispatch process, including, for example, automatically assigning a case number, date, and time to each incident that is dispatched; maintaining the status of records as units are assigned and cleared; and checking input data for possible duplicate cases and invalid addresses.
- To more effectively manage police resources through the use of better data and through a better understanding of the command and control process--the CAD related data provide the potential for better decisions and more effective management.

PCCC response time is one area of service performance where some benefits have been documented. A survey in San Diego indicated that before CAD the emergency telephone sometimes rang 80 or 90 times before being answered. Now 90 percent of the calls are answered within 25 seconds, and operator talk time has been reduced to 60 seconds.³⁹ Phoenix reports improvements,⁴⁰ and Huntington Beach claims a communications response time reduction of an average of 45 seconds a call.⁴¹ In all of these reports, however, it is extremely difficult to judge the quality of the response time analysis. In many cases, reliable before-and-after figures are absent and estimates tend to be somewhat subjective based on a very limited time analysis.

Furthermore, we noted earlier (and other studies indicate) that decreasing response time may result in very little real impact.⁴² Only in Phoenix did officials claim that their computer-assisted system had helped reduce the crime rate, and they hedged by saying that the 2 percent reduction may have been the result of other factors.⁴³

As to outcome measures, little information is available. No surveys of citizen attitudes toward CAD were reported in the literature, and there were no indications of changes in police behavior. Most articles indicate that police attitudes have been positive toward CAD, especially when police departments have been sensitive to the needs of individuals.⁴⁴

Systemic measures have also received little consideration in the literature. Again it is a question whether the additional information CAD delivers is actually being used. In San Diego, for example, a large number of reports are generated, but there is no regular schedule of distribution. In addition, the San Diego system seems to have generated an "information overload." New York City had a similar problem with too much information.

Essentially no costbenefit analyses of CAD appear to have been made, although several documents discuss how to perform an overall evaluation of CAD⁴⁵ and factors involved in a costbenefit analysis.⁴⁶ Huntington Beach projected more than \$29,000 in manpower savings for their first year of CAD,⁴⁷ but little documentation was given on how these savings were calculated.

Finally, regarding the transferability of CAD systems, it appears that a number of systems are in operation around the country, but little actual transfer has occurred. In fact, except for some verbal references regarding the transfer of the Charlotte, N.C., CAD system, there is little discussion in the literature regarding the transfer of CAD from one department to the other, and this is an area which seems to require further evaluation and analysis.

In conclusion one must note that CAD has spread more slowly than expected, but that in recent years the number of systems has increased rapidly and continues to do so. Thus it is important

to develop a nationwide comparative data base regarding the use of CAD systems. At present time such a base is not available, and to the extent the kind of information it should have does exist, it generally focuses on input and process measures, failing to consider outcome and systemic impacts. Several conclusions seem appropriate:

- A comparative evaluation is essential to understand the actual results and impact of CAD, but it must be based upon a common framework applicable to all cities and departments.
- Performance guidelines are necessary at the beginning of each project, and the relationship between CAD user and vendor must be clearly defined.
- CAD implementation requires more than technical expertise. It is where there appears to be a sensitivity to human and behavioral factors that systems seem to be working.
- The potential seems to exist for better police management through the use of CAD-generated information. However, the ultimate impact will depend on the ability of law enforcement administrators to analyze and use this information effectively. Police chiefs have not often considered themselves managers, but as balancers of pressures and promoters of police resources. Thus it remains uncertain whether they will be able to channel the potential technological talent of the computer to do more than routine operations.

Automatic Vehicle Monitoring

The President's Commission in 1967 cited studies suggesting that automatic vehicle monitoring might achieve cost-effective reductions in response time and even improve apprehension rates. An AVM system provides a police dispatcher with real-time location estimates of each vehicle in a fleet and provides additional information on status or mission. (An automatic vehicle location (AVL) system provides only location estimates without status information.) The monitoring systems often are linked to a CAD system.

Different methods have been developed for AVM, compared in a number of references.⁴⁸ Four methods are generally discussed:

Navigation (hyperbolic) systems: Principal AVM application of this type is Lorraine C, used for nearly 20 years in navigation of ships at sea. Utilizing radio location techniques, shifts in location are identified by

keeping track of the arrival time of patterns of pulses emitted from various transmitters.

Trilateration systems: Such systems also utilize radio location with three or more fixed sites used to determine the vehicle distance from each site by measuring radio signal travel time. A computer then uses this data to locate the vehicle.

Signpost-proximity systems: Electronic signposts are positioned in fixed locations throughout an area and a vehicle is located when it comes within the proximity. Signposts can either be a receiver or, in most cases, a transmitter.

Dead-reckoning systems: Much like the inertial guidance system used in missiles, aircraft, and submarines, this system requires that the vehicle's starting position first be put into the computer. Then instruments are used to track the vehicle's distance and direction. The computer translates these into the vehicle's location using an advanced geocoding system.

Development of commercial AVM systems was not begun until the late 1960's, and only a handful of cities have installed police AVM systems. The first, in the early 1970's, was in Montclair, Calif., a city of approximately 30,000. Montclair utilized a signpost system where the transmitter radiated the signpost identification to headquarters. An evaluation was completed in 1976.⁴⁹ However, the system is no longer in operation.

St. Louis was the first urban department to implement fully an AVM system. A computer-assisted dead-reckoning system was established experimentally in one police district in 1976 and citywide in 1977. Evaluations were conducted of both phases.⁵⁰ The Dallas police are beginning to implement a trilateration system.⁵¹ It has been installed in one district and an evaluation is underway.

A few other cities tried AVM. Stamford, Conn., had a signpost system in which the post was a receiver, picking up the identification of the vehicle and forwarding it to headquarters. It is no longer in operation. Orlando, Fla., considered establishing an AVM system,⁵² and Huntington Beach now has a signpost system.

Based on the few evaluation results available, it is clear that results have been mixed, although the potential has been identified. Figure VI summarizes the primary objectives.

AVM is designed primarily to support the status-monitoring function of PCCC. Primary focus of much of the early literature on AVM has been on the reduction of response time, both in reduced dispatch time (because the dispatcher could make a quicker decision) and reduced travel time (because the closest car would be assigned).⁵³

Some hypothesized that reduced response time would increase apprehension rates.⁵⁴ It was anticipated that officer safety would be improved, since an "officer in trouble" call, with its location, could be sent merely by pushing a button.

However, several AVM objectives also relate to the response/adjustment function, through dispatch efficiency and improved tactical command and control.⁵⁵ For example, with an AVM system, the dispatcher would be able to supervise directly the movements in a high-speed chase or a bank robbery, through rapid response and adjustment using real-time location information.

AVM objectives also may relate to resource management through patrol effectiveness and improved supervision. However, many of the benefits related to these objectives are subjective; that is, they are very hard to measure.

The final AVM objective relates to reducing voice-band congestion and improving police communication. This becomes possible when AVM is linked with some kind of mobile digital communication (MDC) system, as discussed later, although such an objective is not limited to AVM systems.

Reported Measures

Despite the fact there have been only two actual evaluations of AVM systems, a number of documents have discussed the potential benefits of AVM and some of the possible measurements of its effectiveness.

Input measures vary depending on the rationale for the program, the method of location selected, and program design and constraints. Additional input measurements include the nonrecurring costs for planning and implementation, system performance specifications, and variables for system planning. Two references⁵⁶ particularly focus on system planning and present guidelines for cities considering AVM. For example, they point out that the required accuracy of an AVM system may vary from 100 to 1,000 feet depending on the purpose and objectives of the system.⁵⁷ Several sources suggest that the full benefits of AVM can only be achieved when it is used in conjunction with a computer-assisted dispatch system.⁵⁸

As for process measures, response time has been heralded as one of the primary benefits of AVM, but the literature indicates disagreement regarding the potential benefits. Doering⁵⁹ estimates a 3 percent reduction in response time. Ritter⁶⁰ estimates 7 percent. Based on a simulation model, R. Larson et al.⁶¹ indicate that under ideal conditions the reduction could be as much as 11 to 15 percent.

Outcome and systemic measures are intended to look at the actual impacts of AVM systems. For outcome, the evaluator focuses

Figure VI

AVM: Stated Objectives

- To reduce response time--through a reduction in both dispatch and travel time.
- To increase apprehension rates--through a reduction in response time.
- To improve officer safety--through continually monitoring the status and location of police vehicles.
- To improve dispatch efficiency and coordination--by providing the dispatcher with precise data and allowing for more effective coordination.
- To improve tactical command and control--through the online direction of such special tactical events as high speed chases, bank robberies, emergency deployment, and support for covert operations.
- To improve patrol efficiency and effectiveness--through the availability of direct information regarding the location and allocation of the patrol force, and through the indirect realization of patrol officers that they are being monitored.
- To improve supervision of the patrol force--through the better online supervision of officers in the field and through the use of the management information generated from the AVM system.
- To reduce voice-band congestion--when AVM is linked with some type of mobile digital communication (MDC).

on police and citizen attitudes, police and citizen behavior, and overall impact on crime. Although police attitudes can be assessed, changes in behavior, performance, or both are far more difficult to measure, and the crime-related impact is almost impossible to identify.

In both the Montclair and St. Louis evaluations,⁶² the systems were found to have encountered difficulties in system performance and reliability. In Montclair, overall reliability was estimated in the 80 percent range with the mobile unit transceiver presenting the lowest reliability. Such problems no doubt influenced officer attitudes, and the system no longer is in operation. In St. Louis, using dead reckoning, the greatest difficulty has been "lost cars." During the Phase I test period, cars had to be relocated or reinitialized in the system at a rate of 11 cars per day, or an average of once every 2.2 hours.

Contrary to earlier expectations, response time reductions do not appear to be a primary benefit of AVM. In St. Louis, a careful evaluation of before and after response times was made. Phase I showed modest but hardly significant improvements in travel time; Phase II, citywide, showed response time on the whole somewhat longer. (There were many reasons other than AVM, such as less experienced dispatchers.)

Similar conclusions regarding response time were also found in Huntington Beach. A sampling of response time data before and after the installation of the AVM system was made for both "dispatcher processing time" and "travel time." Data were taken for operations over two 1-month periods, 1 year apart, in order to minimize the effect of seasonal variations in patrol activity. Based on the study, "no significant change was found to have occurred in response time before and after the installation of the AVM systems."⁶³ In Montclair, some response time benefits were noted in 1974, but the findings were based on a relatively small sample. Also, in a city the size of Montclair, response time does not appear to be a significant difficulty and McLean⁶⁴ suggests the benefits of improved response time seem minimal when compared to the cost.

Officer safety results have also been disappointing so far. Poor accuracy was the major contributor to loss of officer confidence in the St. Louis system, and in emergencies officers now prefer to use the voice radio in combination with the AVM emergency button.

Evaluation concerning AVM's potential to improve police supervisory capabilities and police productivity is still lacking. The St. Louis study concluded that "reasonable levels" of productivity improvement could bring a very attractive return on investment, but the key is to establish a link between AVM and such returns. Precise measures regarding changes in behavior are very difficult to document.

Regarding reductions in voiceband congestion, favorable results were reported in St. Louis. However, the primary benefits appear to come not from uncluttering the voice band but from the fact that digital transmission offers the officer instant access to the communications center. In a smaller city such as Montclair, congestion had not been a problem, so AVM offered few benefits in this area.

A number of conclusions can be reached in spite of the fact AVM has been established in few cities and evaluated in only two.

- There is a close tie between the technical performance of the system and the overall results and attitudes toward it. St. Louis and Montclair both experienced a number of technical and operational difficulties which led to a decline in officer attitudes toward the system. The lesson seems to be again the need for clear performance specifications to the vendor.
- The expected response time benefits were not achieved. If AVM is to be justified, other benefits must be shown. Most promising appear to be tactical control and police productivity, but further evaluation and experimentation is essential.
- Such a system involves such behavioral impacts that failure can be expected without careful training and officer involvement.
- There is an important link between AVM and other PCCC applications. In St. Louis, AVM was installed without a CAD system and thus required such manual operations as manually placing a cursor to identify the location of an incident. In Dallas, the CAD system has been operational since 1969, and this may help more effectively to implement AVM. Dispatch personnel trained to the ways of the computer may also be less resistant to AVM.
- Further evaluation is essential.

Mobile Digital Communications

Figure VII shows some typical hardware and related elements of a mobile digital communications system. It encompasses the traditional police radio system. Those elements which are exclusively associated with MDC are briefly summarized:

- Mobile digital terminal (MDT) is the input-output device. Depending on the sophistication of the system, it could be a simple set of lights and status keys, or a general alphanumeric keyboard with a hard-copy printer, a cathode ray tube display, or both.

- Encoder is a device which converts a message into computer language, automatically adding necessary identification and other routine information.
- Decoder is a device which converts the computer language back to displayable words and numbers.
- Mode select is a device which determines whether the message is digital or voice and switches it accordingly either to the encoder-decoder or the speaker-microphone. Priority is given to voice transmissions.
- Minicomputer controls message switching and MDT displays. Depending on the sophistication of the particular system, its software could be expanded to allow automatic queries of remote data bases, computer-assisted dispatching, message logging, traffic data collection, etc.
- Modem, an acronym for modulator-demodulator, permits computers to communicate in computer language over transmission channels also used for voice. A modem is required at each end of the transmission line.

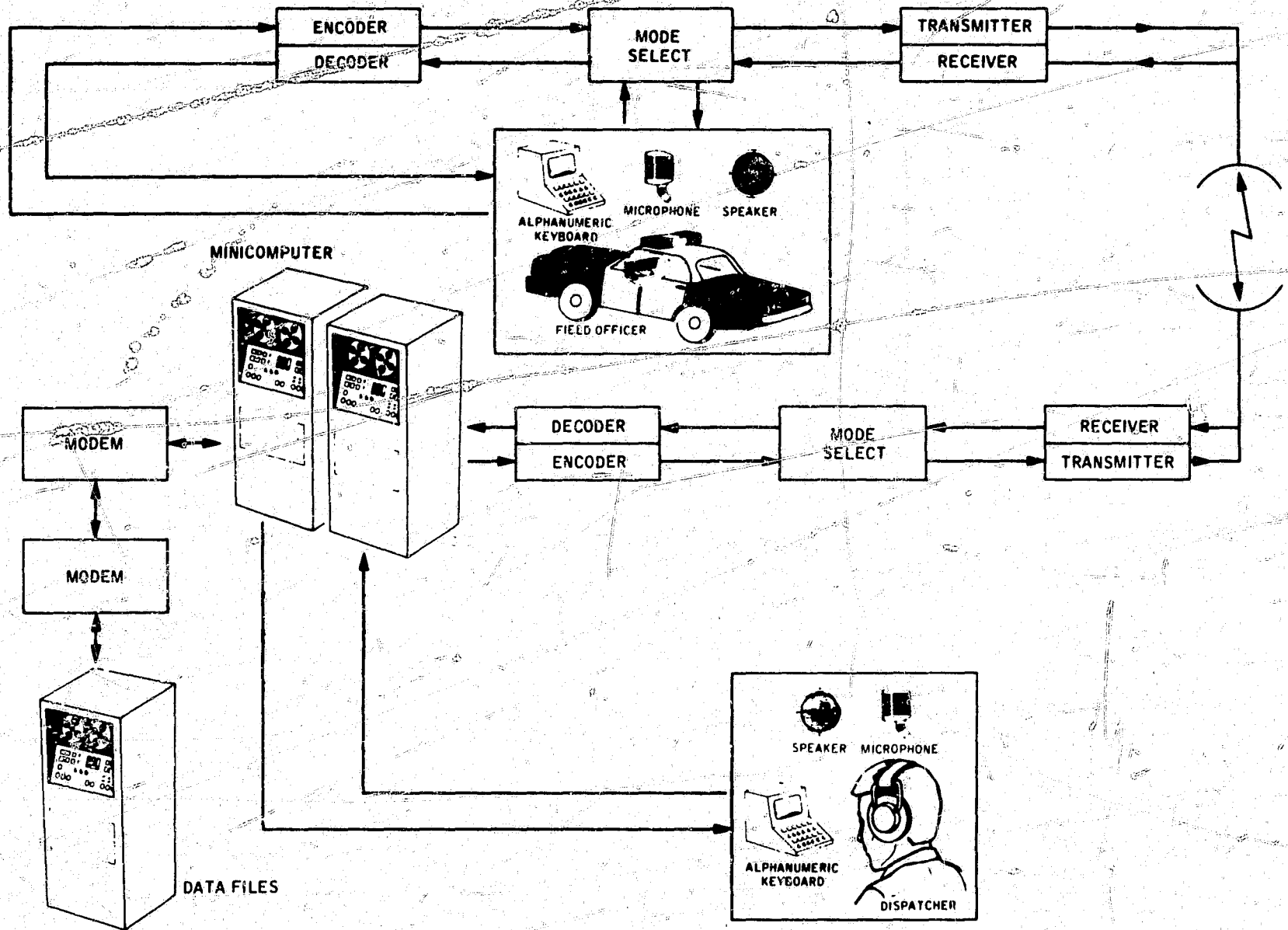
The MDC system in Figure VII is actually quite a sophisticated system. For example, a system permitting officers only to report their status would not require a minicomputer. These are options in MDC sophistication:

- Message capability: status-only v. full-text v. full-text plus automatic data-base query.
- Transmission flexibility: mobile unit only v. mobile unit combined with out-of-car digital transmission capability.
- Communication mode: simplex, half-duplex, or full duplex. In simplex, all transmissions use the same frequency and if either any patrol unit or the base is transmitting, no one else can. In half-duplex, the base can transmit and receive simultaneously, but patrol units use simplex. In full duplex, all can receive and transmit simultaneously.
- Mobile input-output device: hard-copy printer, cathode ray tube, or both.

Military application of MDC date back many years, but in police work only to the early 1970's. Of nine police applications of which we know, most have been quite limited in scope, experimental in nature, and undertaken, at least initially, with LEAA funding. There have, however, been several studies of the potential of MDC in law

Figure VII

Typical Elements of a Mobile Digital Communications System



317

enforcement⁶⁵ as well as somewhat formal evaluations of three test systems: Oakland, Minneapolis, and Hennepin County, Minnesota.⁶⁶

Congestion of police radio frequencies has been the primary reason for adoption of MDC. The National Advisory Commission on Criminal Justice Standards and Goals⁶⁷ predicted in 1973 that MDC could reduce congestion by 40 to 50 percent. A second major reason for interest in MDC is that many law enforcement data bases are computerized, and a patrol unit with digital communications (and a suitable switching unit, a minicomputer, at base) can automatically access these data bases.

These two objectives, however, are somewhat contradictory. Transmitting digital rather than voice signals may alleviate frequency congestion, but providing automatic access to remote data bases would tend to increase such inquiries and aggravate congestion.

A list of MDC objectives appears in Figure VIII. They show that MDC supports both the status monitoring and response/adjustment functions of PCCC. But though all implemented MDC systems have status monitoring potential, only a few are capable of effecting such response/adjustment functions as the direct dispatch of patrol units using only digitally transmitted instructions.

Sohn et al.⁶⁸ indicate some problems in realizing the potential benefits of the objectives:

- Use of MDC can create significant interference on existing voice links; in many cases it is desirable to use dedicated channels.
- Digital links are susceptible to significant error rates in urban telecommunications environments, and the repeated transmissions or elaborate detection codes thus required detract from the goal of short transmission times.
- Installing bulky digital terminals in patrol cars may detract from the convenience of their use and possibly officer safety.
- High cost and specialized maintenance can be a burden.

Other problems, such as the high voltage and explosion danger of a cathode ray tube terminal, have been overcome. But the problems, especially cost, have slowed MDC applications.

The following review of actual MDC experience concentrates on the more sophisticated applications. Status-only MDC systems are not considered. (The latter are usually a small part of a larger

Figure VIII

MDC: Stated Objectives

- To reduce radio air time and expand the message handling ability of a radio channel--using digital signals which have a higher transmission rate than voice signals.
- To improve message security--using digital signals which are harder and/or costlier to decipher or monitor than voice signals.
- To improve message accuracy resulting in less need for repetition--using mobile digital terminals, especially if they can provide hard copies.
- To allow for selective routing of messages on an "as need to know" basis--using mobile digital terminals which could be addressed either collectively or individually.
- To allow for unattended message reception--using mobile digital terminals which could record the transmitted message(s) while the officer is away from his vehicle.
- To increase officer safety--using a timed alarm key and/or using mobile digital communication which facilitates and increases information flow, including information concerning possibly dangerous situations.
- To increase officer effectiveness--using mobile digital communication which facilitates and increases remote data base inquiries, resulting in potentially more "hits" (i.e., apprehensions and recoveries).
- To increase dispatcher effectiveness--using digital signals which could relieve the dispatcher from routine data inquiries, patrol status updates, and message repetitions, as well as dispatches of certain noncritical calls for service.

automatic vehicle monitoring system, as in St. Louis⁶⁹ or a more sophisticated radio system, as in Orange County, Calif.⁷⁰)

The existing evaluations show that the objective of reduced radio air time has not been achieved. Not only did the number of calls to remote data bases increase, but "channel time, made available when a portion of the voice load was converted to digital..., was immediately replaced by...messages which previously could not be transmitted due to...congestion."⁷¹ Also, digital transmissions create interference on voice transmissions. Oakland at first tried to have a combined digital and voice channel, but had to add a digital-only channel to overcome interference. Hennepin County and Minneapolis also used a separate digital channel. Thus, instead of saving radio air time, the interference problem may require a whole new channel. Technically, and at moderate costs, it is possible to minimize the interference problem.

Another interesting finding was that on officer effectiveness. When a patrol officer receives data-base information giving him probable cause to make an arrest or recover a vehicle, it is termed a "possible hit." Oakland found that its MDC-equipped patrol units averaged 14.5 times as much "possible hit" information as its non-MDC units, resulting in 2.8 times as many actual warrant arrests and vehicle recoveries.⁷² On a relative basis, however, MDC units are less efficient by 20 percent. The easier access to remote data bases has caused a greater number of inquiries without a corresponding increase in "hits."

In regard to officer effectiveness, it is unfortunate that the evaluations did not report on the human factors in MDC application. Sgt. Marin of the Los Angeles County Sheriff's Department writes that "license numbers can be exceedingly difficult to compose while also trying to drive through traffic...[and typing] requires the officer to divert his attention from the suspect and this could have fatal consequences."⁷³

Oakland purchased its 32-unit MDC system for \$149,980 with an associated yearly maintenance cost of \$49,815 or about one-third of the initial purchase price. This is indeed a high maintenance cost; it should decrease as the MDC technology improves. Despite the MDC costs, Oakland felt the system cost-effective.⁷⁴

Three points must be made in conclusion:

- MDC does not decrease air time and may increase it.
- The impact of MDC inquiries of remote data bases must be carefully assessed. The data bases' facilities may be overwhelmed as more and more patrol units can automatically access them.

- All three evaluation studies reviewed in this section base their conclusions on small samples of data, sometimes less than one week's worth. Their objectivity is suspect in that all were conducted by police personnel who were closely associated with the system studied. The Minneapolis and Hennepin County studies are only 5 and 12 pages, respectively, and must be considered "process" evaluations at best. Although the Oakland study can be considered an "impact" evaluation, it lacks depth in methodology. It tabulates first-order results but rarely derives second-order, performance figures. Finally, all three studies are based on systems implemented and tested prior to 1976.

An intensive impact evaluation of an upcoming or ongoing MDC application is needed now to update available information.

Management Information System

The four applications already discussed primarily support the tactical PCCC functions. Although they can provide necessary data to support the strategic PCCC function of resource management, the experience to date has been dismal: the computer-based data from these applications have been only intermittently analyzed, if at all.

What is missing is a computer-based management information system which could automatically and continually integrate the tactically generated data and analyze it to derive results that could be used to make strategic resource management decisions.

We know of no such PCCC-related systems today.

There are several systems for general police management or crime analysis, but they typically are limited in scope and not linked to PCCC functions. For example, although the Virginia Beach Police Department has both a CAD system and an Automated Police Management System,⁷⁵ the two computer-based systems are not linked. The management system is limited to providing information on personnel, budget, vehicular maintenance, uniform inventory, evidence control, and found property.

A general MIS should also perform functions other than resource management. In order to assist in resource management, however, it must include resource allocation algorithms which could identify appropriate allocation alternatives. There are several resource allocation models which have been tested and intermittently used for resource management decisions, though not yet PCCC-related.

Only three of the police-related models identified by Sohn and Kennedy and by Chaiken⁷⁶ are in current use. They are the Law Enforcement Manpower Resource Allocation System (LEMTRAS), the Patrol Car Allocation Model (PCAM), and the Hypercube Queuing Model.

LEMTRAS was marketed by International Business Machines (IBM) in the late 1960's.⁷⁷ Its computerized solutions to workload problems are basically deterministic: fixed answers to fixed problems. This creates a number of drawbacks because neither calls for service nor responses to such calls are deterministic in nature. PCAM and Hypercube are more realistic and probability-based, and IBM has withdrawn LEMTRAS from the market. Several police departments continue to use it, however--probably because it is relatively easy to use and, in fact, reflects the way⁷⁸ police departments have traditionally allocated their resources.

PCAM (1972) and Hypercube (1975) both were developed by Richard C. Larson, the former having been later refined and computerized by Chaiken and Dormont.⁷⁹ Although they overlap in some respects, the two models are actually complementary. In general, PCAM determines a time allocation of patrol units and then Hypercube is used to determine the units' spatial positioning during different periods. It was, for example, in this manner that the Wilmington split-force patrol experiment was designed.⁸⁰ The models can only assist the police administrator charged with making resource allocation decisions. Many variables cannot be modeled and the administrator must make final decisions.

Chaiken⁸¹ has identified some 35 agencies which have received copies of the PCAM program; a similar number have received Hypercube. However, the authors have been unable to find any agency which has integrated PCAM, Hypercube, or both with its CAD or AVM systems, so that the generated data can be automatically fed into either or both of the management models. Probably, more agencies soon will access the models, since NILECJ recently funded The Institute for Public Program Analysis to make the models available on microcomputers and programmable hand calculators.

There is a need to develop an MIS which integrates with the other elements of a PCCC system, as is detailed later in this paper.

An encompassing resource management program must not only allocate police units on a temporal and spatial basis, but also schedule them. Heller and Stenzel⁸² have developed a pertinent scheduling algorithm, but it requires further refinement. The degree to which computer-based resource allocation and scheduling algorithms can be effectively used by decisionmakers must remain an area of continued research and evaluation.

Regional Communications System

Regional communications systems are developed for different reasons. In Muskegon County, Michigan, a frequency congestion and interference problem led to a central police dispatch system. Before, eight dispatch centers in the County competed for use of a single radio frequency, leading to garbles and frustration.⁸³ An interference problem also led Orange County, California, to adopt a computer-controlled regional radio and status-only MDC system.⁸⁴ On the other hand, the Boston Area Police Emergency Response Network (BAPERN) was developed because of the need in certain emergencies to coordinate police from the many geographically small jurisdictions.⁸⁵

Cost is another reason. An RCS could result in cost savings, a mechanism for securing Federal and State subsidies, or both. California law provides a tax incentive for jurisdictions to consolidate their services; Santa Clara County was advised to take advantage of this in its effort to develop a coordinated emergency services communications system.⁸⁶ Record-keeping costs led Erie County, New York, to develop the Erie Municipal Police Information Retrieval Enhancement (EMPIRE) system.⁸⁷

An RCS does not necessarily mean a central communications center. The proposed Alameda County 911 system, for example, will have a number of public safety answering points.⁸⁸ Similarly, the Boston Area system is still dependent on the communications centers of its 23 police agencies. What BAPERN has provided is a common radio channel as well as a formal, agreed-on set of procedures for coordinating its agencies' actions in times of emergency.

The stated objectives of an RCS are:

- To decrease planning and operating costs.
- To diminish radio channel congestion.
- To provide a means of coordinating police actions during certain emergencies.
- To establish a joint entity which would be more able to attract Federal and State subsidies.

But a number of issues must be resolved:⁸⁹

- A new legal entity, separate from any of the participating communities but under their control, may be useful. Frequencies can be assigned only to a legal entity and it is easier to assign Federal grants to a single body.

- Procedures for financing and governing must be agreed on, a difficult process because the different communities will have different sizes, tax structures, and crime rates.
- Centralized dispatching should use civilian personnel, both for economy and because the new agency probably will not have authority over sworn personnel. The change in dispatching personnel is likely to cause personnel discontent.
- Each community must retain command of its force.

Perhaps the only evaluation that addresses RCS is the Mu County study⁹⁰ which reviews an RCS which was not computer-based. The key finding was that the 8 member agencies formerly had assigned 19 police officers to dispatching and that after consolidation 13 civilians were able to meet the task. This represented a 32 percent saving in personnel time and a 42 percent saving in personnel cost.

Although there is a lack of supporting evaluations, the RCS concept seems sound. It is well-known that the sharing of resources can result in benefits and efficiencies. Elsewhere in this volume, however, Roger Parks suggests that there is little evaluative evidence that consolidation of police departments is efficacious. And practice of the combinative principle has shown that benefits and efficiencies can accrue up to a certain point after which there are diseconomies.

What is the optimal size of an RCS? Obviously, only careful future systemic evaluations can tell us.

FUTURE CONSIDERATIONS

Figure IX summarizes the reported impacts of the systems we have discussed upon the four PCCC functions, subjectively defining three categories: significant, moderate, minimum. The needs identification function is, as expected, significantly impacted by 911, but it should also be significantly impacted by CAD (which can better record and verify the needs), MDC (which can be used to transmit needs identified while on patrol), and by MIS (which, through analysis of historical data, could better identify and carefully record those characteristics of need important for response purposes). The status-monitoring function is the most impacted, but it still should be impacted by MIS--again through the MIS analysis of historical data. Similarly, the response-adjustment function could benefit from MIS analysis. The resource management function is the least impacted function: a PCCC-oriented MIS could automatically analyze the 911, CAD, AVM, and MDC data with tested algorithms. In sum:

Figure IX

Reported Impacts of Specific Applications on PCCC Functions

Specific Applications ¹	PCCC Functions			
	Needs Identification	Status Monitoring	Response/Adjustment	Resource Management
Nine-One-One	Significant			
Computer Assisted Dispatching	Minimum	Significant	Significant	Minimum
Automatic Vehicle Monitoring		Significant	Minimum	Minimum
Mobile Digital Communications		Significant	Moderate	
Management Information Systems				Minimum

¹Regional Communication Systems are not included since they differ widely in focus.

The PCCC functions of needs, identification, status monitoring, response adjustment, and resource monitoring must be further supported before an effective PCCC system can be realized. What research and evaluation are required?

Research

PCCC must be integrated, proactive, flexible. Research must be undertaken to develop systems which are all of those. It is assumed such a system would be computer based.

The PCCC applications reviewed here not only are relatively distinct in focus, but have been implemented only on an independent basis and in an independent manner. The authors know of no police agency which has implemented all the various PCCC applications. Figure XI identifies an integrated system. Briefly, it is indicated that 911 information is entered into the CAD computer, which also (1) interacts digitally with the AVM and MDC systems; (2) serves if necessary as a switching machine for MDC inquiries to remote data bases; (3) continually provides data to the MIS element. It may be cost-effective for the MIS to share with the other systems the same CAD computer.

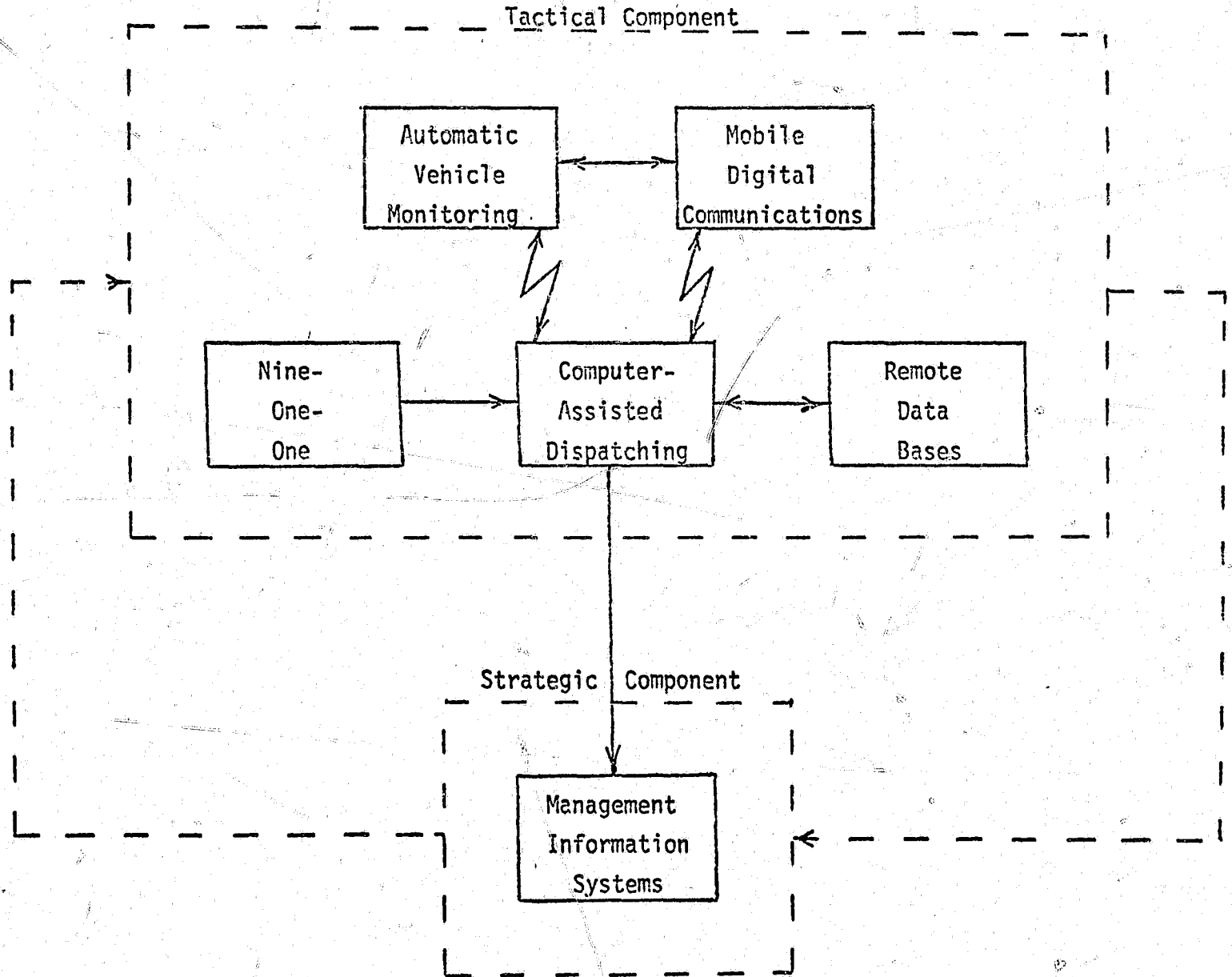
It is important to recognize that there are two components to a PCCC system. All but one of the elements are tactical in function; the second component--comprised of MIS--primarily supports the strategic function. The two components are interactive, in fact, synergistic. A study is needed to assess whether the integrated system is greater than the sum of its parts.

Police departments have to face budgets; none can implement a total PCCC system at once. But each element should be added with the idea that other elements, to be added later, should link with the earlier. CAD can furnish the host computer for other elements. Thus, CAD should be the first PCCC application, followed perhaps by 911, AVM and MDC, and MIS. Proper sequencing bears costbenefit. As earlier noted, implementing AVM before CAD requires manual posting of cursors. Dispatch personnel trained in the ways of CAD have learned to live with a computer, which makes AVM easy.

Except for basic 911, CAD is the most predominant of PCCC applications. In theory, police agencies which have installed CAD can expand it to an integrated PCCC system. In practice, system-compatibility problems develop. Therefore, a detailed system plan should be developed before any PCCC element is implemented.⁹¹ It is interesting to note that in a 5-phased plan, Los Angeles implemented CAD in phase 4, after MDC and before AVM.⁹²

Figure X

An Integrated PCCC System



Although the depicted system of Figure X is conceptually relevant for all departments, not all agencies require such a sophisticated system. Some very small departments may require only an automated telephone-answering system which would allow officers to answer calls from their patrol cars.⁹³

Proactive PCCC

So far, computer-based PCCC has largely only replaced previously manual operations. The potential of a computer in police work has not been fully realized; it has, however, been recognized.⁹⁴ Johns⁹⁵ argues that a proactive or intelligent CAD could:

- Check the validity of calls for service;
- Efficiently allocate patrol resources in a dynamic manner to account for manpower fluctuations;
- Standardize dispatch assignments;
- Improve police strategies by incorporating strategies which, for example, increase neighborhood identity or provide better emergency response;
- Incorporate complex mathematical models for such purposes as estimating patrol unit location and managing queues of calls;
- Collect statistics useful to administrators.

Some 86 percent of all calls for police service is noncritical. The Kansas City studies showed that citizen satisfaction is a function of expectation. Thus, Tien et al.⁹⁶ proposed that cost-effective methods alternative to that of dispatching a patrol unit can be used to respond to noncritical calls provided the citizens are forewarned and advised. Proactive management of police demand can reduce or shift demand peaks and may even lower the demand level. NILECJ is now financing projects on this subject in Wilmington, Del., and Birmingham, Ala. Whatever management procedures are identified by these projects can, of course, be translated into computer algorithms.

FLEXIBLE PCCC SYSTEM

Modern PCCC is totally dependent on the computer. With computer technology advancing daily, any PCCC system must remain flexible. Changing technology will affect PCCC in at least three ways:

- The decreased cost and increased capability of computers will make PCCC systems available even to smaller departments. Solid-state devices are challenging the traditional storage media of tapes, disks, and drums. The new devices have much faster access times.
- Systems can be decentralized and yet coordinated, using the new technology. Data communication and data base management are performed by a number of small computers connected together.
- Sophisticated terminals facilitate interaction with PCCC. The terminals include intelligent interactive graphics terminals, magnetic card readers, optical character readers, and voice data-entry terminals.

Although vendors will always speak highly of PCCC advances, it is important that these applications be carefully evaluated before widespread adoption.

Evaluation

This paper was to review completed evaluations of new technologies in PCCC. However, there have been very few formal PCCC evaluations. A number of articles have been written about the operation of PCCC applications, but this literature generally focuses on the inputs and operating processes--not, as is needed, evaluation of outcomes and systemic impact.

There have been studies of the relative merits of alternative 911 configurations, but no documented formal evaluation of either a basic or advanced 911 system. A number of articles describe the CAD system of a particular city, and a few discuss the broader elements of establishing a CAD system and conducting a project evaluation. But there are no national or regional comparisons and, in essence, no formal evaluations reported.

Three evaluations of AVM have been noted. But both systems were hampered by technical difficulties and personnel attitudes were affected. A complete evaluation thus was impossible. In MDC, there have been several studies of its potential and three fairly formal evaluations. Here again, however, the studies focus on one objective of the technology--to reduce air time--and few efforts have been made to measure the impact of other MDC objectives. The authors know of no formal evaluations of management information systems and of only one of the regional communications system concept. We have listed several evaluations now underway. NILECJ's grant to Public Systems Evaluation, Inc., to conduct a "National Assessment of Police Command and Control Systems," has helped to facilitate this paper, since both the authors are working on the larger study. More important, the larger study should be able

to begin collecting a common data base for further comparisons of PCCC application.

Despite the fact the number and quality of PCCC evaluations have been disappointing, one can draw some conclusions:

- The use of PCCC technology is still in its infancy. The first commercially sold computer, Univac I, was built only in 1951. Perfection should not be expected instantly in an area so young and rapidly changing. Still, a certain mystique about computers has contributed to oversell. This paper has shown that the reality of the state of the art is often far less than the general impression.
- To the extent that evaluation has been conducted, it has shown that actual performance has fallen below initial expectations. For example, the response-time benefits of AVM and the reduced air time under MDC were not achieved. This does not mean that AVM or MDC should be discarded, but that the initial objectives must be modified to emphasize objectives of greater costeffectiveness. This requires meaningful evaluations.
- A common study framework, such as that suggested by this paper, is necessary. When one police department compares its PCCC evaluation with that of another department, they should be talking the same language even if they are seeking different goals.

Finally, it should be clear that the implementation of new technology, such as in the PCCC area, involves more than technical and quantitative expertise. Too often it is assumed that the diffusion of technological innovation can be initiated from above; it is believed that the mere existence of the technology will prove its worth. Failure to recognize many of the complexities and motivations surrounding the implementation of technology and the interaction between the technology and the nature of police work can prove disastrous. In implementing and evaluating PCCC applications in the future, it should be clear that technological changes often result in behavioral and power shifts. Evaluation studies must consider these shifts by extending beyond measures of input and process and beginning to examine outcome and systemic impacts.

NOTES

1 Kansas City Police Department, Response Time Analysis--Volume I: Methodology (Washington: U.S. Government Printing Office, 1978). The paragraph (p.4) from which this definition is quoted proposes a more sophisticated definition, but this paper uses "response time" in the ordinary sense.

2 Marc W. Tobias, Police Communications (Springfield, Ill.: Thomas Publishers, 1974).

3 President's Commission on Law Enforcement and Administration of Justice, The Challenge of Crime in a Free Society (Washington: U.S. Government Printing Office, 1967).

4 Ibid., p. 252.

5 Kent W. Colton, ed., Police Computer Technology (Lexington, Mass.: Lexington Books, 1978).

6 O. E. Dial et al., "Municipal Information Systems: The State of the Art in 1970," (Long Island, N.Y.: Long Island University Public Administration Center, 1970).

7 Kenneth L. Kraemer, "How City Departments Use Computers and Communications Equipment," Urban Data Service Report, (Washington: International City Management Association), November, 1975; Kent W. Colton, "Uses of Computers by Police: Patterns of Success and Failure," Urban Data Service Report April, 1974; and Kent W. Colton, "Computers and the Police: Police Departments and the New Information Technology," Urban Data Service Report, November, 1974.

8 President's Commission on Law Enforcement and Administration of Justice, Task Force Report: Science and Technology (Washington: Institute for Defense Analysis, 1967), pp. 21-28.

9 The bases for this section are studies by James M. Tien and Vincent F. O'Donnell, A Guide to Single Project Evaluation Design (Cambridge, Mass.: Public Systems Evaluation, Inc., 1978); and James M. Tien and Richard C. Larson, "Police Service Aides: Paraprofessionals for Police," Journal of Criminal Justice 6:117-131 (1978).

10 Roger W. Reinke, "911 Grows Up: Four Years Old and Acting It," The Police Chief (November, 1971).

11 Roger W. Reinke, "A Universal Police Telephone Number," The Police Chief (February, 1968).

12 Kansas City Police Department, Response Time Analysis, (Kansas City, Mo.: Board of Police Commissioners, 1977).

13 James M. Tien, James W. Simon, and Richard C. Larson, An Alternative Approach to Police Patrol: The Wilmington Split-Force Experiment (Washington: National Institute of Law Enforcement and Criminal Justice, April, 1978).

14 Law Enforcement Assistance Administration, "Request for Proposal: Dial 911 Systems Assessment and Transfer," (Washington: January, 1978).

15 Scott W. Hovey, Jr., Study for Alameda County 911 (Washington: Law Enforcement Assistance Administration, October, 1974).

- 16 William L. Miller, "Chicago's 911 Emergency Service: A Computer-Based System for Fast Response," Law and Order (February, 1978), 45.
- 17 Marc Kalech, "911 Shakeup," New York Post (December 13, 1978), p.1.
- 18 Colton, "Uses of Computers by Police," cited in note 7 above.
- 19 Colton, "Computers and the Police," cited in note 7 above.
- 20 R. L. Sohn et al., Application of Computer-Aided Dispatch in Law Enforcement, An Introductory Planning Guide (Pasadena, Calif.: Jet Propulsion Laboratory, December, 1975).
- 21 Kent W. Colton, Margaret Brandeau, and James M. Tien, National Assessment of Police Command and Control System: A Preliminary Report (Cambridge, Mass.: Public Systems Evaluation, Inc., January, 1979); and Jet Propulsion Laboratory, Technical Assistance Project for Law Enforcement Communications (LEAA Grant 7755-99-6013, Ongoing Project).
- 22 Colton, ed., cited in note 5 above.
- 23 Ibid.
- 24 Frank D. Campbell, "Indianapolis Activates New \$5 Million System," AFCO Bulletin (June, 1977), 20-24.
- 25 Joseph Farmer, "Phoenix Police Department's Communication Links: Technology/ Equipment/Systems/Personnel," Law and Order (February, 1977), 16-25.
- 26 W. R. Partridge, Evaluation of San Francisco Police Computer Assisted Dispatch Project (San Francisco, Calif.: San Francisco Police Department, September, 1975).
- 27 Earl W. Robitaille, "A Command and Control System," Law and Order (May, 1974), 22-30.
- 28 J. J. Waterfield, "Computer-Assisted Dispatch," Law and Order (February, 1977), 32-36; and William W. Davis et al., "The Virginia Beach Police Information Systems," The Police Chief (October, 1977), 285-290.
- 29 W. J. Wiseman, Las Vegas Metropolitan Police Department, Communication Automatic Command and Control System (Las Vegas, Nev.: Digital Systems, December, 1975).
- 30 R. L. Sohn, E. A. Garcia, and R. D. Kennedy, Multi-Community Command and Control Systems in Law Enforcement (Pasadena, Calif.: Jet Propulsion Laboratory, March, 1976).
- 31 Colton, ed., cited in note 5 above; and Raymond L. Hoobler and Ken Fortier, "For the San Diego Police Department, A Computer-Aided Dispatch System," The Police Chief (October, 1975), 22-24.
- 32 Colton, ed., cited in note 5 above.
- 33 Scott Hebert, "Command and Control in the Boston Police Department: A Technological Approach to Reform," in Kent W. Colton, ed., Police Computer Technology (Lexington, Mass.: Lexington Books, 1978), p. 169-196.
- 34 Frank Moya, "Police Computer: \$2 Million Headache," The Denver Post (October 22, 1978).
- 35 Sohn et al., cited in note 20.
- 36 Colton, ed., cited in note 5.
- 37 Moya, cited in note 34.
- 38 Hebert, cited in note 33.
- 39 Hoobler and Fortier, cited in note 31.

40 Farmer, cited in note 25.

41 Robitaille, cited in note 27.

42 See Kansas City Police Department, cited in note 12; and Tien, Simon, and Larson, cited in note 13; President's Commission on Law Enforcement and the Administration of Justice, cited in note 3; and Sara Edmundson, Technology Assessment: Project Program on Science, Technology and Society (Ithaca, N.Y.: Cornell University, 1977).

43 Farmer, cited in note 25.

44 For example, Hoobler and Fortier, p. 23, cited in note 31 above, had this to say about the implementation of the San Diego System: "In many systems, much emphasis is given to equipment and, in some instances, not enough importance is placed upon people within the user group ultimately required to operate the system....An important point considered in the implementation of the San Diego CAD System was that ultimately it was people, not machines, who would make the system successful."

45 James W. Simon and Richard C. Larson, An Evaluation Plan for the Philadelphia Computer-Assisted Dispatching (Cambridge, Mass.: Public Systems Evaluation, Inc., June, 1978).

46 Sohn et al., cited in note 20.

47 Robitaille, cited in note 27.

48 Gilbert C. Larson, Alternative Automated Vehicle Monitoring Technologies (Cambridge, Mass.: Public Systems Evaluation, Inc., May, 1976); Gilbert Larson and James W. Simon, Evaluation of a Police AVM System: A Phase II City-Wide Implementation (Cambridge, Mass.: Public Systems Evaluation, Inc., June, 1978); G. R. Hansen and W. G. Le Flang, Application of Automated Vehicle Location in Law Enforcement--An Introductory Planning Guide (Pasadena, Calif.: Jet Propulsion Laboratory, January, 1976); R. D. Doering, AVL System: Orlando Police Department (Orlando, Fla., September, 1974); Mitre Corporation, Urban Field Tests of Four Vehicle Location Technologies (McLean, Va.: Mitre Corporation, April, 1973); and Aerospace Corporation, Systems Design (Vol. II), in Equipment Systems Improvement Program Survey and Technical Automated Vehicle Location Systems for Law Enforcement Applications (Washington: Law Enforcement Assistance Administration, July, 1976).

49 Montclair Police Department, Vehicle Location and Status Reporting System (LOCATES), Final Report Phase II, Project Number 182, (Montclair, Calif.: Montclair Police Department, March, 1974).

50 Richard C. Larson, Kent W. Colton, and Gilbert C. Larson, Evaluation of a Police Implemented AVM System: Phase I with Recommendations for Other Cities (Washington: Law Enforcement Assistance Administration, June, 1976); and G. C. Larson and J. W. Simon, cited in note 48.

51 Seymour M. Weinstein, "Automated Vehicle Monitoring in a Command and Control Environment," Electronics and Aerospace Systems Conference, New York, September, 1978.

52 Doering, cited in note 48 above.

53 See Hansen and Le Flang and Aerospace Corporation, both cited in note 48 above.

54 Stephen W. Riter, Summary of a Panel Discussion of Operational Aspects of AVM, Texas A & M, AVM Conference, November, 1972.

55 R. C. Larson, K. W. Colton, and G. C. Larson, cited in note 50; and Raymond L. McLean, "Letter to Mark McKnew of Public Systems Evaluation, Inc.," Montclair, Calif.: Montclair Police Department, October, 1974 (unpublished).

56 See Hansen and Le Flang, cited in note 48, and R. C. Larson, K. W. Colton, and G. C. Larson, cited in note 50 above.

57 Accuracy considerations are also discussed in Doering, see note 48, and in William E. Farragher, Equipment Systems Improvement Program, A Study of Integrated Location Systems (Washington: Law Enforcement Assistance Administration, September, 1976).

58 Aerospace Corporation, cited in note 48; Riter, cited in note 54; and R. C. Larson, K. W. Colton and G. C. Larson, cited in note 50.

59 Doering, cited in note 48.

60 Riter, cited in note 54.

61 R. C. Larson, K. W. Colton, and G. C. Larson, cited in note 50.

62 The Montclair system is reviewed in Montclair Police Department, cited in note 49, and in McLean, cited in note 55 above; St. Louis is discussed in the Phase I report by R. C. Larson, K. W. Colton, and G. C. Larson, cited in note 50, and the Phase II report by G. Larson and Simon, cited in note 48.

63 Otto A. Reichard, "The Effects of Automatic Vehicle Location on Police Patrol Efficiency," (thesis, Malibu, Calif.: Pepperdine University, 1977).

64 See McLean, cited in note 55.

65 Telcom, Inc., Milwaukee Police Teleprinter Study (McLean, Va.: Telcom, Inc., June, 1969); Urban Sciences, Inc., A Study of Digital Communications Equipment for Law Enforcement Use (Wellesley, Mass.: Urban Sciences, Inc., October, 1973); and R. L. Sohn et al., Application of Mobile Digital Communications in Law Enforcement (Pasadena, Calif.: Jet Propulsion Laboratory, May, 1975).

66 Reported in Oakland Police Department, Oakland Police Department's Digital Communications System (June, 1975); Minneapolis Police Department--An Evaluation of Two-Way Communication (Minneapolis, Minn., 1975); and Hennepin County Sheriff's Department, Hennepin County Coordinated Mobile Teleprinter System (Minneapolis, Minn.: Governor's Commission on Crime Prevention and Control, 1975).

67 National Advisory Commission on Criminal Justice Standards and Goals, Police (Washington: Law Enforcement Assistance Administration, January, 1973), p. 566.

68 Sohn et al., p. 4, cited in note 65 above.

69 G. C. Larson and Simon, cited in note 48.

70 Hal Wand, "Computer Controlled Two-Way Radio Network," Law and Order (April, 1973), 68-74.

71 Oakland Police Department, cited in note 66.

72 Ibid.

73 Glenn A. Marin, "Police Communications in Transition," The Police Chief (July, 1971), 61-64.

74 Oakland Police Department, cited in note 66.

75 William W. Davis et al., "The Virginia Beach Police Information Systems," The Police Chief (October, 1977), 285-290.

76 Two different reports are referenced: R. L. Sohn and R. D. Kennedy, Patrol Force Allocation for Law Enforcement (Pasadena, Calif.: Jet Propulsion Laboratory, February, 1976); and Jan M. Chaiken, Implementation of Emergency Service Deployment Models in Oper (Santa Monica, Calif.: The Rand Corporation, P-5870, May, 1977).

77 Sohn and Kennedy, previous note.

78 Colton, ed., cited in note 5 above, pp. 116-121.

79 See Richard C. Larson, Urban Police Patrol Analysis (Cambridge, Mass.: The MIT Press, 1972) for the PCAM design and Richard C. Larson, Hypercube Queuing Model: User's Manual (New York: The New York City Rand Institute, R-1786, September, 1975).

80 Tien, Simon, and R. C. Larson, cited in note 13 above.

81 Chaiken, cited in note 76.

82 Nelson B. Heller and William W. Stenzel, "Design of Police Work Schedules," Urban Analysis, 2:21-49 (1974).

83 John J. McDonnell, Central Police Department Dispatch: Exemplary Project (Washington: National Institute of Law Enforcement and Criminal Justice, 1974).

84 Wand, cited in note 70 above.

85 Timothy C. Coogan, "BAPERN Revolution in Law Enforcement Communication," The MPA Sentinel, Boston, Mass., (Summer, 1978), 22-27.

86 T. I. Dayharsh, G. E. Barker, and S. C. Ivy, Joint City-County Coordinated Emergency Services Communication System Implementation Study: Analysis of Alternative Systems (Menlo Park, Calif.: Stanford Research Institute, October, 1974).

87 Stanley M. Bolas, Lester W. Grawunder, and Edward V. Regan, "Eris County Benefits from Centralized Police Services," Law and Order (February, 1974), 53-55.

88 Hovey, cited in note 15 above.

89 Sohn, Garcia and Kennedy, pp. 14-5, cited in note 30.

90 McDonnell, cited in note 83.

91 Several police agencies including Los Angeles, as reported by Daniel R. Sullivan, "Los Angeles Police Department Emergency Command Control Communication System" Project Search: Second International Symposium on Criminal Justice Information and Statistics System (San Francisco, Calif., April, 1974), pp. 457-462; in Phoenix as reported by Farmer, cited in note 25 above; and in San Francisco, as reported by Louis H. Feder, "Cable Car City's Cable System," Project Search, supra, pp. 463-469.

92 Sullivan, previous note.

93 William G. Klein, "The Automated System for the Very Small Department," Law and Order (February, 1978) 22-23.

94 See Joseph H. Johns, "Intelligent Computer-Aided Dispatching for Urban Police Patrol Units," (thesis, Cambridge, Mass.: Massachusetts Institute of Technology, September, 1975); and Anthony Vastola, "Police Communication's Systems: A Need for Alternatives," Law and Order (October, 1977), 33-69.

95 Johns, p. 115, previous note.

96 Tien, Simon, and R. C. Larson, cited in note 13 above.

OTHER REFERENCES

Aerospace Corporation, Equipment Options and Cost in 911 Emergency Phone Systems (Washington: Law Enforcement Assistance Administration, July, 1974).

Judson Breslin and C. Bradley Tashenberg, Distributed Processing Systems: End of the Main Frame Era (New York: AMACOM, 1978).

Daniel R. Koshey, "Mobile Digital Terminals: Police Communications of the Future," Law and Order (February, 1978), 26-38.

George W. Kellam, "Mobile Digital Communication Systems," Law and Order (April, 1974), 52-54.

Richard C. Larson, "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment," Journal of Criminal Justice, 3:267-297 (1975).

Newark Police Department, "Newark Computerized Communications Command and Control Systems," (Newark, N.J.: Request for Proposal, 1973).

Public Systems, Inc., Feasibility of a Coordinated Records and Communications System for County of San Diego, 8 June, 1971.

Stephen W. Riter, W. B. Jones, and R. W. Elliott, An Evaluation of the Benefits of Automated Command and Control Systems, 1974.

R. L. Sohn, Voice Data Entry System Demonstration (Pasadena, Calif.: Jet Propulsion Laboratory, April, 1977).

Delinquency Prevention and Control Programs

The Need for a Conceptual Framework and Evaluation Strategies

Albert P. Cardarelli *is assistant professor of sociology at Boston University and former codirector of its Community Sociology Program. Dr. Cardarelli specializes in studies of social and political violence and juvenile justice.*

Charles P. Smith *is director of the National Juvenile Justice System Assessment Center. Dr. Smith is former director of management services for the State of California and a former California Youth Authority parole agent.*

The Need for Evaluation

The need for evaluation of juvenile delinquency prevention and control programs cannot be overstated. An examination of the juvenile justice system indicates a decided lack of comprehensive information on the effectiveness of such programs. Existing evaluation materials are generally of poor quality and provide little information on the effectiveness of intervention strategies. Project directors in many cases are unaware of how to design evaluation measures and generally resort to descriptive accounts of program strategies based largely on "intuitive" measures.

The lack of evaluation materials on delinquency prevention was confirmed by William Wright and Michael Dixon after a comprehensive review of delinquency prevention efforts published in the literature over 1964 to 1974. After reviewing some 6,600 abstracts of programs, the authors were able to find only 96 reports which contained some form of empirical data on project effects. While these projects included both prevention and treatment strategies, the authors conclude:

...that the evaluation literature is low in both scientific validity and policy utility, and that no delinquency prevention strategies can be definitely recommended.¹

The findings of Dixon and Wright are further supported by those of Lundman, McFarlane, and Scarpitti.² After examining some 1,000 citations of delinquency studies published professionally, the authors were able to find only 25 that contained information on the nature and results of the prevention venture. Of these, most could be described as "corrective,"³ and that "in no specific prevention program examined was attention paid to the establishment of punitive or mechanical prevention policies."⁴ In their analysis of these projects, the authors' findings are similar to those of Dixon and Wright, and in their conclusions they argue:

Copyright (c) 1979 Albert P. Cardarelli and Charles P. Smith

Delinquency prevention programs have been largely unsuccessful because of inadequate data, inaccurate or incomplete theories, and compromised intervention strategies. We believe, therefore, that the solutions include direct field observation of delinquents, construction of integrated theories which reflect these field data, and assessment of the constraints which currently compromise prevention efforts.⁵

In the following pages, we will discuss some of the issues associated with delinquency prevention and control and suggest a conceptual framework for evaluating such activities.

The Growing Role of Government

Regardless of one's faith in, or skepticism about, crime statistics, there is near universal agreement that the volume of crime increased during the last 2 decades, and that much of this increase resulted from crimes originated by juveniles and young adults. While this increase has to be balanced against the fact that many juveniles "phase out of" delinquent activities when they enter their late teen years, the need to prevent the onset and persistence of delinquency is obvious.⁶ The more successful society is in reducing the number of juveniles entering the juvenile justice system, the greater the likelihood of reducing the number of juveniles who become persistent or chronic offenders.⁷ The importance of prevention in the reduction of the number of such offenders is reflected in the growing concern of the Federal Government in juvenile delinquency and youth crime during the 1960's.

In 1961, the President's Committee on Juvenile Delinquency and Youth Crime was established. This committee soon recommended the enactment of the Juvenile Delinquency and Youth Offenses Control Act of 1961. Originally authorized for 3 years, the act was later extended through fiscal year 1967. In 1967, the importance of prevention in dealing with crime and delinquency was highlighted by the President's Commission on Law Enforcement and Administration of Justice, which stated:

In the last analysis, the most promising and so the most important method of dealing with crime is by preventing it-- by ameliorating the conditions of life that drive people to commit crimes and that undermine the restraining rules and institutions erected by society against antisocial conduct.⁸

In the following year (1968), Congress enacted both the Juvenile Delinquency Prevention and Control Act and the Omnibus Crime Control and Safe Streets Act. These were followed by the Juvenile Delinquency Prevention Act of 1972 where, in a similar fashion to the earlier Prevention and Control Act of 1968, funding remained under the auspices of the U.S. Department of Health, Education, and Welfare (HEW). Two years later, Congress enacted the Juvenile Justice and Delinquency Prevention Act of 1974, which established the Office of Juvenile Justice and Delinquency Prevention within the Law Enforcement Assistance Administration (LEAA) of the Department of Justice. The concern of the Federal Government regarding delinquency and the importance directed toward "prevention" is evidenced not only in the title of the act of 1974, but in the attention given to prevention as an important strategy for forestalling antisocial behavior among adolescents and young adults.

In pointing to the major objectives of the Act, Congress noted that its declared policy was:

...to provide the necessary resources, leadership, and coordination (1) to develop and implement effective methods of preventing and reducing juvenile delinquency; (2) to develop and conduct effective programs to prevent delinquency, to divert juveniles from the traditional juvenile justice system and to provide critically needed alternatives to institutionalization; (3) to improve the quality of juvenile justice in the United States; and (4) to increase the capacity of State and local governments and public and private agencies to conduct effective juvenile justice and delinquency prevention and rehabilitation programs and to provide research, evaluation, and training services in the field of juvenile delinquency prevention.⁹

Defining Prevention—Some Conceptual Issues

The purpose of the following discussion is to review some of the varied meanings associated with the concept of "prevention." Before doing so, however, it is important to realize that the confusion and ambiguity associated with prevention is further exacerbated by joining it with the term "delinquency." Both concepts, because of their inclusiveness, have incited a good deal of controversy and debate at the policy or programmatic level.

In an early review of delinquency prevention programs, Witmer and Tufts point to three major conceptions of prevention that dominated the field of delinquency up to the 1950's.¹⁰ The first cat-

egory includes efforts aimed at promoting the "healthy personality development" of all children. Within this broad conceptual level, the prevention (or control) of delinquency was directed toward improving those aspects of society that affect the personality development of children. Obviously, the range of such activities is extremely broad and encompassing, and while commendable, includes a greater array of behavior than delinquency. A second category envisions delinquency prevention as efforts directed primarily toward potential delinquents before they become involved in delinquent behavior. Proponents of this viewpoint not only believe that community resources can be more effectively utilized with predelinquents, but argue that such individuals can be identified through predictive measures. The third category includes programs which stress the reduction of recidivism by lessening the possibility of serious offenses. Prevention efforts under this orientation are directed toward preventing the continuance of delinquency rather than its onset.

After reviewing the efforts based on these categories, the authors argue for a definition of prevention as referring "both to the forestalling of delinquent behavior and also to the reduction in its frequency and seriousness."¹¹ In this way, they have included the essential elements of all three definitions above as opposed to arriving at a more precise definition. Rather than discuss the merits of these approaches, several additional interpretations of prevention will be examined.

In a further critique of delinquency prevention programs prior to the 1960's, John Martin found that delinquency prevention programs correspond to one of the following definitions:

- Delinquency prevention is the sum total of all activities that contribute to the adjustment of children and to healthy personalities in children.
- Delinquency prevention is the attempt to deal with particular environmental conditions that are believed to contribute to delinquency.
- Delinquency prevention consists of specific preventive services provided to individual children or groups of children.¹²

Although the first category is a restatement of Witmer and Tufts' classification, the second definition emphasizing "environmental conditions" reflects the increasing attention paid to the importance of the social system as a causal factor in increasing delinquency. The last definition indicates a growing recognition of the varied types of behavior classified as "delinquency" and

the search for differential treatment strategies. As with Witmer and Tufts, each of these program orientations can be traced to the varied theoretical perspectives oriented toward the etiology of delinquency.

The confusion in delinquency prevention that prevailed in the 1950's and early 1960's led Lejins to write in 1967:

...the field of prevention is by far the least developed area of criminology. Current popular views are naive, vague, mostly erroneous, and for the most part devoid of any awareness of research findings; there is a demand for action on the basis of general moralistic beliefs, discarded criminological theories of bygone days, and other equally invalid opinions and reasons. In scientific and professional circles the subject of prevention has received remarkably little attention. Even the basic concepts in the field of prevention lack precision. There has been very little theory-building, and attempted research under such circumstances has failed to produce any significant results.¹³

In his review of the field of prevention, Lejins argues for the need to distinguish between "prevention" and "control." For Lejins:

Prevention is a measure taken before a criminal or delinquent act has actually occurred for the purpose of forestalling such an act; control is a measure taken after a criminal or delinquent act has been committed.

Since "control measures" may also help to forestall further criminal offenses, Lejins argues that it is a difficult and confusing task to differentiate between control and prevention, unless the concept of control is restricted to:

...any action concerning an offender taken as a result of his having committed an offense...even if it interrupts the continuation of criminal behavior and thereby forestalls future criminal acts.¹⁵

Using this distinction as a base, Lejins describes three types of prevention: punitive, corrective, and mechanical.¹⁶ Punitive prevention, he notes, relies on the threat of punishment to forestall the criminal law, and is based on the premise that a potential offender's awareness of the prospective punishment for an offense will deter him from that behavior. Corrective prevention,

on the other hand, is based on the premise that certain conditions "lead to" or "cause" criminal behavior and it is these conditions which must be eliminated if delinquency is to be prevented. The last category, that of Mechanical prevention, is directed toward making it difficult or impossible for an individual successfully to commit a limited range of offenses. Under this orientation, the primary goal is to "harden the target" to make it inaccessible to the offender. The recognition that prevention includes a wide range of activities is also noted by Eleanor Harlow, who in 1969, distinguished three major prevention strategies:¹⁷

- Primary Prevention is directed toward the criminogenic environment without distinguishing between those persons who have responded criminally and those who have not.
- Secondary Prevention includes programs concerned with delinquency-prone individuals and emphasizing early identification and treatment of predelinquents.
- Tertiary Prevention is corrective in that it is concerned with preventory recidivism.

An examination of Harlow's categories indicates little difference between the early classification of Witmer and Tufts; each interprets prevention as being directed at three types of youth: (a) general population, (b) predelinquent, and (c) delinquent. Obviously, the use of the term "prevention" to include activities associated with all three categories of youth can only add to the confusion associated with the concept.

In a more recent analysis of prevention, Kenneth Polk and Solomon Kobrin argue that the tendency in the past has been to search for the "causes" of crime and to develop prevention programs addressed to these causes.¹⁸ In their analysis the authors argue for an approach that specifies how legitimate rather than illegitimate pursuits are pursued. Delinquency prevention, they argue, should emphasize institutional reform rather than individual change. For them, both from a practical and strategic matter:

The approach to the problem of adolescent deviance, and to delinquency prevention and control, must focus on institutional malfunction.¹⁹

Based on this approach, efforts would be directed toward restructuring existing institutions, while discarding those features that tend to foster delinquent behavior and identities.

Polk and Kobrin go on to argue that since prevention consists of activities developed to reduce the incidence of those be-

haviors that lead to the imposition of the label of delinquency, the most appropriate manner in which this can be accomplished is through the restructuring of the present institutions or the creation of new ones.²⁰

The growing attention to institutional change is also noted by LaMar T. Empey who argues that "...any serious effort at crime prevention would have to consider ways by which socialization per se might be made more effective."²¹ Given this direction, Empey argues that if these institutions are to be more effective, prevention programs should consider the following assumptions as crucial to prevention:

1. the primary focus of prevention efforts should be upon the establishment among young people of a legitimate identity;
2. a legitimate identity among young people is most likely to occur if they have a stake in conformity;
3. the cultivation in young people of a legitimate identity and a stake in conformity requires that they be provided with socially acceptable, responsible, and personally gratifying roles; and
4. a rational strategy of delinquency reduction and control must address the task of institutional change.²²

In an attempt to deal with existing confusion regarding the definition of "prevention," and the most appropriate strategies to accomplish the prevention of delinquency, the National Advisory Committee on Criminal Justice Standards and Goals recommended the following definition in 1976:

Delinquency prevention is a process of problem identification, resource analysis and strategy building aimed at lowering rates of delinquency through the provision of services to persons or groups with specific and demonstrated needs.²³

While the above definition indicates the importance of "process" in the prevention of delinquency, it also emphasizes the "provision of services" as the major strategy for accomplishing prevention.

These conclusions clearly indicate the need to differentiate among varied patterns of delinquency in both criminological research and delinquency prevention programs.²⁴ In addition to these

definitional issues, it is also important to recognize that a majority of juveniles engage in delinquent acts sometime during their adolescence, most of which are never officially observed or recorded.²⁵ Further, while most adolescents with official arrest records are from lower socioeconomic backgrounds, the evidence that delinquency exists among all levels of the economic system is quite substantial.²⁶ The implications of these findings for delinquency prevention programs need to be given serious consideration.

In addition to the confusion generated by the meaning of "delinquency" in research studies, the problem of defining delinquency may also affect the way in which delinquency prevention programs are developed. Spergel, for example, argues that the term delinquency:

...is just as variable and complex as community. In fact, it may be even more variable within a community than across the juvenile-justice system. There appears to be no across the board operational definition of delinquents or predelinquents in community based programs. It is not at all clear what a community program to prevent or to treat delinquents really is or should be.²⁷

Studies, evaluations, and descriptions of prevention and treatment programs nationwide rarely define the range of behavior included under the definition of delinquency nor, except in a few cases, are there well-developed strategies to deal with the differential character of delinquency.²⁸ Further, while many administrators and program staff are able to distinguish between prevention and control in theory, the distinctions are not always explicit in practice. Some agencies, claiming to be "preventive" organizations, are actually involved in the treatment of adjudicated delinquents, while some "treatment" agencies are involved in the development of preventive activities for children considered to be either "predelinquent" or as having high probabilities of being defined as delinquent at some future time. Further, since the term "prevention" is not always precisely defined by practitioners, it is difficult to establish mutually exclusive categories of preventive and treatment activities, thus adding to the confusion that has dominated the field of delinquency prevention for the last 40 years.²⁹

Despite recent efforts to provide overall profiles of the types of projects aimed at preventing juvenile delinquency, little is known about the effectiveness of the wide range of interventions to prevent delinquent behaviors.³⁰ Similarly, there have been virtually no efforts to describe both the actual interventions and the chains of theoretical reasons and assumptions linking these

interventions to underlying principles of causality and social change.³¹ In short, little is known about either what actually takes place nationally under the rubric of juvenile delinquency prevention or how those practices and projects are justified by sound theoretical and operational principles.³²

A CONCEPTUAL FRAMEWORK FOR EVALUATING DELINQUENCY PREVENTION AND CONTROL PROGRAMS

A Synthesis of Delinquency Prevention Projects

Although findings from each of the above mentioned studies are based on analyses of reports contained within the professional literature, they are consistent with those obtained in an assessment of delinquency prevention projects conducted by the Center of Vocational Education at Ohio State during 1975, under a grant provided by the National Institute of Law Enforcement and Criminal Justice.³³ The purpose of this assessment was to provide an information base for policymakers by examining and assessing current strategies of delinquency prevention nationwide. The following pages provide a synthesis and assessment of the major findings of this effort.

In order to arrive at a representative sampling of delinquency prevention programs in operation, an extensive search of existing data bases was undertaken. This data search yielded several reference sources to project specific information ranging from approximately 200 to over 5,000 citations. The Law Enforcement Assistance Administration's Grant Management Information Service (GMIS), by far the largest printout, included 2,100 pages of project summaries, with an average of 2.5 summaries per page.³⁴ Unfortunately, the GMIS did not distinguish between primary and secondary prevention programs and consequently further manual screening was necessary.³⁵ When possible, staff keyed decisions to catchphrases such as "predelinquent" and "primary prevention." Eventually, 1,486 programs were identified as prevention programs; of this figure only 20 percent were currently active. After an extensive phone survey, 120 projects were selected as possible candidates for site visits.

Efforts were also directed toward establishing contacts with information sources other than LEAA. Specifically, the staff searched for Federal, State, and local agencies of government involved in delinquency prevention. To add to these difficulties, no centralized source of information exists for cataloging privately funded delinquency prevention projects or programs. Therefore, it is important to note that the analysis is skewed toward federally supported efforts.

Actual site selection procedures for assessing delinquency prevention programs were initiated by an extensive telephone interview survey which focused on individual projects. The purpose of these interviews was to confirm the type of program and to gather program information prior to the final site selection. Many of the projects contacted in this manner nominated additional primary prevention programs with which they were familiar. All projects reviewed by telephone were classified according to the following criteria:

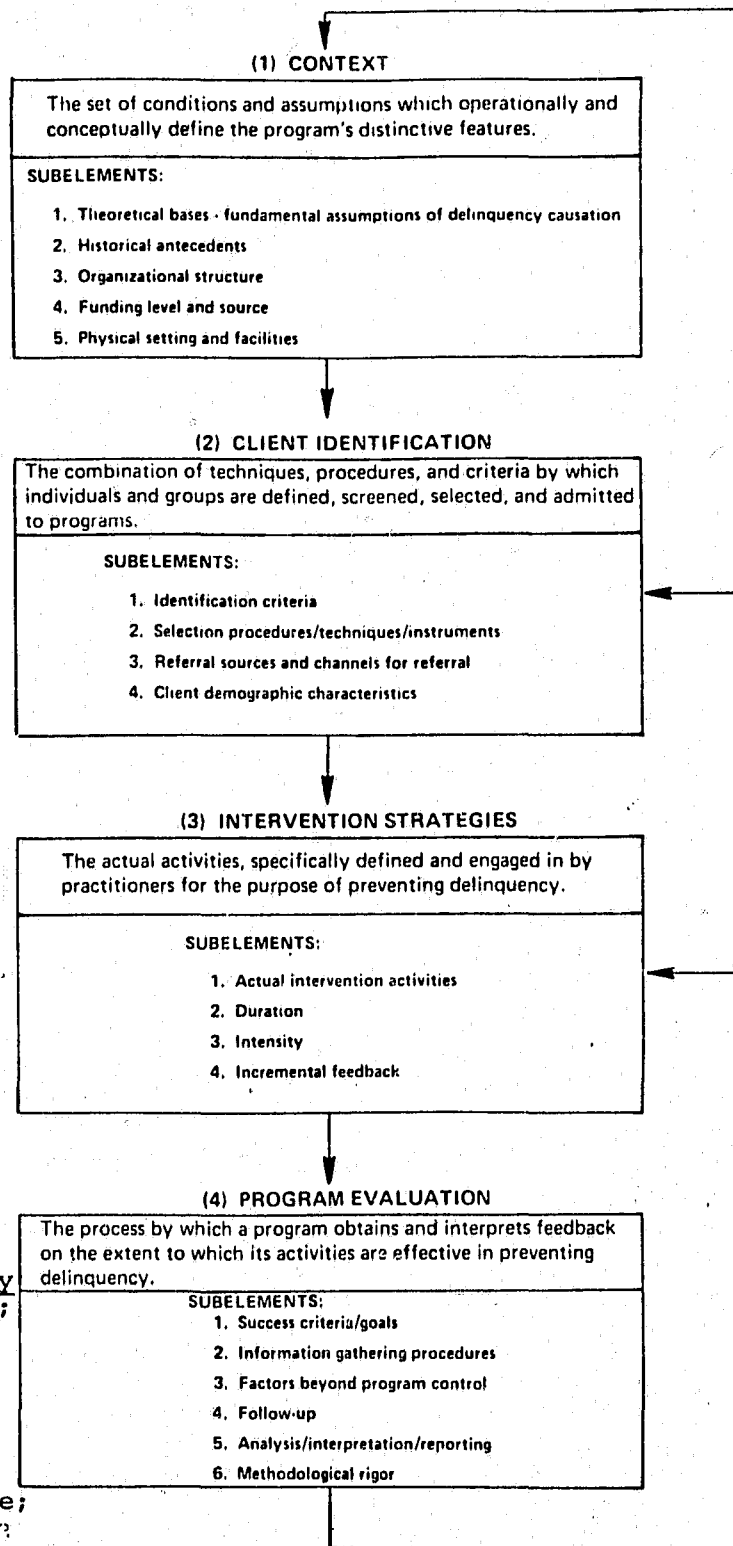
- the type intervention by cluster type (see descriptions below)
- geographic location
- locus of intervention (e.g., family, schools)
- predominate mode of intervention (e.g., case work, counseling)
- target group characteristics (e.g., age, sex, race, and approximate percentage of clients having police or juvenile court records)
- funding source
- maturity (i.e., years in existence)
- staff size
- evaluation.

It is important to emphasize that one crucial criterion for site selection was evaluation. Extreme care was given to visiting sites claiming formal, external evaluations, which offer the most potential for gathering systematic information. From the list of 120 projects, 35 were selected for site visits.

In addition to the importance of evaluation, efforts were made to examine a wide range of projects representative of the dominant strategies being utilized by program staff. Based on an extensive review of literature, research, and evaluation reports of delinquency control and prevention, a conceptual framework for classifying strategies and techniques of juvenile delinquency prevention was developed.³⁶ This framework classifies specific program techniques under one of six major intervention strategy clusters described below.

Counseling: Defined as that range of interventions from psychiatric analysis and psychotherapy to simple advice giving and "active listening." It is the most prevalent

Figure I
Juvenile Delinquency Prevention
Program Elements ³⁷



Source: Principles and Guidelines for State and Local Administrators of Juvenile Delinquency Prevention Programs; by Dennis L. Billingsley, Albert P. Cardarelli and Jerry P. Walker; National Institute of Law Enforcement and Criminal Justice; January 1976, p. 347

and widely accepted of all intervention strategies, employed by licensed psychologists, social workers, and lay persons alike.

Recreation: Defined as those semiorganized and organized nonwork activities engaged in by youth for pleasure and considered by practitioners as an appropriate youth development activity to prevent destructive behavior.

Police-School-Community Relations: Defined as those programs designed to promote positive attitudes, opinions, and thoughts of individuals in society toward the law in general and law enforcement activities in particular. Most such programs involve law enforcement officials who attempt to minimize the alienation between youth and the law.

Instruction: Defined as those activities that manipulate given resources and skills to bring about an improved grasp and comprehension from diverse sources. Information may be either formal or informal knowledge. It is a strategy that has as its objective an increase in competence for the adolescent, those in traditional positions of instruction (e.g. teachers), or both.

Opportunity Enhancement: Defined as those efforts to generate resources that would provide opportunity to acquire and apply given skills. In practice, opportunity enhancement programs involve job training, vocational counseling, job development and job placement activities to increase the life changes of an individual youth. The key concept of these programs is the provision of opportunity.

Youth Advocacy: Defined as those activities which involve the direct interface with institutions on the behalf of youth. While it may be undertaken for an individual youth, it is usually associated with "class action." Advocacy programs favor changes in organizational structures, procedures, and service patterns as well as changes in the legal institutions. Emphasis is on innovative change and effective delivery of services to clients.

A series of pretests of local prevention programs in Ohio was initiated to determine problem areas in gathering information during site visits. Based on these experiences, a site-review manual was developed which had as its objective the development of a descriptive profile of each project in terms of four broad categor-

ies. These include: (1) the context of the program; (2) the methods of client identification; (3) the intervention strategies used to help prevent delinquency; and (4) the methods of evaluation utilized in the program (see Figure I for a summary of these categories).

Context is defined as the set of conditions and assumptions which operationally and conceptually define the distinctive features of delinquency prevention programs. Included are the physical, financial, historical, organizational, and theoretical characteristics of the program.

Within "context," the matter of fundamental assumptions represents an area requiring special attention and documentation. It was felt that the fundamental assumptions would define the bases upon which target audiences are identified, the intervention strategies which are selected and implemented, and the evaluation logic and procedures which are to be employed.

It should be strongly stressed that it was not expected that any single program would (or should) attempt to articulate and document the fundamental assumptions that account for the full range of all delinquent behavior. Rather, it was anticipated that programs would either "specialize" in mediating particular causative factors within a well-defined and documented range, or would involve staff in prevention practices, exclusive of causative factors, but within an equally well-defined range of activities.

Client Identification is defined as the combination of techniques, procedures, and criteria by which individuals and groups are defined, screened, selected, and admitted to program participation. As previously indicated, the fundamental assumptions of a project indicate the problematic characteristics or causative factors from which the identification criteria and procedures are derived. A review of project summaries prior to site visits indicated that many programs are much less individualized or targeted in identifying characteristics of causation which are to be dealt with by the intervention process. Often termed "nontargeted," these programs are more general in setting criteria than so-called targeted programs. Relative juvenile crime rates, scholastic ability, school dropout rates, crime victim surveys, self-report instruments, socioeconomic status, ethnicity, or area of residence are all criteria used for nontargeted group selection.

Intervention Strategies include the full range of actual activities engaged in by practitioners for the purpose of preventing delinquency. Included within the program element of intervention are the subelements of duration, intensity, and sensitivity to incremental feedback. While duration and intensity

are self-explanatory terms, it was anticipated that within programs both would vary by client, and that such variance would be determined by individual client characteristics, staff sensitivity to incremental feedback, and the unique needs of both staff and clients. Sensitivity to incremental feedback presupposes the ideal existence of planned and implemented measurement points during the intervention process. Such "midstream" measurement points allow for decisions to be made regarding the success or appropriateness of the intervention activity, changes in intervention methodology (i.e., possibly referral to another program), client recycling, termination from all intervention efforts, or simply changes in duration or intensity.

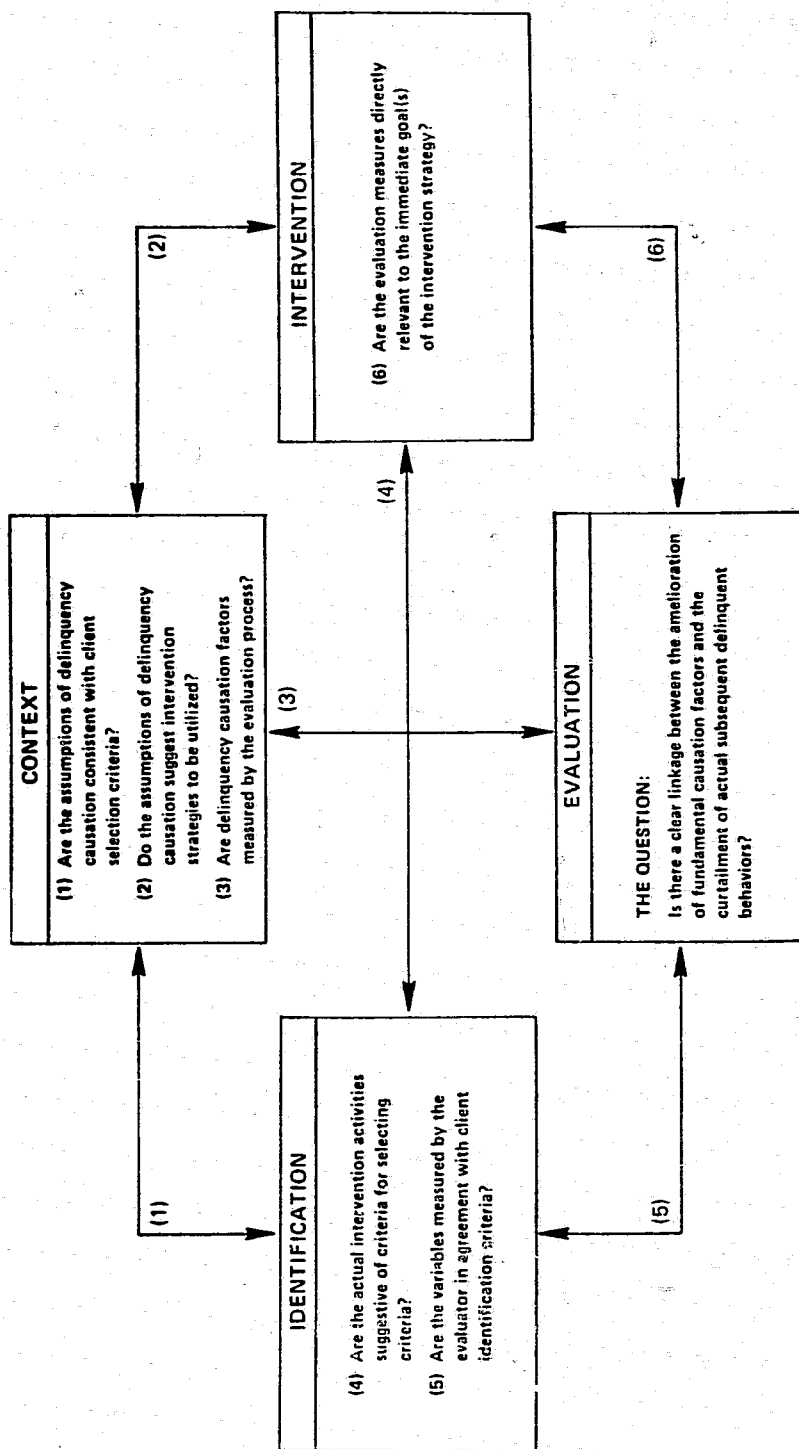
Evaluation is defined as the process by which a program obtains and interprets feedback on the extent to which its interventions have been successful. An ideal program evaluation would attempt to explain both its successes and failures in terms of implications for program improvement. Measures of costeffectiveness or administrative efficiency, although helpful, are not considered to be sufficient evaluation measures. Similarly, monitoring practices, incorporating numbers of clients served by age, sex, ethnicity, education level, and reporting problems are not, in and of themselves, evaluation.

In effect, the evaluation of most programs should be twofold. The evaluator should first focus on the changes and processes that develop throughout the history of the project and whether these changes affect the character and direction of the intervention strategies. Second, the evaluator should determine the impact of the strategies on the extent or character of delinquency.

The difficulties in adhering to these "ideal" evaluation considerations are enormous. When one considers the fact that few programs have managed to conduct impact evaluations which offer convincing evidence that delinquent behaviors have been prevented, it becomes clear that obstacles to a methodologically sound evaluation abound.

The above discussion provides a general description of the four broad elements of juvenile delinquency prevention projects --context (emphasizing fundamental assumptions), client identification, intervention strategies, and evaluation--which were utilized in the site visits to assess delinquency prevention programs. Implied in each of these definitions are several major logical linkages or interrelationships of program elements (see Figure II). Throughout our assessment the undergirding principle was that if these linkages were clear and strong in their logic, the projects would be much more effective and efficient in their activities.

Figure II
Logical Linkages of Juvenile Delinquency
Prevention Program Elements ³⁸



Source:
Principles and Guidelines for State and Local Administrators of Juvenile Delinquency Prevention Programs; by Dennis L. Billingsley, Albert P. Cardarelli and Jerry P. Walker; National Institute of Law Enforcement and Criminal Justice; January 1976, p. 8.

An Assessment of Delinquency Prevention Projects

Some of the major findings associated with the site visits are summarized in terms of the four major categories of context, client identification, intervention, and evaluation.³⁹

Context-Fundamental Assumptions

For almost all site visits it was necessary to infer many of the projects' fundamental assumptions about delinquency causative factors. Members of the prevention programs interviewed either did not explicitly state the basic assumptions or philosophies underlying their programs, or simply did not give serious consideration to the issue. Many programs intervene only with strategies with which they are most familiar and proficient--usually counseling--but do not or cannot directly link what they are doing to any basic assumptions about the causes of delinquency. Thus, while the family is seen by almost all projects as a "causative" factor in delinquency, less than one-half of the projects emphasize family intervention strategies. Further, although economic factors are viewed as important causal elements, they are seldom dealt with directly. For example, many program staff argue that attempts to ameliorate the "causes" of delinquency are not feasible within their programs, and therefore are not productive avenues to pursue. In this respect, it becomes clear why many programs which report the same fundamental assumptions differ substantially in the intervention strategies they utilize.

Overall, the site interviews revealed a pattern of ill-defined or inconsistent fundamental assumptions and intervention strategies. Often, prevention program staff paid lip service to particular types of intervention strategies but did not implement them. Lastly, in several cases it was evident that the fundamental assumptions about delinquency causation often were prepared for purposes of grantsmanship and were really not manifested in any part of the program.

Client Identification

In terms of the weaknesses of the prevention program, the process of identifying potential delinquents is second only to the illogic and inconsistency between context and intervention and is due in part to the many dilemmas and paradoxes faced by project directors. Because of the virtual absence of consistent identification criteria, most delinquency prevention projects not only work with nondelinquents but with those who are arrested or adjudicated as delinquent. Although these programs are able to provide an indication as to the proportion of clients within each category, the intervention strategies within the programs

do not discriminate between the so-called nondelinquent youth and those who have "penetrated the system." In fact, several of the projects made a point of not asking questions related to the clients' previous arrests or apprehensions. Inappropriate or neglected identification procedures and criteria result in the inappropriate selection of clients and account in part for the failure of many delinquency prevention programs to demonstrate their success.

The lack of congruency or logical linkages between the identification of clients and fundamental assumptions or intervention strategies brought to light several other important findings. Primary among these is the topic of client selectivity or what is referred to as "the skimming process." Through a variety of means, many of the programs manage to work with "the best of the bad kids," referring the more belligerent, hostile, and aggressive youths to other agencies, or leaving them to fend for themselves. This is an important finding for social policy. The exclusion of those youths most in need of services may preclude the prevention of more serious forms of delinquency and may be instrumental in the development of criminal careers.

Intervention Strategies

Actual intervention techniques were found to conform to the six cluster types as previously speculated. Programs in the field exhibited intervention activities which may generically be titled counseling, instructional, recreational, youth advocacy, opportunity enhancement, and police-school relations.

The general pattern of linkages between a project's fundamental assumptions and its intervention strategies is one in which practitioners espouse a "social-institutional" or "social-interactive" basis for delinquency, but actually utilize intervention strategies which are individual/psychological in nature. Practitioners in all clusters, with the possible exception of youth advocacy, not only are at a loss in effectively addressing the admittedly complex interactive set of social/individual variables giving rise to delinquency, but believe there is little they can do to directly impact on these variables. These findings have important implications for those who wish to establish social policy based on a theoretical framework that is system-specific in nature.

Evaluation

The national evaluation staff found that, without exception, efforts at evaluation are either lip-service or inadequate in execution. In many cases evaluation is nonexistent, with few proj-

ect directors viewing it as a much-needed process. There are only haphazard designs concerning the intensity of program interventions, with program staff concerned more with the duration of the staff-client relationship rather than its intensity. Many project staff resort to the argument that one cannot adequately measure "subjective behavior or attitudes," whereas others claim that there is not enough expertise at their disposal to execute adequate evaluation.

External Linkages and Program Constraints

In addition to the above elements, programs were also assessed in terms of their "external linkages" with the outside world, as well as the external program constraints.

Site visits indicated that most external linkages are based on some formal or contractual agreement. The most common is one with the juvenile justice system for the purpose of referral or transmission of information concerning delinquent individuals. In the vast majority of cases these formal connections are for establishing some type of intervention strategy. This usually takes place when clients are being identified for acceptance to the program and generally terminates immediately thereafter. In several instances, the juvenile justice system contracts and pays for program services. During the intervention and evaluation stages, seldom is there further communication between the programs and the juvenile justice system.

Other formal linkages exist within the community agency referral system, and include welfare, employment, educational, and law enforcement organizations. Again, these are generally utilized at the identification or entry stage of the program, and are usually formal in that services are contracted and paid for either directly or indirectly.

Overall, the current status of linkages of delinquency prevention programs to external agencies, community resources, and other prevention programs can be best characterized as: (a) substantially lacking in cooperation for referral, feedback, and followup purposes; (b) riddled by mistrust and suspicion; (c) competitive (for both clients and funding); and (d) ill-conceived and haphazardly maintained. Program linkages with the juvenile justice system are typically contractual arrangements, serve to "widen the net" of the juvenile justice system, and serve only as a referral channel since little or no subsequent information flows between the system and the program.

In addition to problems resulting from inadequate program linkages, virtually all delinquency prevention projects visited were faced with a wide variety of explicit constraints. Two of

the most important were funding and public relations. In terms of funding, many program staff indicated that the issues of continuity, timing, and criteria were even more pervasive constraints than the actual level of dollar amounts per grant. The issues of shifting funding resources, massive paper work, contradictory criteria, and annual uncertainty were complicated and bitter topics of discussion, and viewed as having important ramifications affecting all program elements.

A further program constraint expressed by community-based practitioners in particular is that of negative political, police, and community relations. It was not uncommon to hear project directors speak of the need to establish better relations with parents and leaders within the community, or to see projects establish "Boards of Directors" from the area's politically powerful residents in an attempt to establish credibility and muster support from the political structure.

CONCLUSIONS

Political and Economic Constraints

It is important to keep in mind that the successful implementation of any program of delinquency prevention may be affected by structural factors of a national character over which program directors have little control. For example, the continued transition of large cities from manufacturing to service economies, coupled with the outflow of manufacturing and retail jobs to the suburbs, has led to higher rates of urban unemployment that is often reflected in family and personal disorganization. The importance of these wider sociocultural and environmental factors in both the causation and prevention of delinquency should neither be ignored or treated lightly, nor be used as "excuses" to avoid the immediate problems associated with their impact. Program staff not only need to be explicit about their domains of competency, but further must be realistic about the changes that are feasible within program structure.

In addition, there are other political or economic constraints that inhibit both program initiation or evaluation. These include the separation of powers among branches, levels, and units of government; fund availability and budgetary process; the nature of bureaucracy; constituencies and special interests; and inadequate policymaker understanding and involvement.⁴⁰

Theoretical Issues: Changing Individuals vs. Conditions

One of the most obvious dichotomies in the field of delinquency prevention derives from practitioners who view the causes and solutions of delinquency in terms of the individual juvenile, in contrast to those who focus on social conditions or factors to which the individual is exposed. Each of these perspectives has direct consequences for the kinds of prevention activities that are undertaken. Either perspective to the exclusion of the other may, however, result in too narrow a focus as noted by Richard Cloward:

The tendency to define the source of many social problems in essentially individualistic terms rather than to focus as well on the way in which institutional inadequacies generate these problems has led us to become too preoccupied with rehabilitation. We have tended to behave as if problems of delinquency, dependency, illegitimacy, and the like can be solved in essentially rehabilitative terms.⁴¹

Although these comments are directed toward perspectives that are essentially "individually oriented," the criticism applies equally well to those sociological perspectives that either do not deal with individual differentiation or responsibility, or with the legal rules that define certain behavior as "deviant" in the first place.⁴²

The Degree of Community Cohesion

One of the major factors to consider in the development and implementation of any prevention program is the degree of community cohesion that exists. During the last 3 decades most large American cities have undergone important economic and cultural changes that raise serious doubts about the existence of any "uniformity of interests" or "communal purpose." Racial conflicts, neighborhood deterioration, and high rates of mobility are only manifest examples of widespread political and economic conflicts in many large cities. They must be given serious consideration prior to the development of any delinquency prevention program. Given the wide variations of neighborhoods throughout the country it is vital for policymakers to recognize that if delinquency prevention programs are to succeed, they must be integrated into community activities rather than developed in isolation from neighborhood residents or agencies.⁴³

Early Intervention vs. Nonintervention

In the field of delinquency prevention, many argue that it is possible to delineate the path of behavior that an individual will take over a period of time and thereby allow increased attention to be directed to those adolescents who are "potentially delinquents." This rationale for early identification is based on the belief that once "potential delinquents" are identified, more effective strategies can be implemented. The nature of the activities will vary from nonintervention to active intervention and treatment.

Attempts at early identification and prediction of predelinquency have not been without criticism from both an empirical and ethical level of analysis.⁴⁴ Thus, the President's Commission advocated in 1967 that prediction studies of individual cases should be approached with caution and concern for:

[I]nherent in the process of seeking to identify potential delinquents are certain serious risks--most notably that of the self-fulfilling prophecy.⁴⁵

Additional arguments against attempts at predicting delinquency gain their impetus from the growing philosophy of "nonintervention" espoused by many practitioners and described by Edwin Schur as that which:

...implies policies that accommodate society to the widest possible diversity of behaviors and attitudes rather than forcing as many individuals as possible to adjust to supposedly common standards.⁴⁶

The belief that early intervention may not only be a negative activity but an inefficient one is not without empirical evidence. Longitudinal studies of juveniles by Wolfgang et al.⁴⁷ indicate that only a small proportion of adolescents become chronic and persistent offenders. These youth not only are responsible for a large proportion of serious offenses, but are much more likely to be arrested as adults. Most youth who engage in some acts of delinquency do not follow this pattern. Official interventions by the juvenile justice system in contrast to positive youth development strategies prior to the onset of delinquency would therefore appear more costly and less effective.

Prevention as Reinforcement of Positive Activities

During the 1960's, with the publication of Delinquency and Opportunity,⁴⁸ and the direction that the President's Committee

on Juvenile Delinquency pursued, greater emphasis was placed on enhancing the opportunities of large numbers of inner-city youth who were seen as most vulnerable to delinquency. Programs similar to New York City's Mobilization for Youth were established and funded in several large cities, with the prime goal of changing those conditions viewed as most conducive to delinquency.⁴⁹ Supported by the political themes of the "New Frontier" and "The Great Society," government spent millions of dollars to provide large numbers of disadvantaged youth with expanded economic opportunities.

Although the initial hopes of these programs were never fully realized for a number of reasons, increased attention was directed toward the ways institutions of social control deal with youthful misbehavior and delinquency. One result of this process has been an increased demand for viewing prevention as a reinforcement of positive attitudes and the development of positive goals rather than the prevention of a wide range of juvenile behaviors, many of which would not lead to an arrest if committed by an adult.

SUMMARY

Given the political and economic constraints that impact on the development and implementation of delinquency prevention programs, much greater priority needs to be given to the expansion of legitimate opportunities and identities in the primary socializing institutions. Linkages between the school, the world of work, and family need to be strengthened and encouraged. In addition, there is a critical need for increased interagency coordination between the schools and neighborhood agencies if delinquency is to be prevented through the encouragement of prosocial activities.

Further, because of the overwhelming evidence showing that most adolescents "phase out of delinquency," special consideration needs to be given to those adolescents who are most in need of services, if they are to make the transition to adulthood with the least amount of negative costs to themselves and society.

The above assessment indicates a pressing need for increased evaluation of delinquency prevention and control programs in the context of a suitable conceptual framework and adequate implementation strategies. Without this we will be increasingly immersed in programs that do not work--or that do work, but we do not know why they work.

It is only through the combination of such competent program design, implementation, evaluation, and modification that the problems of juvenile crime and delinquency will be properly prevented or controlled.

NOTES

1 For a discussion of delinquency in terms of its pattern and seriousness over time, see Marvin E. Wolfgang, Robert M. Figlio, and Thorsten Sellin, Delinquency in a Birth Cohort (Chicago: University of Chicago Press, 1972); Albert P. Cardarelli, "Socio-Economic Status and Delinquency and Adult Criminality in a Birth Cohort." diss., (University of Pennsylvania, 1973).

2 For an insightful analysis of delinquency as a "temporary" or "episodic" aspect of adolescence, see David Matza, Delinquency and Drifts (New York: John Wiley & Sons, Inc., 1964); Edwin Schur, Radical Non-Intervention: Rethinking the Delinquency Problem (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973).

3 The President's Commission on Law Enforcement and Administration of Justice, The Challenge of Crime in a Free Society (Washington: U.S. Government Printing Office, 1967), p. 58.

4 Juvenile Justice and Delinquency Prevention Act of 1974, Public Law 93-415, 93d Congress, S. 821, September 7, 1974, p. 2.

5 Helen L. Witmer and Edith Tufts, The Effectiveness of Delinquency Prevention Programs, Children's Bureau, U.S. Department of Health, Education, and Welfare, 1954.

6 Ibid., pp. 3-5.

7 John M. Martin, "Three Approaches to Delinquency Prevention: A Critique," in Simon Dinitz and Walter C. Reckless, ed., Critical Issues in the Study of Crime: A Book of Readings (Boston: Little Brown & Co., 1968), pp. 161-164.

8 Peter P. Lejins, "The Field of Prevention," in William E. Amos and Charles F. Wellford, ed., Delinquency Prevention: Theory and Practice (Englewood Cliffs, N.J.: Prentice-Hall, 1967), pp. 1-21.

9 Ibid., p. 2.

10 Ibid., p. 2

11 Ibid., p. 3

12 Eleanor Harlow, "Prevention of Crime and Delinquency: A Review of the Literature," Information Review on Crime and Delinquency, 1, 6:3,4, (Hackensack, N.J.: National Council on Crime and Delinquency, April, 1969).

13 Kenneth Polk and Solomon Kobrin, Delinquency Prevention Through Youth Development, U.S. Department of Health, Education, and Welfare, 1972.

14 Ibid.

15 Ibid.

16 LaMar T. Empey, "Crime Prevention: The Fugitive Utopia," in Daniel Glaser, ed., Handbook of Criminology (New York: Rand McNally, 1974), p. 1106.

17 Polk and Kobrin, Youth Development, note 13 above, pp. 2-3.

18 Juvenile Justice and Delinquency Prevention: Report of the Task Force on Juvenile Justice and Delinquency Prevention, prepared by the National Advisory Committee on Criminal Justice Standards and Goals, Washington: 1976, p. 25.

19 Wolfgang, Figlio, and Sellin, Delinquency in a Birth Cohort.

20 See LaMar T. Empey and M. L. Erickson, "Hidden Delinquency and Social Status," Social Forces 44:546-554 (June 1966); Martin Gold, "Undetected Delinquent Behavior," The Journal of Research in Crime and Delinquency, 3:27-46 (January 1966); Thorsten Sellin and Marvin E. Wolfgang, The Measurement of Delinquency (New York: John Wiley & Sons, 1964); H. L. Voss, "Socio-economic Status and Reported Delinquent Behavior," Social Problems, 13:314-324 (Winter 1966); and Robert H. Hardt and George E. Bodine, "Development of Self-Report Instruments in Delinquency Research," Youth Development Center, Syracuse University, 1965.

21 For a collection of materials see, in addition to the above sources: Edmund Vaz, ed., Middle-Class Juvenile Delinquency (New York: Harper and Row, 1967); J. J. Conger et al., Personality, Social Class and Delinquency (New York: John Wiley & Sons, 1966); Jerome G. Miller, "Research and Theory in Middle-Class Delinquency," The British Journal of Criminology, 10, 1:33-51 (1970); Bertram Spiller, "Delinquency and Middle Class Goals," The Journal of Criminal Law, Criminology and Police Science, 56, 4:463-478 (1965); John P. Clark and Eugene P. Wenninger, "Socio-Economic Class and Area as Correlates of Illegal Behavior Among Juveniles," The American Sociological Review, 27:826-834 (December 1962); and Charles P. Smith, "The Relationship of Socio-economic Characteristics to Parole Success," thesis, (University of Arizona, 1961).

22 Irving A. Spergel, "Community-based Delinquency Prevention Programs: An Overview," Social Service Review, 47 (1):18 (1973).

23 See Marguerite Warren, Correctional Treatment in Community Settings: A Report of Current Research, Center for Studies in Crime and Delinquency, National Institute of Mental Health, 1972; idem, "The Case for Differential Treatment of Delinquents," Annals of the American Society of Political and Social Science, 381:49-59 (January 1969).

24 For a discussion of some of these problems, see Lejins, in Amos and Wellford, ed., note 8 above, pp. 1-21; and Empey in Glaser, ed., note 16 above.

25 Michael C. Dixon and William E. Wright, An Evaluation of Policy Related Research on the Effectiveness of Juvenile Delinquency Prevention Programs (Nashville: George Peabody College for Teachers, 1974).

26 For a review of the major theoretical perspectives and assumptions associated with delinquency prevention projects, see The Theory and Practice of Delinquency Prevention in the United States: Review, Synthesis and Assessment, prepared by Jerry P. Walker, Albert P. Cardarelli, and Dennis L. Billingsley for the National Institute of Law Enforcement and Criminal Justice, Office of Juvenile Justice and Delinquency Prevention, January 1976, especially pp. 27-112; and A Comparative Analysis of Delinquency Prevention Theory, vol. I, National Institute for Juvenile Justice and Delinquency Prevention, Law Enforcement Assistance Administration, 1977.

27 William E. Wright and Michael C. Dixon, "Community Prevention and Treatment of Juvenile Delinquency," Journal of Research in Crime and Delinquency, January 1977, p. 60.

28 Richard J. Lundman, Paul T. McFarlane, and Frank R. Scarpitti, "Delinquency Prevention: A Description and Assessment of Projects Reported in the Professional Literature," Crime and Delinquency, July 1976, pp. 297-308.

29 Ibid., p. 300.

30 Ibid., p. 300

31 Ibid., p. 308.

32 For the full results of this study, see The Theory and Practice of Delinquency Prevention in the United States . . ., note 26 above.

33 In addition to the GMIS resource, the following data bases were searched and yielded varying quantities of references, abstracts, articles, and project summaries: The National Criminal Justice Reference Service, the National Technical Information Service, The National Council on Crime and Delinquency, The National Science Foundation, The National Institute of Mental Health, and The

Mechanized Information Center, through which the following data bases were searched: (a) Psychological Abstracts, (b) Educational Resources Information Center, (c) Social Science Information Service, and (d) Multidisciplinary.

34 Primary prevention was defined as those programs functioning in the "pre-apprehension" realm, although it was clearly realized that many such programs would also cater to youth who have committed delinquent acts, as well as those who have been adjudicated as delinquent.

35 See The Theory and Practice of Delinquency Prevention in the United States . . ., note 26 above.

36 Ibid., pp. 145-164, for a detailed review of the findings.

37 Richard A. Cloward, "Social Problems, Social Definitions, and Social Opportunities," in Robert Schasse and Jo Wallach, ed., Readings in Changing Interpretations of Behavior (Los Angeles: Youth Studies Center, University of Southern California, May 1965), pp. 24-51. See also Nathan Caplan and Stephen D. Nelson, "On Being Useful: The Nature and Consequences of Psychological Research on Social Problems," American Psychologist, March 1973, pp. 199-211.

38 See Schur, Radical Non-Intervention, note 2 above.

39 See Amos and Wellford, Delinquency Prevention, note 8 above, pp. 22-36.

40 Challenge of Crime in a Free Society, p. 59.

41 Schur, note 2 above, p. 154.

42 Wolfgang, Figlio, and Sellin, note 1 above.

43 Richard A. Cloward and Lloyd Ohlin, Delinquency and Opportunity, (Glencoe, Ill.: The Free Press, 1960).

**PART III:
HOW WELL DOES IT WORK?**

... and Where Do We Go From Here?

Gentle reader, by this time you have persevered with us through the previous 361 pages of discussions of the state of evaluations and corrections, community crime prevention, courts and police, and juvenile justice. Sixteen prominent authors have reviewed several hundred criminal justice program evaluations and have synthesized the findings within the particular programmatic areas. While we, the Editorial Board, feel that the authors are to be commended for their diligence, they have, nonetheless, raised almost as many questions as they have answered. In this section we will be discussing these questions, some of the problems, and we hope appointing some general directions for the future.

The works in this volume show some general central tendencies about the ultimate audience for criminal justice evaluation. Only an occasional comment or isolated evaluation effort appears to be directed towards the Congress. The public is similarly ignored. The needs of the program operator are given short shrift and only recognized after the fact in the outcome summaries of the typical evaluation report. The program policymaker is occasionally considered, but rarely are the evaluation results articulated in a fashion that is comprehensible or usable for the policymaker. The success-failure judgments found in some evaluations are helpful to a policymaker in making some decisions about the program, but these are generally overgeneralized and nonspecific. The recognition of the sponsoring agency as a potential audience is reasonably widespread. Indeed, in the minisurvey conducted by the Editorial Board and discussed below, this recognition of sponsoring agency needs raises critical, ethical, and operational problems for the evaluation community. The system is even more clear as one notes the competition between the academia-based evaluator, the nonprofit organization-based evaluator, and the profit-making organization-based evaluator. An examination of the documents and the conclusion of many of the studies reviewed by the authors clearly indicates that the principal anticipated audience is the academic community. The language, the format, and the arguments of the evaluation materials are consistent with the argot of this subcommunity. Clearly it is an audience which understands and appreciates the efforts and results of the evaluators. It is clearly a comfortable audience to address, not only because of the compatibility between the evaluator and his audience, but also because of the mutual utility of the evaluation results.

If it is not true that the optimal desirable audience for evaluation should be the academic community, then there are major strategic and tactical decisions that have to be made in the future concerning the organization, design, funding, and management of future evaluation efforts. What other audiences?

- The Congress. Almost since the start of its involvement in the criminal and juvenile justice area, the

Congress has expressed an interest in evaluation results. The interest stems from a need for information on fiscal and programmatic accountability as criminal justice programs reach for a state of the art. Although Congress has expressed an interest in evaluation, it does not seem to support social science evaluation at the same level as it supports physical science evaluation. Whether this is caused by the general lack of credibility for social science within the United States or is a result of the inherent "softness" of social science methods and answers is unknown. Clearly, however, Congress and its parallel State legislatures represent a potential audience.

- LEAA and NILECJ. Although evaluation existed before LEAA, the largest sponsor of criminal justice evaluation work is the Law Enforcement Assistance Administration. The vast majority of the evaluations studied in this volume were funded by one of LEAA's programs. The agency's agenda for sponsoring evaluation has changed frequently over the years. Sometimes its demands for rigorosity have been subservient to the demands of expediency; sometimes the level of sophistication has been overcome by transitory naivete. But the question remains, is the evaluation done for the sponsoring agency or for yet another group?
- Academia. Most of the authors contributing to this volume are either members of the academic community or are the closely allied partners of university-affiliated or university-sponsored evaluation organizations. This community speaks the same language and understands the same values and, thus, the evaluator of a given project has a ready audience. The academicians also have their own interests. Frequently they are seeking new knowledge, new intellectual fodder for dissemination to students and colleagues, and theoretical breakthroughs. By the nature of their community, they communicate readily among themselves and to some other specialists. We suspect that the academic community represents the greatest single audience for evaluation reports.
- Program Policymakers. The programs that are being evaluated and the evaluations themselves are generally brought into being by the decision of a State or local programmatic policymaker. This is true even when Federal funds are used. This person or group makes a decision to initiate a program and supports that initial decision by supporting an evaluation effort. The sponsors' interest frequently is to improve a partic-

ular function or a particular part of the system towards making it more efficient and more effective and, not infrequently, more humane. They are an intelligent and educated community but not necessarily completely current about either criminal justice or evaluation issues. However, because of their area of responsibility, they represent the most significant of the potential audiences.

- The Program Operator. The administrator or manager of the host program in which the evaluation is taking place represents a critical potential audience. The operator has an interest in making the program successful not only in terms of an immediate return upon his effort, but also towards the long-term continuance of the program. His interest in new products and evaluation are both professional and self-interest. As such his interest in evaluation is potentially keen.
- The General Public. The general public, as measured by repeated public opinion polls, etc., has a strong interest in criminal justice programs. This interest is not well articulated nor organized in a traditional operational pattern. But, nonetheless, it serves to provide the general programmatic design goals for all of the groups discussed above. It is the public that ultimately tolerates the existence of an experimental program. It is the public that in the end supports the funding and continuation of the program past the initial startup phase. The public's interest in keen evaluation is sometimes difficult to discern, identify, and communicate with, yet experience demonstrates that it exists.

The corollary question to "Who is the audience for evaluation?" is the question of, "Why do we do evaluations at all?" It is the notion of the Editorial Board that the impetus for doing evaluations may well lie within five major categories. These categories are:

- Curiosity. Curiosity can range from a relatively simple inquiry as to what works or doesn't work to the finding of new knowledge or new items to teach or publish. This curiosity may exist for any of the above audiences, but its identifying characteristic is that it is satisfied with knowledge alone.
- Fine Tuning. Fine tuning is a managerial concept. It refers to an evaluation effort designed to help the program manager tune his operation to better meet the ongoing program goals. In this sense, fine

tuning evaluations share a common history with management information systems. Fine tuning efforts share the common characteristic that they are designed to make an ongoing program work better.

- Major Programmatic Decisions. As a program evolves through time, significant programmatic decisions have to be made which can radically change the nature and future of the program. Evaluation can play a vital part in this process. It can provide the programmatic operator or policymaker with information about alternatives that is vital to making major programmatic decisions. Common to this category is the characteristic of evaluations pointing out desirable and undesirable options for action within a programmatic area.
- Guns-or-Butter Evaluations. The criminal justice program competes for finite resources with a host of other programs. Decisions of whether to continue or discontinue a program are constantly and repeatedly made. Evaluation can play a vital role in these decisions. Common to this kind of evaluations are conclusions compared rigorously to some pre-stated outside goal such as cost or specific behavior modifications.
- Evaluation for Gain. Conducting evaluation can be a rewarding experience. The reward can be financial, in the case of contract research tasks, or in personal advancement and recognition, as among academicians. Evaluations within this area may entail ethical questions, organizational questions, and validity questions.

The Editorial Board, in reviewing the authors and from their knowledge of the studies considered by the authors, agreed that the overriding reason for the studies reviewed in Part II was curiosity. There were very few references to fine tuning, major programmatic decisions, or guns-or-butter judgments.

If this is true, then the next question is, "Is curiosity a sufficient justification for the expenditure of resources for criminal justice evaluations?" Knowledge for knowledge's sake has its ritual, but perhaps it is too subtle or too diffuse to be relevant for the dynamic, operating world.

WHERE ARE WE NOW?

The Editorial Board, in attempting to address the issues of this section, decided to sample the opinions of its peers and contemporaries on these four questions:

- What are the most important findings in criminal justice evaluation?
- What are the most important questions to be answered next?
- What are the critical issues facing criminal justice evaluation?
- What are the dangerous pitfalls?

Respondents included most of the authors and outside Editorial Review Board members of this volume and its companion, Crime and Justice: An Annual Review of Research. These respondents also nominated other respondents. Admittedly, it was a small and, in techno-statistical terms, biased sample. Also, the questionnaire was not multiple-choice; respondents answered in their own words. Thus the coding used to classify the answers inserts a subjective element which the editors thought unavoidable. Not all respondents gave five answers to every question, and some gave the same answer twice (according to the coding of their responses). Thus, with only 27 respondents, the last two questions received similar answers 31 times each!

A. What are the most important findings in criminal justice evaluations?

- | | |
|--|----|
| 1. Preventive Patrol doesn't work. | 15 |
| 2. Rehabilitation doesn't work. | 15 |
| 3. Police response time is irrelevant. | 6 |
| 4. Money or job on release positively effects recidivism. | 6 |
| 5. Employment status affects bail and sentence outcomes. | 5 |
| 6. Patrolmen can handle investigations. | 4 |
| 7. Quick short interventions affect criminal behavior. | 4 |
| 8. Training does (doesn't) affect criminal justice staff behavior. | 4 |
| 9. Women are as effective as men in criminal justice work. | 3 |
| 10. We don't know anything. | 3 |
| 11. Diversion programs widen the criminal justice net. | 3 |
| 12. There are disparities in sentencing. | 2 |

B. What are the most important questions to be answered next?

- | | |
|--|----|
| 1. What are the impact(s) of determinate, indeterminate, flexible, or proscribed sentencing? | 15 |
| 2. What is deterrence? | 13 |

* Chicago: University of Chicago Press, forthcoming 1979.

3.	Does rehabilitation work?	13
4.	What are optimal patrol strategies?	12
5.	What programs are costeffective?	9
6.	How does agency organization affect program out-comes?	8
7.	Effect of money or job on successful survival of ex-inmates.	5
8.	What is the role of the citizen in criminal justice operations?	5
9.	What is criminality?	3
10.	Can evaluation results be believed?	3
11.	Impact of prosecutor discretion.	3
12.	What is incapacitation?	2
13.	What are optimal assignments of police manpower?	2
14.	What is the utility of parole?	2
15.	What is the total system impact of alternatives?	2
16.	Validate National Evaluation Program studies.	1

C. What are the critical issues facing criminal justice evaluation?

1.	Improving the methodology of evaluations.	31
2.	Finding a way for the findings to be used.	21
3.	Getting more money for evaluations.	19
4.	Finding good staff resources to conduct evaluations.	19
5.	Finding ways to have the evaluation results affect program operations.	12
6.	Credibility of evaluations.	7
7.	Results must be disseminated.	5
8.	Evaluation and research must be recognized as different but equal.	3
9.	Evaluation and research should have equal status.	3
10.	Evaluation should be costeffective.	3
11.	Recognition of urban/rural/cross-cultural values in evaluations.	2

D. What are the dangerous pitfalls?

1.	Poorly done evaluation design and methodology.	31
2.	Unsound and/or poorly done data analysis.	28
3.	Unethical evaluations.	20
4.	Naive and unprepared evaluation staff.	10
5.	Poor relationships between evaluation and program staff.	10
6.	Cooptation of evaluation staff and/or design.	8
7.	Poor quality data.	6
8.	Poorly done literature reviews of subject area.	4
9.	Focusing on the method not the process.	1

To no one's surprise, 19 respondents saw the need for more money and of finding better staff resources as a "critical is-

CONTINUED

4 OF 5

sue." Unsoundness of data analysis and naivete of staff were frequently cited pitfalls.

In general our sample of evaluators is quite unimpressed with the products of their work. That is not to say that individual evaluators don't think their individual products are good, but in the main they find other evaluations unimpressive, unconvincing, and not particularly relevant. There are a handful of exceptions, notably the Kansas City work and the work of Robert Martinson.

Three perceived needs have emerged among evaluators, constituting a central tendency, namely:

- The need for improved methodologies and methodological practices. These comments focus on the need for an improved technology in the design, development, and implementation of methods of evaluation in a given setting. Relatively little emphasis is placed on the analytical components.
- Greater training and understanding of evaluations. This point is generally expanded to include greater understanding on the part of evaluators, program operators, and policymakers. The sense of the position is that if these three groups were better trained in what evaluation was and how it worked, they would understand it better and by inference use it more and be influenced more by it.
- Greater appreciation and understanding of evaluation results. The sense of this proposition is that few people of significance outside the evaluation community, (and not infrequently inside the community) have any real appreciation or understanding of evaluation results. In general this position is postulated from the perspective that it is not the evaluation community's fault that it is not appreciated or understood, but rather the ignorant outer community.

In examining the apparent central tendencies emerging from the survey, the Board was struck by the strong strain of self-serving attitudes and beliefs expressed by the polled evaluation community. Also:

- The Board is also impressed by the disbalance of what was being evaluated. The overall majority of the evaluations were outcome evaluations. The Board feels that this focus on outcome virtually to the exclusion of process and impact reflects poorly on the evaluation community. It is certainly desirable to seek a balance among the three.

- The Board was struck by the fact that the evaluations were almost exclusively aimed at a very limited audience which had very narrow interests. The primary audience of the evaluation effort was the evaluation community itself rather than the the greater world within which it operated.
- The Board was disappointed by the generally unimaginative approaches used in evaluations. With the wide pallet of available evaluation methodologies, it appeared odd that so little originality was present. The Board also ascertained intellectual rigidity in many of the evaluations. The classical evaluation design would be clung to when its utility had been exhausted. Evaluators insisted that only apples can be compared to apples and oranges only to oranges, and made no effort to give recognition to the fact that apples are going to have to be compared to oranges. Quantification, while of great value, not infrequently seems to become an end rather than a means.
- The Board was also struck by the general ignoring of the Hawthorne Effect. Only one of the authors discussed the Hawthorne Effect (see note 26, p. 256) and few of the evaluators appeared to be cognizant of it.
- There was a strong sense in the evaluation community that the ethics, and by inference, the validity, of the community itself are open to question. The existence of unethical practices are recognized but there does not appear to be any concomitant effort towards reform or control.

WHERE SHOULD WE BE GOING?

From the work that has gone into the preparation of this volume, there does emerge some general marching orders which can be transmitted to the evaluation community. These marching orders may serve as a guide for the future evaluator. They are as follows:

- The evaluation effort should be honest, ethical and competent.
- Evaluation results should be translated into policy alternatives, programmatic alternatives, or both, better to serve a wider audience.
- The evaluation community must become more collected. It must suppress the argot of evaluation and expand its interest to embrace the greater community.

- The evaluation findings must be related to the real world both in the past, the present, and the future.
- Potential replicability of programs which are evaluated should always be addressed. The impact, the outcome, and the process of the evaluation should be recognized and reported.

The National Institute of Law Enforcement and Criminal Justice, among its duties, is charged with aiding jurisdictions seeking to achieve proper evaluation goals. How does it do this?

- The Office of Research Programs sponsors exploratory approaches to criminal justice problems.
- The Office of Research and Evaluation Methods sponsors methodological research, focusing on evaluation measurement problems.
- The Office of Program Evaluation sponsors evaluations of selected projects, including State and local criminal justice initiatives, many of which are described in this volume.
- It is when these ideas reach the Office of Development, Testing, and Dissemination that their National impact is felt. ODTD:
 - Identifies and develops program models,
 - Designs and sponsors field tests,
 - Supports training workshops,
 - Spreads information on the findings of prior research and evaluations, and
 - Tests and develops standards for equipment used by criminal justice agencies.

Full circle for this volume: "Does it work?" and "How well does it work?" NILECJ can be a willing partner to jurisdictions seeking answers to these evaluative questions.

Justus Freimund, Manager,
Criminal Justice Services,
National Criminal Justice Reference Service,
for the Editorial Board

**U.S. DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
WASHINGTON, D.C. 20531**

**OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, \$300**

**POSTAGE AND FEES PAID
U.S. DEPARTMENT OF JUSTICE
JUS-436**



**SPECIAL FOURTH-CLASS RATE
BOOK**

END