RANDOMIZED FIELD EXPERIMENTS IN CRIMINAL JUSTICE AGENCIES: WORKSHOP PROCEEDINGS

108719

RANDOMIZED FIELD EXPERIMENTS IN CRIMINAL JUSTICE AGENCIES: WORKSHOP PROCEEDINGS

Richard O. Lempert and Christy A. Visher, Editors

Committee on Research on Law Enforcement and
the Administration of Justice

Commission on Behavioral and Social Sciences and Education

National Research Council

August 1987

NOTICE: This project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

This project was sponsored by the National Institute of Justice, U.S. Department of Justice, under Grant No. 84-IJ-CX-0082. The contents do not necessarily reflect the views or policies of the grantor agency.

Available from: National Institute of Justice through the National Criminal Justice Reference Service, Rockville, Maryland, 301-251-5500.

WORKING GROUP ON FIELD EXPERIMENTATION IN CRIMINAL JUSTICE

Richard Lempert (Chair), School of Law, University of Michigan

Richard Berk, Department of Sociology, University of California, Santa Barbara

Robert Boruch, Department of Psychology, Northwestern University

Anthony Bouza, Chief of Police, Minneapolis Police Department

Shari S. Diamond, Department of Psychology, University of Illinois at Chicago, and Sidley & Austin, Chicago

E. Michael McCann, District Attorney, Milwaukee County

Albert Reiss, Jr., Department of Sociology, Yale University

Daniel Rubinfeld, School of Law, University of California, Berkeley

Jeffrey A. Roth, Study Director

Christy A. Visher, Research Associate

COMMITTEE ON RESEARCH ON LAW ENFORCEMENT AND THE ADMINISTRATION OF JUSTICE

1986-1987

Norval Morris (Chair), School of Law, University of Chicago

Richard Lempert (Vice Chair), School of Law, University of Michigan

Anthony V. Bouza, Chief of Police, Minneapolis Police Department

Jonathan D. Casper, Department of Political Science, Northwestern University and American Bar Foundation, Chicago

Jacqueline Cohen, School of Urban and Public Affairs, Carnegie-Mellon University

Shari S. Diamond, Department of Psychology, University of Illinois at Chicago and Sidley & Austin, Chicago

David P. Farrington, Institute of Criminology, Cambridge University

Charles F. Manski, Department of Economics, University of Wisconsin

Albert J. Reiss, Jr., Department of Sociology, Yale University

John Rolph, The Rand Corporation, Santa Monica, California

James F. Short, Jr., Social Research Center, Washington State University

Kurt L. Schmoke, State's Attorney, Baltimore, Maryland

Patricia McGowan Wald, U.S. Court of Appeals for the District of Columbia Circuit

Stanton Wheeler, Amateur Athletic Foundation of Los Angeles, on leave from School of Law, Yale University

Samuel Krislov (ex officio), Cochair, Panel on Statistical Assessment as Evidence in the Courts; Department of Political Science, University of Minnesota

Ann Witte (ex officio), Chair, Panel on Taxpayer Compliance Research; Department of Economics, Wellesley College

CONTENTS

THE WORKSHOP

Over the years, scientific knowledge has been a major weapon in the fight against crime. One technique, the randomized field experiment, has recently achieved some prominence for its potential to transform scientific findings into immediate practical knowledge. To illustrate: in 1982 the Minneapolis Police Department conducted an experiment that required police officers responding to domestic assaults to follow a predetermined procedure that ramdomly selected one of three actions: arrest the offender, offer advice to the offender or victim, or order the offender off the premises. Arrest appeared to be more effective in forestalling future assaults than either of the other two strategies. Other experiments have tested the efficacy of different sanctions for drunk driving, the value of providing financial support to parolees, and the use of electronic supervision of arrestees pending trial.

In randomized experiments, the intervention is assigned to persons, cases, or other units according to a random schedule. That is, each person, case, or other unit has the same probability as any other of being assigned to either the intervention or the control group. This procedure ensures that the differences in outcomes can be attributed to the interventions rather than to any characteristics of the units themselves. Quasi-experiments are another technique used to compare outcomes for units that received an intervention with outcomes for units that did not, but there is an important difference: randomization is not used in their assignment. Statistical techniques such as regression analysis, time-series analysis, and analysis of variance must be used to adjust for systematic differences between the units themselves that may contribute to differences in outcomes.

To explore these techniques, on March 13-14, 1987, a distinguished group of criminal justice researchers, practitioners, and policy makers attended a workshop on randomized field experiments conducted by criminal justice agencies. The workshop was convened under the aegis of the National Research Council's Committee on Research on Law Enforcement and the Administration of Justice, with support from the National Institute of Justice (see Appendix A for the workshop program). A total of 90 people attended the workshop (see Appendix B). The workshop was planned by the Working Group on Field Experimentation in Criminal Justice, which consists of four committee members and four other members selected for their knowledge of randomized experiments and quasi-experiments in criminal justice research (see Appendix C for biographical sketches).

The workshop had four major purposes:

(1) For researchers and policy makers to share perspectives on promising experimental and quasi-experimental techniques for research aimed at providing a practical knowledge base for improving criminal justice administration;

(2) To disseminate and discuss results and insights from randomized field experiments and quasi-experiments in policing, prosecution, court decision making, and offender supervision;

(3) To discuss issues that arise in planning and conducting field experiments, including choosing topics, potential legal and ethical problems, the exigencies of day-to-day management, methodological issues, and the interpretation of results; and

(4) To provide a forum for practitioners, research sponsors, and researchers to exchange views on future prospects for experimental research in criminal justice.

The workshop presentations in plenary sessions focused on specific experiments as well as on such broad topics as "Uses and Limits of Experiments," "Legal and Ethical Issues," and "Alternatives to Randomized Experiments." Transcripts of the presentations are included in this report. In two follow-up discussion periods, the participants worked in smaller groups that focused on four specific areas of criminal justice practice: policing; pretrial release, prosecution, and diversion; adjudication and sanctioning; and offender supervision. The goal of each discussion group was to identify promising experiments and to discuss design, implementation, and management issues that could arise in conducting experiments.

A central message emerging from the presentations and discussion sessions was that changes in the criminal justice system occur both gradually, sometimes in response to changing social science understanding of crime problems, and on occasion more rapidly, as the system is pressed to do something about perceived crises. In either case, however, the changes typically reflect commonsense views of reality rather than the systematic observation and testing of alternative programs and procedures. These commonsense views need to be tested and complemented by systematic, reliable knowledge to guide crime control.

Sociologists, economists, psychologists and others have traditionally applied their disciplines' usual methodological approaches to expand knowledge about criminal justice. These methodologies include laboratory experiments, ethnographic investigations, regression analysis, and various techniques of

causal modeling. Each technique has its place, and together they are essential to a full understanding of crime and the criminal justice system. The randomized field experiment has two major advantages: The first is that treatments are delivered by and to the same types of people who are involved in the ordinary workings of the criminal justice system. The second is that randomization is an efficient control for many idiosyncratic factors that might otherwise plausibly explain the apparent impact of a criminal justice intervention. One reason that experimentation has been rare in the past is that, more than any other technique, it requires close cooperation between social science researchers and those who operate criminal justice organizations. For these reasons, a conference bringing together criminal justice researchers and practitioners to discuss this topic was timely, appropriate, and potentially useful.

The group attending the workshop was somewhat larger than originally anticipated, a reflection of the burgeoning interest in field experiments in the practitioner and scholarly communities. The 25 speakers and discussion leaders included researchers, state and local criminal justice practitioners, and federal officials. Attendees from the practitioner community included police officials, pretrial services managers, district attorneys, judges, court administrators, and probation officials. The audience included practitioners and researchers who had conducted randomized field experiments or who were actively considering this option as well as people who were less knowledgeable about but interested in this research technique.

Among the background materials for the workshop were 20 commissioned synopses of criminal (and a few civil) justice experiments. Each synopsis provides a two-paragraph summary of the purpose and results of the experiment and other details, including target group, type of treatments, assignment

method, outcome measures, and reported implementation problems. This work provided attendees with a convenient assembly of information on criminal justice experiments (see p. 220). The other materials distributed in advance of the workshop were published articles about research design (Diamond, 1987), nonexperimental criminal justice research (Chambers, 1977; Pierce and Bowers, 1981), and ethical issues in experimental research (Meier, 1972; Shapard, 1985).

Nine themes emerged from the workshop, which collectively might be considered guidelines for running randomized field experiments or important lessons for future experimentation in criminal justice. The next few pages present these themes and summarize the closing speakers' views on priorities for future research. These themes are not intended to address all the substantive and methodological issues that must be considered in designing and carrying out field experiments. Rather, they provide a brief perspective on the issues that emerged as especially important during workshop discussions. The literature on social experimentation is extensive. For the interested reader, several general references on experimentation are included in the bibliography that follows this summary.

WORKSHOP THEMES

(1) **Choose an interesting problem--a policy question that people really care about or an existing procedure that clearly needs improvement.**

Although experiments need not be costly, they often are, imposing burdens on administrators, usually requiring the assistance of personnel in criminal

justice organizations, and possibly imposing burdens on experimental subjects. Field experiments should therefore not be undertaken on unimportant issues. Field experiments are particularly helpful for choosing among competing options and for resolving uncertainty about the right choice. Choices of questions and options should reflect community attitudes and values as well as the interests of criminal justice practitioners and researchers.

(2) **Legal and ethical issues that may arise in criminal justice experiments can usually be solved by creative thinking, alternative design strategies, compromise, and some foresight about potential problems.**

Many of the legal and ethical issues that surround experiments arise first at the design stage and can usually be cooperatively resolved within the experimental setting. Although the problems seldom admit of standard solutions, there are certain basic principles that apply. One is that the research design should impose the fewest burdens on those whose cooperation is needed to answer the questions it addresses. A second is that the impact and costs of the proposed intervention must be weighed not only against the costs of having only anecdotal or more equivocal knowledge about the effectiveness of interventions, but also against the costs of continuing current programs and policies that are potentially ineffective or counterproductive.

Perhaps the most common ethical and political hurdle faced by investigators is that the nature of field experiments demands that similarly situated actors be treated unequally. It is generally conceded that this is not a serious problem if the intervention treats subjects less harshly than the status quo

(e.g., a random sample of persons who ordinarily would be jailed are given intensive supervised probation). Indeed, if an agency's resources are inadequate to deliver a treatment to all members of an eligible population, randomized allocation of the treatment is often regarded as a fair approach, totally apart from the fact that it creates the possibility of important gains in knowledge. Even randomized imposition of strong sanctions may sometimes be justified, especially in situations in which, without the experiment, a new, harsher approach might be extended to all.

(3) **To achieve the full advantages of experimentation, the random assignment of persons, cases, or other units into treatment and control groups must be rigorously maintained throughout the experiment. Deviations from strict randomness should be monitored and noted.**

Random selection is not biased, is not arbitrary, and in a variety of circumstances has been upheld by the courts as an appropriate research tool to achieve certain goals, including assessing and improving program effectiveness. But complete randomization is difficult to achieve. Ethical or legal considerations may pose barriers to certain kinds of randomized manipulations, and efforts under day-to-day pressures to carry out randomization inevitably involve some errors in assignment. Nevertheless, considerable success may be achieved by close cooperation between researchers and practitioners. To the extent the randomization efforts fall short, statistical techniques may sometimes produce valid conclusions despite the

errors. The potential for damage control is greater if possible failures of randomization are anticipated so that they can be taken into account in the experimental design.

(4) Not all research problems are suitable for randomized field experimentation. The choice of designs and methods of investigation turn on both the questions to be answered and the available data. Quasi-experimental and other retrospective techniques may be more economical or more viable ways of addressing some research problems, and nonquantitative methods such as ethnographic investigation may be appropriate for other questions.

The appropriate research strategy will depend on what must be evaluated, the conditions of the intervention or innovation, and previous theoretical and empirical work. New laws, for example, typically apply unconditionally, thereby precluding randomized experiments, but quasi-experimental time-series designs may provide a good measure of their effects. Ethical concerns may preclude some randomized treatments (e.g., those involving very severe sanctions), but the systematic study of natural variation may reveal salient effects. For some questions, very different research approaches may be more appropriate. For example, if we wish to learn how to break up drug dealer networks, the infiltration and observation of such networks may be an important first step.

(5) From the outset of study design and planning, field experiments require continued teamwork and close cooperation between researchers and practitioners. Failure can be unwittingly designed into an experiment before it begins.

Field experiment designs need to fit local practices. Cooperation must therefore begin with the initial design of the experiment. It is important to involve all key personnel in the project as early as possible; they should understand the nature of the experiment and its potential importance. Since experiments will require some organizational change, at least temporarily, incentives should be built in to secure the cooperation of the staff who must apply the experimental treatments. The success of an experiment is directly related to the commitment of all participants from beginning to end.

However, close involvement of researchers and practitioners may have an unanticipated consequence. Awareness of being involved in an experiment may lead people to perform differently, perhaps more effectively, than they would if they were administering the intervention as part of a daily routine. This phenomenon is known as the "Hawthorne effect," named after a series of experiments at the Hawthorne Electric plant, that revealed that the experience of being separated out for experimental treatment in itself could improve performance, apart from whether the intervention effect was positive or negative. If subjects or implementors of a field experiment are affected by the novelty of the experience, then it is less likely that the experimental effects will recur in jurisdictions that adopt the policy or in the test jurisdiction once the intervention becomes a part of normal operations. A careful design will check for the existence of Hawthorne-like effects. When these effects are identified, statistical modeling may allow the researcher to correct for their implications.

(6) What one gets out of experiments depends on what one puts into them.

Experimentation is not an easy research method; it takes time, effort, and careful attention to anticipate and avoid problems and to emerge with results that can stand up to scrutiny. One must begin with a sensitive understanding of the field situation and the wisdom that practitioners have acquired over the years. Good experimental design requires that practitioners and researchers work together from the start. The requirements of good design include: the selection of problems that lend themselves to experimentation; procedures to ensure close adherence to the randomization plan (perhaps being carried out by a neutral or "blind" observer); the need to understand the proposed treatment and what it entails before the experiment begins; possible pilot tests of treatment procedures or a short trial-run period; the measurement of relevant outcomes in multiple ways (e.g., through records, interviews); a clear definition of the proposed target group and advance investigation of its size (the number of potential experimental subjects is often overestimated). When these requirements are met, the resulting payoffs are likely to be especially high because one has special confidence in attributing effects to policies being tested.

(7) The purpose of experimentation in criminal justice is to inform policy, not to make policy. Experiments provide information about policy options, but policies often have several goals, so experimental results are rarely sufficient for selecting the "correct" policy.

A common result in field experiments is that the new program or procedure being tested has no measurable effect. Such negative results must, like positive results, be treated cautiously because potential effects may not have been adequately measured. If the finding that an innovation has no effect proves valid, it can be extremely useful. For example, an experiment may demonstrate that the existing policy is actually superior to proposed innovations or show that two programs are equally effective, although one might be less costly. At the other extreme, finding a substantial experimental effect of a new program also does not resolve policy problems. The policy maker must still weigh costs and values. For example, an experiment might reveal that jailing speeders rather than fining them has a substantial deterrent effect, but the policy maker still must decide whether incarcerating speeders is a wise use of scarce jail space.

(8) **Experiments involve political risks, which must be understood and confronted.**

Those who conduct experiments will face accusations that they are playing with people's lives--introducing differential treatment into a system that is supposed to vigorously pursue equality of treatment. Such accusations are legitimate expressions of fundamental social values and may call attention to experiments whose objectives fail to justify the social or individual costs they impose. But such objections should not be either made or accepted mechanically. It is too easy to overlook the fact that maintaining the status quo is also a form of treatment, and that inconsistent responses to people and situations (e.g., discretion) occur throughout the criminal justice system.

Experimentation may systematize decision making and also allow policy makers to learn from variations that would in the usual course occur haphazardly. While field experiments may seem radically innovative, introducing innovations experimentally rather than universally is often a conservative procedure, for it allows the criminal justice system to abort plausible policy changes that turn out to be wasteful or counterproductive. The decision not to experiment, like the decision to experiment, has its political aspects, and objections to experiments should be weighed against the costs of refusing to proceed this way.

(9) **Replicating studies is critical to ensuring that the results will generalize to other locations and is usually necessary to justify widespread implementation of experimentally successful treatments.**

Communities differ in terms of their demographic, organizational, and political characteristics, as well as in their decision making about criminal justice policies. The results of research in one community may reflect the interaction of a treatment with these specific characteristics. This is particularly difficult to control in criminal justice, because the effects of an innovation by one agency, such as the police, may be intimately tied to the behavior of other agencies, such as the prosecution or the courts. Replications help to point out such possible links. Moreover, if experimental results are due to Hawthorne-like effects, as discussed earlier, generalizations of findings to other communities may be misleading and the results may even disappear in the experimental community when the treatment is introduced on a larger scale and implemented by the regular operating system. If an experiment fails to

replicate, researchers and policy makers alike are alerted to the need to search for those factors that led to the initial experiment's success (or failure). If experimental results are replicated in a variety of settings, new sites can implement the innovation with some confidence that it will work as intended.

PRIORITIES FOR FUTURE RESEARCH

The closing session of the workshop addressed future prospects for randomized experiments in criminal justice from three perspectives: a practitioner (Malcolm MacDonald, American Probation and Parole Association and Texas Adult Probation Commission), a federal research sponsor (James K. Stewart, National Institute of Justice), and a quasi-governmental corporate sponsor of state and local justice innovations (David Tevelin, State Justice Institute).

Although these speakers emphasized slightly different priorities for future research, their messages contained similar themes. All three speakers stressed that the need to ascertain whether successful experimental programs are effective in different types of communities means that researchers, practitioners, and funding agencies should place a high priority on replication. They reminded the attendees that a thorough knowledge of the specific target population, the program objectives, and the treatment or intervention strategies is important not only to successfully conducting a randomized field experiment, but also to implementing programs based on others' research results.

Important topics for randomized field experiments in criminal justice appear almost limitless. Some of the areas that the workshop participants

identified as particularly promising include: procedures for imposing, collecting, and enforcing fines; educational programs for convicted offenders; the use of offender classification systems in decision making; police resource allocation; and alternatives to incarceration. Since tackling these and other criminal justice problems will often require coordinated interventions involving more than one criminal justice agency, programmatic experimental research and effective treatment strategies are likely to involve the extensive cooperation of key actors in the criminal justice system.

In his concluding remarks, Stewart emphasized that it is in the best interests of the criminal justice community--both practitioners and researchers--to be receptive to experimentation. In this era of tight fiscal resources, we must avoid the costs of adopting new technologies and policies that may not be effective. The time has come, Stewart said, to move the criminal justice community from a craft--which bases its knowledge on tradition, "seat of the pants" technologies, and intuition--to a profession in which decisions are based on sound research involving testing and replication. Criminal justice field experiments are one of a variety of research techniques that are tending to move the criminal justice community in this direction. The conference attendees recognized that, even as a technique for acquiring knowledge, the field experiment is not a panacea, but they agreed that among the research techniques likely to add to usable knowledge, the criminal justice field experiment is a particularly promising one.

BIBLIOGRAPHY

Archibald, R. W., and Newhouse, J.P.
1980 Social Experimentation: Some Whys and Hows. R-2479-HEW. Santa Monica, Calif.: The Rand Corporation.

Chambers, D.L.
1977 Men who know they are watched: Some benefits and costs of jailing for nonpayment of support. Michigan Law Review 75:900-940.

Cook, T.D., and Campbell, D.T.
1979 Quasi-Experimental Design Issues in Field Settings. Skokie, Ill.: Rand McNally.

Diamond, S.S.
1987 Methods for the empirical study of law. Pp 637-696 in L. Lipson and S. Wheeler, eds., The Handbook of Law and Social Science. Russell Sage Foundation.

Federal Judicial Center
1981 Experimentation in the Law. Report of the Federal Judicial Center Advisory Committee on Experimentation in the Law. Washington, D.C.: United States Government Printing Office.

Fienberg, S.E., Singer, B., and Tanur, J.M.
1985 Large-scale social experimentation in the United States. Pp. 287-326 in A.C. Atkinson and S.E. Fienberg, eds., A Celebration of Statistics. New York: Springer-Verlag.

Gilbert, J.P., Light, R.J., and Mosteller, F.
1975 Assessing social innovations: An empirical base for policy. Pp. 39-193 in C.A. Bennett and A.A. Lumsdaine, eds., Evaluation and Experiment: Some Critical Issues in Assessing Social Programs. New York: Academic.

Haggstrom, G.W.
1976 The pitfalls of manpower experiments. In H.W. Sinaiko and L.A. Broedling, eds., Perspectives on Attitude Assessment: Surveys and Their Alternatives. Champaign, Ill.: Pendleton.

Lempert, R.O.
1966 Strategies of research design in the legal impact study: The control of plausible rival hypotheses. Law and Society Review 1:111-132.

Meier, P.
1972 The biggest public health experiment ever: The 1954 field trial of the Salk poliomyelitis vaccine. In J. Tanur, F. Mosteller, W. Kruskal, R. Link, R. Pieters and G. Rising, eds., Statistics: A Guide to the Unknown. San Francisco: Holden-Day.

Pierce, G. and Bowers, W.
1981 The Bartley-Fox gun law's short-term impact on 1981 crime in Boston. The Annals of the American Academy of Political and Social Science 455:120-137.

Riecken, H.W., and Boruch, R.F., eds.
1978 Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York: Academic Press.

Rossi, P.H., and Freeman, H.E.
1982 Evaluation: A Systematic Approach. Beverly Hills, Calif.: Sage.

Shapard, J.E.
1985 The ethics of experimentation in law enforcement. Pp 418-429 in W. Geller, ed., Police Leadership in America. New York: Praeger.

P R O C E E D I N G S

Friday, March 13

Welcome and Introduction

PROF. LEMPERT: I would like to call this Workshop to order, if I may. This Workshop that you are attending is the fourth invitational meeting in a series that is a joint venture of the National Research Council and the National Institute of Justice. The Workshop has been planned by a working group of a committee of the National Research Council, the Committee on Research on Law Enforcement and the Administration of Justice.

I am Rick Lempert. I am Vice chair of this committee, a title that acquired new meaning when I walked along Bourbon Street last night, and I have, also, been the Chair of the working group that has planned this particular Workshop.

The committee institutes or sponsors a wide variety of investigations in the areas of criminal justice and the administration of justice more generally. Over the years, the National Research Council through this committee has produced many reports on matters of great concern to all who work on criminal justice issues. You have at your tables a brochure which describes the various reports and other projects that the committee has been responsible for. Some of the reports are available to you for purchase, if you are interested in learning what research has to tell us about various subject matters. That is the word from our sponsors or at least one of our sponsors. You are going to hear another sponsor very shortly.

As I said, I have chaired a working group that planned the agenda and selected speakers. This was truly a working group, not just in the sense that we met in places like Chicago in late November, but, also, in the sense that, as you will see, the various members of the working group are on the program because all of us either agreed to or were dragooned into agreeing to chair a particular session or give a particular talk. Just to both alert you to some of the people you will be seeing and to thank those people who served in the working group, let me simply give you their names; those who were appointed by the committee are: Dick Berk, Bob Boruch, Tony Bouza, Shari Diamond, Mike McCann, Al Reiss and Dan Rubinfeld. Many of you know the people from the National Institute of Justice who worked with us: Joel Garner and Dick Linster.

Those of you who have been involved in anything like planning a conference or chairing a working group realize that while your name goes on the program, and you get the credit, you don't really do the hard work; the real work is done by the staff behind the scenes. We have been very lucky in connection with this conference, as we have been generally in connection with the committee's activities to have an absolutely first-rate staff, and I would like just briefly to introduce them to you. I will ask them to stand up so that you can see who really gets the credit here.

First of all, specifically assigned to this project and working very hard has been Christy Visher. Christy?

(Applause.)

PROF. LEMPERT: And then the general staff leader, if you will, of the parent Committee on Research on Law Enforcement and Administration Justice who is involved in everything that goes on under its auspices is Jeff Roth. Jeff?

(Applause.)

PROF. LEMPERT: Then of course, the staff has its staff because they, too, rely on others, and there are two people who are important in terms of the staff's staff. Let me introduce them also. You have in your books that were handed out to you, a series of very, very nicely done summaries of major experiments in the criminal justice research area. This was done by Michael Dennis, a graduate student in the doctoral program in psychology at Northwestern University, in cooperation with us. Would you stand, please, Michael?

(Applause.)

PROF. LEMPERT: Thank you. And then, finally, we always get to the really key person last, the administrative secretary of the committee working with the working group who makes everything happen, Gaylene Dumouchel.

(Applause.)

PROF. LEMPERT: These are the people without whom we would not be here. But we are here, and on behalf of the committee and of the working group, I am very happy to welcome you to this gathering. The gathering is a relatively large one. It is relatively large because much to our delight, we had a very high acceptance rate; when people are interested, it is natural to wonder why. I wondered, and of course the first solution I came up with is that we did a very good job planning the program. We have a lot of interesting speakers and even I am on the

program. Thus, my favorite explanation for the turnout is that there are probably some people here who saw the names of the working group members on the program sent out with the invitations, and they want to hear us speak. The only problem, of course, is that this explanation is wrong. I doubt if any of you came to hear us specifically or, indeed, because the conference seemed especially well planned. My next hypothesis focused on New Orleans in the spring. Maybe that is closer to the mark, but I don't even think this is the explanation.

I believe that the real reason why so many of you are here is simply the topic that we are going to look at, the idea of field experiments in the law. It is a tremendously exciting topic, and I think intuitively all of us realize its excitement and its potential for teaching us, giving us information which will help us all in our work.

The image of the classic practicing professional, the medical doctor, for example, is the image of somebody who relies on academic researchers for basic knowledge and then goes out and applies that basic knowledge in the field. The image of the academic researcher, with respect to those professions is someone who works on an important problem in mind in his or her laboratory, develops a way to solve it and furnishes that information to the practicing professional. We visualize Dr. Salk on the one hand, and the doctor who gives your child the polio shots on the other, working as a team to make life better.

In many circles, the image of the criminal justice professional is somewhat different. It is the image of somebody who has a good deal of practical knowledge and practical learning but is, to some degree, disdainful of the kind of learning that comes out of the universities. Similarly the popular image of the academic researcher in criminal justice

is often far removed from Dr. Salk. It is not a portrayal of a researcher who identifies field problems and says, "Let me generate knowledge to solve them," but rather it is the person who has very abstract theoretical interests which perhaps 100 years from now will tell us something about crime but don't expect information that can be put to work tomorrow.

Now, as a social scientist, as an academic researcher, I have always realized that the latter image is wrong. Social scientists are interested in, deeply interested in, theory; but they, also, are deeply interested in being relevant to the problems that confront society; in working on problems which we must solve to make this country a better place.

One of the happiest aspects, I think, for me of my now almost 6-year association with the Committee on Research on Law Enforcement and the Administration of Justice has been working with practicing professionals and coming to realize how profoundly mistaken the other image is, also, the image of the criminal justice professional at work in the field.

Indeed, it is my experience that you people crave knowledge, seek to apply knowledge, are often sophisticated consumers of knowledge, as you try to find ways to do your job, whether it is policing, corrections, prosecution, that are more effective, to better working, more economic, whatever. Indeed, we have an agency which I think is built on that perception, and that the knowledge needed to attack criminal justice problems can only come about through cooperation between academics and the criminal justice professionals. Indeed, one of the reasons why the area of field experiments is so exciting is precisely because there is probably no place in the knowledge-generation area that requires more teamwork

between those who are in the field confronting day-to-day problems and academic researchers who wish to deepen our understanding of these problems and develop usable knowledge.

This observation brings me to my last and perhaps most pleasant chore of these introductory remarks which is to introduce our first speaker. It brings me to that because our first speaker, I think more than anyone else, combines in one person those two kinds of interests, a sense of the value of knowledge and a sense of the value of practicality, of application. He has over the last 5 years worked with an agency, the National Institute of Justice to bring it more than it ever was into the forefront of those sponsoring research which has genuine theoretical, scientific rigor and value and at the same time genuine, often immediate practical implications. The speaker is no stranger to any of you. So, I will be brief in my introduction.

He is someone who has worn three hats in criminal justice. He has been a practitioner, a researcher and a research sponsor. He has served as a police officer in Oakland, California for 15 years. He moved up the ranks from patrol officer to commander of their criminal investigation division. During his police career he, himself, carried out research on a range of topics, including juvenile diversion, police selection criteria and detective productivity. In 1981, he moved to Washington to become a White House fellow, and he later served as a Special Assistant to then Attorney General William French Smith. In that position his potential was immediately spotted, and he was appointed Director of the National Institute of Justice where since 1982, he has led that agency in sponsoring scores of criminal justice experiments and many more non-experimental research projects. I think he is -- not just in the

sense of providing the money for us, but in the sense of providing a knowledge base for us to discuss -- the person most responsible for our being here.

So, without further ado, it gives me great pleasure to introduce Chips Stewart.

(Applause.)

MR. STEWART: Thank you, Rick, and thank all of you for that very warm introduction. I would like to say that the research community can learn and does learn. After all, given the weather in November, I think that is probably why you selected New Orleans so that all of you could come down here in the beautiful time of Mardi Gras. Let me say that our relationship -- the National Institute of Justice's relationship with the National Academy of Sciences -- has been a longstanding one, and one which I have vigorously pursued to be sure that our relationship remained strong and enduring over the past 6 years.

I have been delighted with the excellent work of the Committee on Research on Law Enforcement and the Administration of Justice, the staff of the National Academy of Sciences and in general, I think that we have had what somebody could call a very beneficial symbiotic relationship. They have not only worked very hard to develop a core of scientists who have a great deal of respect amongst the academic community, but I think that they have -- much to their credit and this audience is representation of that -- brought together, at my request, the practitioners, and the practitioners have worked very hard at not only providing test beds for experiments but, also, ideas. The researchers have provided some excellent analysis and design, and together we have formed a partnership of collaboration that has benefitted all of us here in America.

I think that there are numbers of people in other countries and other places looking to democracies who actually can help manage their populations and governance in society in a much more honest way, in a way that creates less friction and more professionalism. I think that today that this particular meeting represents a very major plateau because essentially you are all sitting as equals here. The recognition that the academics devoted their life to scholarship and to discovery of better ways of measuring events, experimenting to come up with better ideas, but also, that the criminal justice system -- these judges that are here in the audience, the police chiefs, the prosecutors, the people in corrections, the people in parole and probation, all of those who have enormous responsibilities in society is being confronted all the time with enormous social dynamics suggests I think that you are searching for a better way, a better way to make a difference in how we really live as a people. Thus this great experiment that Lincoln talked about and our forefathers talked about lives, and I think it lives very much because of the criminal justice system.

Let me take just a few minutes to talk about my enthusiasm over experimentation. They say that invention is the child of necessity, and I think that to a large extent that what happened in the sixties and then the tail of the seventies was to say, "Listen, is there something we can do that is better about bringing justice in America?" We worked hard, and we experimented a little bit in the sixties and seventies, but it was more or less a social dynamic that was, I think, essentially unplanned. Since then there have been some disciplines that have grown up, some measurements and some things like that, and dedicated individuals, both in the academic sphere and, also, importantly in the practitioner sphere, the idea that can we do our job better was more important.

My idea about experimentation is we have to find out what works in crime control. How do we protect the innocent all the way through our system? How do we fairly treat those transgressors so that we can either recover or do it in such a way as they don't hurt us or our future chances, but that we do it in a fair and humane way?

One of the things that I insisted is that we try to get policy-relevant research. James Q. Wilson talked about that. It was the idea that research on the shelf alone makes no difference. It is like spinning prayer wheels in Tibet. It doesn't change our lives, but if we can get research to change our actions, our activities and our policies, that can make an enormous difference -- we ought to change it in a way that we can understand what the consequences would be. In the past we changed it in a way that we felt was the best idea, and unfortunately, our best intentions and our sense, our visceral reactions turned out in many instances to be wrong, and in fact, our intuition turned out to be wrong.

Now, criminal justice is supposed to be a profession. The idea is "are we a craft or are we a people of professionals who begin to accrete a series of knowledge about the consequences of our actions?" Let me just say a couple of quick things now. It is that I expect that when these panels go on that all of you in this audience engage in vigorous discussions with the panel members. I would like to see this be a very yeasty debate that happens inside here. This is not a teaching session nor is it an authoritative session where the judges lecture to the jury; the academics lecture to the students or the police tell the crowd to move on.

(Laughter.)

MR. STEWART: But what I want to see here is a group of equals who come together to begin to debate from the very best of their knowledge and experience and their judgment to test your ideas. So, I expect vigorous discussion to come out of this.

Also, just in a quick highlight, somebody said that as an example, one of the best field experiments that has been done, and we will talk about that in just a few minutes, was a field experiment in Minneapolis, Minnesota. It is said that the way the police act, and eventually the way that criminal justice acts, the prosecutor acts and the judges act have future consequences, and those consequences could increase or reduce violence and eventually murder. That is a very important thought because if your idea and our idea is to make a more peaceful society, and we want to cut the murder rate, one of the things that we ought to do is a better job with family fights.

That is the most discouraging job that the police have is to go to a place in which they are having a family fight. Judges and prosecutors feel that the wife is not vigorous in her prosecution because she usually bails the guy out because she feels desperate, and the fact is that she may not have an income. In fact, she might have to move out and lose her own house, and it is a terrible burden for her to bear, etc. But in Minneapolis they tried an experiment by surrendering their discretion over whether to arrest. They, in a sense, gave up their judgment -- this, many professionals didn't want to do, but you can see the same thing with doctors who say, "We want to make patients better. We have got to randomize. We have got to give fair and equal treatment to make comparisons as to what happens because of certain judgments we make." I think that if you worry about family fights, you think, well, that is

something you have to handle; that is tough, but if you are worried about murders, and you are worried about assaults, and you are worried about the future of America in terms of the family, all of a sudden, this routine call takes on enormous implications for the rest of us.

The other thing is how about something as simple as drug testing? All this research that has been done, we found out that about 2 percent of your police department, maybe less, is devoted to narcotics enforcement, maybe only 1 percent or 1/2 percent, and yet we hear that is a major problem in many of our cities. Is that a good policy? We hear that there are lots of cases that aren't charged because there is not enough of the evidence there or there are too many people who come in charged with those crimes.

You talk to the police, and they say, "We want to concentrate on burglaries and robberies. We don't have time to take care of the victimless crimes." Studies supported by the National Institute of Justice have shown, remarkably that in fact, high-cost drug use is one of the principal indicators of the career criminal, the high-rate criminal. The National Academy of Sciences helped us review criminal careers. It helps identify objectively and reliably much more than our visceral look or our sense of the person's past record or his current crime, and the question is how do we bring these together?

I was just talking to Judge Bruce Beaudin just before we came in today, and it turned out that about 3 years ago we had a group of pretrial services people a little larger than this that met and I said then that it is very important to try to bring together policy and research because we could make a difference in the way we conduct business in our system.

Afterwards he came up to me, and he said, "You know, I have thought that the pretrial services gathers informations about who to let out on OR release," and he said, "You know, if we tried this idea that you talked about, about trying to reduce the danger to the community in drug testing, we might be able to do that in Washington, D.C." The incredible part was that this was one person who was sitting in a room like this who was struck with the idea that something could be done that is different and better. I am pleased to say that that has been an enormous success, and we have Judge Fred Ugast here, as well, from Washington, D.C. We also have people from New York who are beginning to try this.

We are getting reliable data for the first time, and we are talking about a major breakthrough, but it was a breakthrough because it brought researchers together with the practitioner who said, "Wait a minute, I do see an opportunity here for an experiment, something that can give us some knowledge that can change the way we do business, something that can make us better in our community and something that doesn't break the system; it doesn't overburden the police or the prosecutors, doesn't cause the judges problems; it doesn't fill up either our prisons and tax our people to death or doesn't turn the wolves loose to the sheep in the community."

It is sensible and inexpensive. You know that is what we ought to look at in terms of research. Let me just say that the drug testing is a very complex issue, and I happened to notice on the way out here that Omni magazine was talking about Star Trek, and it talked about assessing the future. I think that is very appropriate here because all of you are taking a look at the future and how we can get our bearings a little better, and one of the headlines was liquid gold.

Now, that was pretty good because I had just heard about a guy who turned gold into lead, and I thought that maybe this might be the other way around. But it happened to say that in response to the growing drug testing phenomenon in both government and private industry, the whimsical, Austin, Texas, entrepreneur with a flair for the irreverent began selling vials of powdered, guaranteed dust-free urine, a business where you would only suspect you could get a sky-high market there. Bird Laboratories began to produce some urine, drug-free urine so that you could use it in your laboratory or you could use it, if asked to supply a specimen, I guess. He switched to powder, and he said that that was much better. He said, "We tried using a freeze-dried process, but this version never panned out." He says, "All you have to do is add water, preferably distilled water," and he said, "The company slogan is," and I hope I don't offend anybody here, but it says, "Pee for pleasure, not for employment."

(Laughter.)

MR. STEWART: The lab is described, are you ready for this, because you scientists here are practitioners, the lab is described as "purveyors of fine urine products. Yellow gowns were de rigueur last winter at the first urine ball to raise money for the urine defense fund." Amazing, here was a great idea we had about trying to protect the community, and we have spawned a derivative industry. Thank God for the entrepreneurship here in America.

What I wanted to try to get across is that decisions have enormous consequences, and many times because we are forced into making the most economical decision in order to process large numbers of people, we want to cut corners or make what we consider to be particular judgments, and we don't realize that this has changed society. In a group

like this here in New Orleans, I hope that 20 years later we will look back on this meeting and say that something important happened that made a change, that we left here with the idea that we could do something different, and that is important, and that is the kind of thing that I am really excited about, and that is the reason that you are all here. All of you have already done something very important. All of you have been major contributors to this accretion of knowledge about a better way to do things, and all of you are important. That is why I am insisting that you be vigorously involved in the discussions and the debate because you are not just here to listen. You have brought with you a constellation of skills, experience and knowledge that could inform all of us, and a small comment from the back of the room may change the way that we all see our particular tasks.

Criminal justice, I want to say, is under a great deal of stress. I think that quite frankly we have a system that has remarkably endured for 200 years, and yet I think it is under a great deal of stress right now because I have a sense that there is an estrangement between criminal justice and the people it serves, its noble purposes and the people who daily try to benefit from what we do. It costs too much, and it does too little. As a result we have had things that have developed like court watchers, people who came in and sat in our courts and began to generate newsletters about the quality of justice.

We have police review boards in many cities. We have prosecutors being "unelected" because of certain ideas and certain knowledge that has gotten out. We have competition for the first time in our lives. I should not say the first time in our lives, but more or less in the last 50 years from the private sector. The police are engaged in very heavy

competition from the private sector and private security because people feel more vulnerable today than ever before.

We find that the prosecutors are getting competition in the private sector, because there are businesses who feel that they are not treated correctly in criminal courts and they go looking for other ways to adjudicate theft, industrial espionage or very complex cases.

We are finding that the courts -- burdened as they are trying to dispense the best justice possible and the envy of many places all over this world -- are, also, looking at things like Adjudication and other private firms, the rent-a-judge firms in California and elsewhere so that you can get access to justice and fair decision up or down in fairly quick time.

I think these are all signs to all of us that this competition is a healthy competition, an acknowledgement that what we do is very important and that, also, people are willing to invest their own money in some alternative.

Now, if we want a society that is based on the rule of law, and if we want a society in which we all work together, then we all have to work together. Right here in New Orleans, we are trying to create the drug-use forecasting system that we have done in so many other cities. We are going to try for the first time in the history of America and maybe any other nation, to get a baseline of real drugs that are being used by the arrestee population.

You think about all of the work that we have done before, NIDA and DAWN, DEA, all of the drug-testing data that we have gotten has all come from surveys, has all come from people who may not be the target group, has all been flawed that way, but if we could take a look through

urine testing of 200 people in each city every quarter, a randomization, we could end up with a baseline that would tell us whether a drug seizure in Colombia made a difference or whether a buy-bust program made a difference or whether release from court on an evidentiary issue made a difference in the quality of life in our community.

We could, also, find out such things as in Lynn, Massachusetts, where we found out surprisingly that when you use drug enforcement at a high rate, a surprising thing happened, the burglaries and robberies began to fall. That's what's interesting about experiments -- sometimes you cannot tell where it is going, but the benefit is when you are looking, you definitely will find something that is more valuable than anything else. Knowledge is the key to our authority and the key to a great society.

Now, I have the great pleasure of introducing to you a man whom I think lives in all of our hearts and all of our minds. He is a great chief of police who begins to show what can be done with experimentation, a certain amount of irreverence and a high commitment -- I have to say that because I get calls from other police chiefs who say, "Gosh darn, I heard him on TV. How can you do these things with him?" I said, "Because he does good work."

The idea is that people who are willing to speak their mind and engage in this kind of debate are extraordinarily important.

I bring to you this morning a man who has made an enormous difference in New York City and in fact, is part of that famous band that has left New York City that has gone on to proselytize throughout America this idea of testing what police do and trying to make a change; a man who is no stranger to debates. He is a man who does not shrink at all from controversy, but a man who spends the time to think before he talks.

He is a man who has gone to the most unlikely place to conduct experiments, and that is Minnesota. It is so cold in the wintertime and so filled with mosquitoes in the summertime, and yet it has produced a member on the Supreme Court. It has outstanding state legislators. It has a great social climate, and it is a surprise because they say, "Why is it there?" So much goes on, and it is a place in which we now all turn our eyes, and we have since about 6 years ago when Chief Tony Bouza came out there. He has also been the police representative on the National Academy of Sciences Committee on Law Enforcement. So, without any further ado, I introduce you to the inimitable, the irrepressible Tony Bouza.

(Applause.)

MR. BOUZA: Thank you, Chips. How to live up to that is going to be a mystery to me, but I want to thank Chips for a generous introduction. No one has done more for experimentation and for research and for analysis and for encouraging the trickle and development of study and progress in the field of law enforcement in my experience than Chips Stewart. We owe him a great debt. I am, also, happy to be on a forum with a man whom I think is the foremost criminologist in America today, Larry Sherman. I can also assure you that Rick Lempert worked our buns off. And Peter Rossi is an eminent scholar I know by reputation, and I am happy to be with him.

If there is anyone in this room who has not been introduced, I think this would be an excellent time to get up and tell us a little bit about yourself.

I am very pleased to have worked on this project - by the excitement that has been generated as evidenced by the turnout, and the fact that there had to be a cutoff, and people who wanted to be here are not here. So, there was a terrific amount of excitement created about this conference, and I think that is very exciting in itself. I will be talking very fast because we Midwesterners tend to do that, hopefully reflectively. I have about four or five outlines available to those benighted spirits who are so thirsting and hungry for knowledge that they must have the pearls of my wisdom firsthand, and I want to remind Stewart that symbiosis happens to be a crime in Minnesota.

(Laughter.)

MR. BOUZA: But he has introduced us, I think, to a fascinating vision, the question of liquid gold and perhaps has brought to us the vision of the tapping of a great natural resource during halftime at a football stadium in this country, and it is a simple question of siphoning the liquids into proper bottling and merchandising it accordingly.

I think there is a great natural industry waiting to be discovered, and perhaps 20 years later that will be the important thing to emanate from this conference.

So, why experiment? That is a good question, really, in my experience. There is a great deal of resistance to change. As Machiavelli reminded us there is a big constituency for the status quo and very little constituency for change. Those who might ultimately profit from change - hard to identify that - hard to touch it and feel it. Anybody who has read Mike Dennis' really brilliant precis of the experiments that have taken place - and I urge you to do that, if you have not already done so because you are never going to find documents that

fascinating and so pithily and briefly described - but it is the precis of the experiments that you encounter the resistance, the critical importance of the practitioners, the individuals who are affected by these experiments and how their willingness, as Chips Stewart reminded us, to work, have produced successes, and those who resist have produced failures. There is very little constituency for change. You are playing with people's lives. There is the media, there is the public. There are politicians. There are pressures, and the easy way is to go along and get along.

The police are an evolutionary ad hocracy, in my experience. They change tactics to suit the time, depending on circumstance and political reality. They create stakeout units. They kill a bunch of people. Political pressures mount, they abandon stakeout units. They create sting operations. They work, and then there are some dysfunctional aspects that appear to surface, and they abandon them. Seat of the pants. We called them soup school experiments in New York, as I was growing up in that department. Seat of the pants stuff, so much of it has evolved. It is an ad hocracy. We had a cop on the beat. Then the telephone got invented, and people stopped opening their windows and shouting for the police. They picked up their telephones. We had to take the cop off the beat, put him in a car. Then we had to give him a computer, and then 20 years ago, almost to the day, March 1967, the President's Crime Commission Report suggested the creation of a three-digit emergency number 911, just 20 years ago. Maybe we ought to be celebrating its birthday, because I think it was March 1967. So, 911 and computers, all of that evolving, rising levels of violence, criminality, traffic problems, service calls and 911 volumes have resulted in us

constantly responding, reacting to these pressures without too much reflection, without too much analysis.

We work in a world of action, and that action must be informed by knowledge, and it isn't, all too often. Experimentation, research, analysis and study are ways to create a solid foundation of knowledge from which to act. We don't know much. There haven't been many experiments in the police field. It is safer to go with traditional methods.

From 1968 to 1980, there was a large flow of funds which, to our eternal discredit, we largely spent buying things like emergency rescue vehicles so that we could go up and down the main streets of our towns in tanks and impress everyone - and threw the whole question of funding into disrepute.

Interestingly enough, now that there is a trickle into the National Institute of Justice, that trickle is kind of golden in the sense that, and I don't mean to make --

(Laughter.)

MR. BOUZA: -- an unintended analogy, but that trickle is finding its way to real experiments and studies. The money did dry up from 1968 to 1980, and really, when you think about it, Jimmy Carter was probably right to kill LEAA in terms of the way it was spending its money which speaks to the question of research and analysis and study now needed, as I speak to you, in the expenditure of funds relating to drug enforcement, for example. The scholars, when the money dried up, left and left the world to darkness and to me.

When scholars and researchers and cinematographers crawl over an agency a great deal of excitement is created by that crawling, an intellectual synergy is created. It encourages the members of that

organization to begin to think creatively. We see that in Minneapolis with the repeat call experiment, the domestic abuse experiment, community crime prevention initiatives, a lot of other stuff, people crawling over the agency, lots of excitement. It looks a bit sloppy and threatening and worrisome, but there is a terrific intellectual vitality and vibrancy.

I am sure that the Renaissance was not an orderly period of enormous creativity. It was a very sloppy business. I can just imagine Michelangelo Buonarroti being summoned by the pope, the pope pointing to the ceiling and saying, "Listen, I would like you to paint the creation, God, Adam, genesis, the whole operation up there on the ceiling," and Buonarroti looking up, I am sure, and I know just what he said. He said, "Sure, Boss, what color?"

(Laughter.)

MR. BOUZA: And this intellectual synergy creates a spillover effect, and it has an effect on others. The important thing about experimentation is the Hawthorne effect, and we have forgotten it. You start experimenting. People begin to think that what they are doing is exciting and interesting, and they begin to take a different view of themselves. They begin to think expansively, creatively, originally, and there is a tremendous amount of intellectual energy lying dormant in police agencies that needs to be inspired. I have only to look to one of my colleagues from Minneapolis sitting in the audience today, Bud Emerson, who has just won a Bush Foundation fellowship, just promoted to sergeant and who carries in his knapsack, as Napoleon Bonaparte reminded us every corporal does, a marshal's baton. So, there is an effect on others, but there is an anomaly here. There is resistance to change, and yet it creates tremendous enthusiasm when it is finally pushed forward. So, you must break through.

To me, the analogy that fits is the breakthrough of the sound barrier. Change is pushing against those enormous forces that begin to suddenly begin to make everything buffet and shake, and what most of us do when that begins to happen is we begin to slow up and pull off the throttle. That is the time to push the throttle forward and burst out and boom into the next dimension. That is what it is all about. There can be no growth without change. There should be no change for change's sake.

You are, also, talking about the personal growth and development of the individual members. We have seen this in Minneapolis, the tremendous enthusiasm, an organization that believes that the chief is a criminal creyin psycho and the great Satan, not to mention the Ayatollah Khomeini brought west, but even so, they are volunteering for new experiments because they sense the excitement, and they want to be a part of it. They want to participate, and we are playing with people's lives, as you must do.

Anyone who read that enormous analysis of the polio vaccine has to understand the tragedy that little boys and little girls had to die in order to bring an oral vaccine that worked against polio and saved the lives of thousands of others.

A few had to die. We do have to play with people's lives, must take risks, must face the press, must face an irate public and educate them. The American people are very tough and very smart, and they can take it, if you give them the truth. There are no panaceas, but it is a way of stumbling to its more useful discoveries rather than to the discovery of "the truth."

An organization needs change in order for growth, as an individual needs change. Organizations have stability. They, also, have stagnation. They must have growth, and they must have progress. Can an organization have too much change? Yes, if it is not carefully thought out and brought about for change's own sake. No, if changes are useful. Organizations can absorb enormous changes. It is my own private, personal view that an organization has reached the appropriate levels of change when every single member of that agency is saying, "Too much is happening too fast, I am going crazy." That is when the changes are just beginning to get to the point where they are useful.

I remember hearing that in New York in 1972 and 1973, when Pat Murphy was adopting one change after another on the basis of analysis and study, and the reality was that organization achieved peaks of vitality and energy and thrust that it never managed before and has not managed since, and I would debate anyone on that question.

So, too much, too fast is the key. I want to just conclude by speaking of the importance of the National Institute of Justice. They have few dollars, but have spent them carefully. They have brought about leverage. It is not my position to be a sycophant, as Chips Stewart is well aware that I speak my own mind and have not always said things that have pleased him or anyone else. He has spent the money well, given us a lot of leverage, holds state-of-the-art conferences on policing that stimulate a tremendous amount of intellectual excitement, and the importance of the intellectual climate cannot be overstressed. We need to take on the tough issues of inefficiency, anti-intellectualism, corruption, brutality, waste, stagnation. We need to challenge. We need the Socratic approach, the questioning approach. There is a tremendous

amount of anti-intellectualism in the police world. The police must learn to marry the enormous values of their experience - and it is enormously valuable - to the rigorous and tremendous importance of intellectuality and science and research and study, because that shapes an organizational value system - the Socratic approach and intellectualism.

My hope for the conference is that the attendees go home committed to the proposition that the pursuit of knowledge and the creation of an organizational climate that fosters it are the most useful initiatives the criminal justice executive can inspire. Thank you. God bless you. Go home and have a good time.

(Applause.)

Discovering What Works: Uses and Limits of Experiments

PROF. LEMPERT: Don't go home just yet. We have some very good speakers for you. Before I introduce the two speakers in the first substantive session, there are two announcement I have been asked to make. The first is that this is being recorded. All of us, and all of you are on candid microphone, and cassettes are being made of the various sessions. If you wish to purchase a cassette, there will be forms outside on the registration table so that you can do that.

The second is that in some of the sessions there will be time for comments or questions from the floor. If you wish to comment, if you wish to ask a question, be sure you take the microphone on the table, bring it over to you and speak directly into it. That way it will be recorded on the tape and, also, other people in the room can hear you.

The microphone is an open microphone. It is always on which means conversely if you wish to say something which you don't want distributed to everyone, please be sure you are not talking into the microphone.

I will introduce our next two speakers who are both going to discuss the uses and limits of experimentation now so that they can just follow each other. They are two very distinguished and very interesting speakers.

Our first speaker is Peter Rossi who is currently the S. A. Rice Professor of Sociology at the University of Massachusetts. Peter has taught at a number of the leading academic institutions in this countrym, Harvard University, University of Chicago, Johns Hopkins. While at University of Chicago, he was the Director of the NORC which is the organization that runs one of our regular national opinion polls. His distinction has been recognized in too many ways for me to list. I will just name one of the most important which is that he has served as the President of the Americal Sociological Association, selected by his peers as their leader. He has done research in a lot of areas related to criminal justice, some very important work on weapons and crime, and he was a researcher in the TARP experiment that is reproduced in your book. Currently as has been the case in his career, he is in the forefront of those researching pressing media problems, right now looking into the problems, dilemmas of homeless persons.

Peter will be followed at the speaker's podium by someone whom Tony Bouza has just said is the leading criminologist in this country, Lawrence Sherman. Lawrence is the author of many, many articles, books, research reports and the like, but he has established a name for himself

in criminal justice field experiments. He has designed and directed 13 field experiments. There are more field experiments than we put in our agenda book, but you see we did try to do a real search, and we came up with 30 we could write about, and Lawrence has been involved in 13 himself, which is an extraordinary number. He was the principal director of the Minneapolis domestic violence experiment of which you have heard already and will hear some more about. He is currently professor of criminology at the University of Maryland and President of the Crime Control Institute in Washington, and at the moment though he has the Maryland appointment, he is the Seth Boyden Visiting Distinguished Professor at Rutgers University. Lawrence will follow Peter to the podium.

(Applause.)

PROF. ROSSI: I now understand why Larry Sherman and I are designated to follow the previous speakers. It is to show us up. It is a very hard act to follow, and some of the topics that I am going to talk about overlap with the previous speakers' presentation.

I am going to talk about social experimentation, what are its essential features, why social experimentation of a formal sort is different from just fiddling around, and why one should undertake experiments. Experimentation of some kind, of both the informal and the formal variety has been around for a long time. Bob Boruch who is sitting over there can regale you to the point of utter boredom with accounts of experiments that he has culled from the Talmud, from the writings of Islamic scholars, medieval manuscript writers and the like, all of which describe activities like experimentation which have the characteristic of trying something new and then observing whether it works. Of course,

there is experimentation in the common language sense of fiddling around. My father, for example, used to call himself an experimenter simply because he used to fiddle with the dials on the shortwave radio trying to get distant places. "What are you doing, Dad?" "I am experimenting. I am trying to get as far as I can."

Now, however, laudatory or foolish these efforts at informal experiments may have been, they are not quite what we mean by formal experimentation, that we are talking about here today, although they do have some things in common. First, they both are motivated by the goal of improving current practice, fueled by accompanying dissatisfaction with those practices or perplexity, just wonderment about whether one is doing the right thing with existing practices. Secondly, both the formal and the informal experimenters are really expressing skepticism and doubt mixed at the same time with hope and optimism, skepticism and doubt about current practices and optimism and hope about improvement.

Experimentation is not the way of a true believer. You have to be a skeptic, at least an agnostic. (On a Friday, the 13th, that is a very good thing to be.) Nor is experimentation for the pessimist. If a person believes that life cannot be improved or that practices cannot be improved, then there is no point to experimentation. So, in other words, both formal and informal experimentation rest upon the faith that active intervention, that one should do something, or treatment can produce improvements. Finally, both formal and informal experimentation share in common the conviction that it is possible to assess whether or not one is achieving any improvement, whether one process is better than another.

The differences are many, however. From this point on, the ways of formal experimentation diverge sharply from those of the informal counterpart. The procedures of formal experimentation were developed in the early part of the century by biological and social scientists, who began to realize that biological and social systems were very tricky and very complicated.

Change, rather than stability was characteristic of such systems. Hence, the observation of some change after you tried something new - after you intervened with a new change - did not necesarily mean that the new practice was the cause of the observed change. The spontaneous changes which are characteristic of social systems and biological systems may produce false good news and, also, false bad news. As the medical aphorism goes, the common cold, if treated lasts two weeks and if untreated lasts 14 days. The fact that colds last 14 days and cure themselves spontaneously makes it very, very difficult to assess the efficacy or the effectiveness of the cold remedy.

Now, don't throw away your favorite cold rememdies. They may not cure the cold, but they make you feel better; the Hawthorne effect or the placebo effect is probably most important.

Social systems, whether they be police forces or court systems or ordinary communities or less organized communities particularly present a special problem when it comes to assessing whether a treatment or an intervention has worked.

Social systems are selective in character, selective in such ways that an intervention can be facilitated by that kind of selection or impeded. This principle is probably best illustrated in Governor Lester Maddox' well-known statement that he made in response to the criticism

that he wasn't doing anything about reforming the prisons of Georgia. He is reputed to have said, "We cannot reform the prisons of Georgia until we get a better class of prisoners." He is right. That is one way of improving the prisons of Georgia or anything else. It is one way of showing that an intervention works is to do it on the best possible case, especially on those cases that would show improvement in any event.

Getting a better class of prisoners into the Georgia prisons is one sure way to improve the prisons but it is, also, the irrelevant way. The real trick is to improve the Georgia prisons without changing the class of prisoners, by changing the organization of the prisons and by changing prison practices and working with ordinary workaday prisoners.

The essential features of formal experimentation consist of a set of procedures. There are ways of acting, ways of behaving designed to enable the experimenter to discriminate between the effects caused by an intervention or a treatment, discriminate them from those that are produced by all the other processes that are at work. These procedures differ in detail from experiment to experiment as experimental designs accommodate to the particular intervention that is being studied. I think Dick Berk will talk about some of the types of experimental design that one can use. The designs all have in common the notion that we are trying to isolate out the special effects of an intervention and discriminate those effects or the size of those effects from what would ordinarily have occurred.

Most important of all an experiment in a formal sense is concerned with establishing the conditions for making a division between the net effects of an intervention and the effects of other processes that could produce the same sorts of changes.

The net effect of an intervention is defined as the difference between what happened under an intervention as compared to what would have happened otherwise had the intervention not occurred.

Now, of course, this is difficult to accomplish because we can never know what would have happened if we hadn't acted in some particular way. Indeed, that is a problem that concerns us all endlessly when we go back over some embarrassing episode of our lives and have said, "Gee, if I only had done X or Y differently, things would would be different now." We cannot do that. We can ruminate over it, but we have to work out some other way of estimating what would happen if we had not intervened. What most of the experimental designs try to accomplish is to construct a circumstance in which there is no intervention, and by contrasting that no intervention circumstance with the intervention circumstance, we can make some kind of estimation of the effects of an intervention.

The clever experimentalists aided by statisticians have worked out a variety of ways to produce these control situations or the non-intervention situations. Although this is not the only solution, the classical solution that involves making the least number of assumptions is to create interventions and control the situation by randomly assigning whomever or whatever you are experimenting with into an experimental and control group, administering intervention to the experimental group and withholding the intervention from the controls.

As used in the vocabulary of experimentation, random selection means following quite specific procedures to ensure that every subject that is in the experiment has known and sometimes equal chance of being picked up in an experimental group or being in a control group.

This procedure is very contrary to the meaning of random in everyday speech. Random selection is not arbitrary. It is not capricious and not subject to whim or whimsy.

Indeed, random selection is highly structured and is fair, unbiased and usually an equal opportunity device. It is this fairness through the "equal opportunity device" that has made it possible for the courts to develop a rational for its use in uncertain circumstances.

What does randomization accomplish? First of all, it prevents the Governor Maddox solution to reform. If the experimental and the control groups are large enough neither of them will contain an excessive number of good or bad subjects. Each will tend to be composed of the same mix that was in the pool from which the subjects were drawn.

This means that random selection applied to persons who are arrested, for example, will produce about the same mix of offenses, ages, ethnic groups, etc., in an experimental and a control group or random selection applied to controls will produce a group of experimental controls and a group of controls that are alike in most discernible ways.

The second important feature of formal experimentation is an obsessive concern with the nature of the intervention being studied. In order to be sure that the net effects that are estimated are correct, one must, also, be sure that the intervention delivered is of high fidelity, that is to say a faithful reproduction of what was intended as an intervention. Specifically that means the following: First the intervention has to take place. Now, that sounds so obvious that you must wonder why I even mention it at all. I stress it because there are entirely too many instances in which an intended intervention did not occur at all, and not necessarily deliberately but by a failure of the

delivery system to produce the intervention in time or frequently enough or whatever it may be.

Second, the mode of delivery should not contradict the treatment. Perhaps the best example of such a contradiction is one of the experiments that is in the compilation by Michael Dennis in which the experiment is supposed to evaluate the utility of group therapy in helping prisoners when they are released to adjust to civilian life. The contradiction involved is the guards were used as the therapists which undercut whatever effect group therapy might have had. Finally, the treatment has to be delivered at an apropriate dosage. If treatment calls for holding a spouse abuser in jail overnight, then that doesn't mean that a 15-minute stay in jail will produce a proportionate effect. If it is not the appropriate dosage then the treatment is not being delivered with very high fidelity. The third important feature of formal experimentation is an obsession with measurement. The importance of measurement has to be stressed because careful recordkeeping and measurement often slips below acceptable levels in actual operation of experiments. Keeping track of what is going on can often appear irritating and burdensome to persons who are in charge of operations. Poor measurement is absolutely catastrophic to experimentation. You will never be able to discern effects, even if they exist, if they are not measured appropriately.

Successful intervention can have its effects so badly masked by poor recordkeeping and measurement that it may not be possible to discern its success at all. This obsession with the quality of measurement and the type of measurement is a burden which falls on persons in operations, as well as persons in research, and it is a burden which should be shared equally or at least to the extent that it is possible. No one should have

excessive measurements, that is to say measure things that you don't really need.

What I was trying to do here in this talk today was to induce in you some kind of enthusiasm for and commitment to the idea of experimentation, but that was done so nicely by Chip Stewart and Chief Bouza before me that I don't really have to do that, but now I can tell you why it will hurt a little bit.

Experimentation takes time and effort. Clearly it is not a plaything that can be fooled around with for short periods of time, nor should anyone experiment with interventions that do not make any sense at all, that is to say fiddling around - like my father experimenting with the shortwave radio - is not something that public officials should do. Fiddling is simply irresponsible behavior.

A good experiment is one that is a careful search for some new way of proceeding that will work better than the status quo in a careful, painstaking demonstration that the new ways work or do not work, and I wish you luck in whatever experiments you undertake.

(Applause.)

PROF. SHERMAN: The first prosecutor turned to me and said, "I don't care. I don't care if we are causing more crime against battered women by prosecuting their batterers. We don't need to know that. We probably shouldn't even have that information."

The second prosecutor said, "I disagree. I think it would be really good to know what effects our decisions have, but I don't want to be the prosecutor who has to do this experiment by charging people according to lottery. I wish some other prosecutor's office would do the experiment so that I could read about the results."

The third prosecutor said, "You are both wrong. Not only is it essential for us to know what effects our decisions are having, but it is our professional and ethical obligation to cooperate with experiments when we have the opportunity."

These comments were made while negotiating with a large prosecutor's office about the potential of doing an experiment on domestic violence. The comments illustrate three key points about field experiments and criminal justice.

The first point is that experiments can tell you what works, or at least tell you the effects of some decisions.

The second point is that experiments cannot necessarily make your policy for you, especially when you have many goals of your policy, and the experiment in question is only addressing one of those goals and not other goals, such as just desserts.

The third and perhaps most important point is that what you get out of experiments depends on what you are willing to put into them.

The premise of all these comments, and throughout the whole morning, in fact, is rather threatening, kind of insulting, if you think about it, to many people in the field. The premise is that we don't know what works. We could, in fact, be doing more harm than good with many of our decisions. We could be doing wonderfully with others, but we need experiments to tell the difference.

There are many examples of how our good intentions have been producing bad results. A famous juvenile delinquency experiment in the 1930's, the Cambridge-Somerville experiment, sort of a big brother program, took 650 poor boys from poor neighborhoods, randomly assigned one-half of them to an experimental group to receive an average of 5 years

of weekly counseling and one half to a control group which basically was left alone.

Over a 30-year follow-up, the results showed that the experimental group who had received the benefits of this counseling wound up committing more crimes, having more death at an early age, suffering more alcoholism, more mental illness and more stress-related diseases than the control group.

It appears that these boys would have been better off, if we had just left them alone. All that effort and good intentions only made things worse. But if we hadn't done the experiment, we wouldn't have known that.

The juvenile diversion experiment in Los Angeles in the early 1970's took a group of juvenile offenders, randomly assigned them to 1) be released completely from further processing, 2) referred to a community treatment program or 3) prosecuted as normal through the system. Though the self-reported acts of repeat crimes did not differ between the three groups, the official number of rearrests was highest for the group that was prosecuted as normal, second highest for the group referred to community treatment and lowest for the group that was left alone altogether.

Again, doing nothing for that group was better than doing something, but we wouldn't have known that if we had not done the experiment.

The drunk driving jail time experiment in Minneapolis (written up in your packet) took advantage of a policy change in the early 1980's when all the judges announced that they were going to use two days in jail as the sentence for all convicted drunk driver offenders. One judge

cooperated with that policy 75 percent of the time; call him the jail judge. The other judge cooperated 24 percent of the time; call him the no-jail judge. The cases were randomly assigned to the two of them at random by a court clerk. There were no differences in background or other characteristics of the cases going to those judges; a two-year follow-up showed that 20 percent of the jail judge cases had another arrest for drunk driving, but only 16 percent of the no-jail judge cases had a rearrest. It was not statistically sigificant, although it was borderline. Clearly we weren't getting any benefit from our $100 a night for putting those people in jail, and from the federal law in which we give more money for highways to states that mandate two days in jail -- at least not on the criterion of specific deterrence or reducing repeat crime by those who are punished.

But this example also illustrates the second point, which is that experiments cannot make your policy for you. At the same time that this jail time policy was not working, and perhaps backfiring, for those who were actually sentenced, there was a countywide reduction in the number of accidents related to drunk driving because of the general deterrent effect of announcing this policy with a lot of publicity.

So, here we had general deterrence without specific deterrence. Who is to say which is more important? An experiment is not going to answer that question. There are many goals, many factual questions and many considerations involved in making policy, even when we find that some of our policies are backfiring with respect to some of our goals.

The purpose of experiments is not to make policy but to inform policy, to provide more specific information about the consequences of policy that officials can use to make decisions in light of community

values. Unlike many studies that gather dust on the shelf, that information has, in fact, been very influential.

The positive examples include the Vera Institute of Justice in New York. In the early sixties Vera tested the new idea of release on recognizance, or personal recognizance as it is called in some places. ROR as we know it today around the U.S. and several different countries is directly attributable to the results of that experiment, which showed that people could be released without greatly increasing their absconding rates, without having bail, if they met certain eligibility criteria for community ties.

The Police Foundation's experiment in San Diego, showing the difference between one-officer patrols and two-officer patrols in terms of law officer safety and other characteristics was another influential success. It found that, in fact, the one-officer cars appeared to be somewhat safer because the officers took fewer chances. That experiment has had a lot of influence on big cities which previously had used only two-officer patrols, and many of them, including Minneapolis, then moved to one-officer patrols.

Our own Minneapolis domestic violence experiment, showing that arrest was more effective than two alternatives in that city at reducing repeat violence by the offenders arrested has in part contributed to an increase among cities over 100,000 from only 10 percent of them encouraging arrest in 1984, when the study was announced, to 44 percent of them encouraging arrest two years later.

Now, some of my academic colleagues are a little nervous when they see experiments having that kind of impact because they are rightfully and painfully aware of the limitations of social science in

discovering the truth, and the substantial chance that even an experiment can be wrong or at least not generalizable to every other city in the country. The question becomes "too much" influence from one experiment as compared to what, and the what is the status quo for which there is often no evidence.

Think of all the medical treatments that have been standard practice for years until somebody conducted a randomized experiment to test them and found that the practice did more harm than good. I am not talking about ancient history like bloodletting. I am talking, for example, about two years ago, an experiment testing an operation designed to prevent strokes was found through a randomized trial to do more harm than good and generally abandoned throughout medicine.

Thus, the key ethical issue to me in experimentation is not whether it is unfair to punish people according to a random formula. The key ethical issue to me is the ethics of ignorance, that is how we can justify <u>not</u> doing experiments when the alternative is virtually dealing with people blindfolded, not having any idea what effect we are having on them.

Now, there may be ethical flaws to randomization, but they can be no worse than the harm caused by untested policies or by policies that deter some people while making other people more criminal than they would have been, and that is something that experiments can also shed light on.

How many Thalidomides are lurking out there in the criminal justice system? We won't know until we test what we are doing. Not that experiments can test all of our methods, of course. There are many questions that are inappropriate for randomized experiments to answer. Other research designs, while limited, must be used, and in general,

experiments need large sample sizes. So, if you talk about large units like police beats or neighborhoods or entire cities, it is very hard to get enough cases for a randomized experiment.

For example, suppose you want to test the general deterrent effects of roadblocks on drunk driving, which is a key question that judges are addressing in deciding whether roadblocks are constitutional. You would ideally want to get 100 cities volunteering to be randomly assigned to do roadblocks or not. It is not very likely you are going to get that kind of commitment from that many cities, and in experimentation commitment is vital -- which brings me to my third ad final point: <u>what you get out of experiments depends on what you put into them</u>.

That is not just a grantee's bid for more money, although certainly putting money into experiments is very often essential. But more concretely, what you do with that money has got to include scientific rigor, especially in how the randomization is done. It has to include sound management of the organizational change process that an experiment entails, and it has to be testing strong and powerful treatments to be very interesting.

The appointment of Chips Stewart has provided a window of opportunity for experimentation. For the first time in the history of federal funding of crime control advances, we have an NIJ Director who views randomized experiments as the top priority method for the research program, but neither he nor NIJ can do it alone.

Improving crime control through field experiments requires three-way cooperation. First and most important, we need <u>practitioners</u> who are willing to conduct experiments who are willing to give up their

discretion, to take the risk of allowing a computer formula to decide how to dispose of the cases that are flowing into that decision point. We need researchers who are willing to make the academic sacrifices and take the career risks of leaving their offices and their secondary data sets to spend a lot of time out in the field negotiating, and perhaps failing, to implement and develop experiments. We need people from the funding bodies to insist in a very tough quality-control fashion that both the practitioners and the researchers do the best job possible.

Now, that is very different from the LEAA of the early seventies when we had "programs" and "evaluations" and where, quite frankly, the programs and the program people did whatever the hell they felt like after they got the grant award. The "evaluators" had to make do to patch up the research design and try to produce some kind of report, full of qualifiers about how "we cannot really conclude very much."

We have a lot less money these days. We can do a lot fewer projects, and our standards have to be a lot higher for the era of experimentation.

Each site needs to think of itself as a national laboratory in which the primary client is not the local policy maker but the people of the United States who are financing or at least leveraging a large part of the experiment.

The experimenting researchers, unlike evaluators, need to think of themselves as dedicated to the experimental design and not to the site. If the practitioners cannot provide sufficient scientific rigor in the implementation of the experiment, the experimenter should simply go elsewhere. They have no other leverage. It is only that option of simply going someplace else that can provide an inducement towards a rigorous design.

Now, that approach means a lot of pain and a lot of failed negotiations, a lot of frustration. But I think it beats wasting money and time on experiments that are doomed to break down. Read in your packet about the Denver judges who promised to do a drunk driving experiment, who accepted all the wining and dining that the principal investigators provided, and then just didn't do it.

Read in your packet about the Kansas City police officers not participating in the preventive patrol experiment which was making history by trying to withdraw patrol from some areas. These officers just wanted to have a little fun by driving through the control area and sabotaging the experiment, thereby helping to prove that patrol didn't deter crime (but I don't think that was their plan).

You won't read in your packet about the Birmingham police who took ample Police Foundation money to conduct an experiment to test the idea of decoys to fight robbery. Once they got the special unit set up, they decided that the best time to catch robbers is really 9 a.m. to 5 p.m., Monday through Friday; very nice hours to work, but I don't think you catch many robbers that way.

Weak treatments, watered down randomization, sabotage -- when practitioners allow these things to happen, they get out of an experiment what they have put into it. But the same can be said for researchers, because if experimentation is to be a true collaboration, everybody is responsible for the proper implementation of the design.

Both practitioners and researchers must recognize that experimentation is not just a study; it is not just a program evaluation; it is a major process of organizational change. No matter how temporary that change may be, it is still going to be intrusive and major.

The same problems should be anticipated with experiments that are anticipated with any other kind of organizational change, and that is a limit on experimentation, the limit of what you can do within the power of internal and external groups opposed to the experiment or in favor of the experiment, a limit of the resources available, a limit of the job tenure of the policy maker sponsoring the experiment. Steve Goldsmith, the prosecutor in Indianapolis, is fond of saying that the true test of the success of any experiment is whether the prosecutor gets re-elected. There is a lot to that.

Like any change, it is better to plan it with those people who must carry it out. NIJ staff, I think, increasingly realize that experiments that come in as grant proposals that are all neat and tidy, and everything is worked out and finalized, may not work out that way in the field.

Policy makers making decisions in headquarters might also expect that some of those decisions may not survive in the field. Perhaps the best process is to have centralized planning of the basic outlines, while the details are worked out in a mandatory retreat. Before every experiment, the people who are going to have to carry it out and the researcher should go off someplace where there are no phones and no interruptions and really thrash the thing out for two or three days. They can explore their own commitment and their own concerns about doing the experiment, and develop some ownership in the creation of the project.

That ownership, I think, is crucial. It must be constantly renewed throughout the life of the experiment with frequent meetings, discussions, attention, maintaining the Hawthorne effect that Tony Bouza told you about and which he insisted that we do in the domestic violence

experiment by having monthly meetings for which he bought the beer. That is why the experiment, I think, accomplished as much as it did.

If you don't put those kinds of elements of good management into an experiment, again, the experiment is less likely to produce valid results. The same is true for the randomization process: the more discretion practitioners want to retain in overriding the randomization process or in screening out cases, the less valid and the less interesting the results.

The same is true for the treatment selected. The more powerful the sanction, or more important the policies, the more valuable the results are going to be for improving crime control. In general we have seen the middle range of questions being tested: misdemeanor domestic violence (not felony), minor juvenile delinquency, and drunk driving. But we have established a precedent for the experimental method with those questions. Now, we can move on to even more serious questions, such as whether second-time robbers have more or less recidivism when they get 2 years in prison or 3 years in prison. A lot of judges and a lot of prosecutors have different theories on that question. It is the kind of question you can subject to experimentation.

Does locking up very young serious offenders reduce their recidivism more than letting them have several "free" robberies or burglaries just because they are 11 or 12 years old? Again, an experimentable question. There are hundreds of middle-range questions awaiting testing, both about existing practices and about policies that have yet to be invented or that are just being invented: ideas like electronic monitoring which NIJ is now testing, or urine surveillance, or the RECAP program in Minneapolis, focusing on 500 high-volume addresses

that get a good chunk of the crime in the city, one-half of which are being subjected to an experimental intervention and one-half of which are being left alone.

In cancer research, 500,000 chemical compounds have been tested for their effects in fighting cancer. Do you know how many of those worked? Forty. We have to expect failure, not in the doing of the experiment, but in the results. Out of all of the ideas we could possibly test as a way of fighting crime, perhaps the very small minority of them are going to make a difference, but it is only, in my view, through that Edisonian approach of trial and error that we are going to find better ways to control crime in this country.

I hope my discussion has not made the limits of experiments seem to be greater than the uses, because the uses are truly enormous. Hundreds of millions of people have been freed on ROR because of an experiment. Billions of dollars in patrol salaries have been saved through one-officer patrols. For better or for worse, hundreds of thousands of spouse abusers have been arrested, in part because of an experiment. If you accept the claim that experiments are essential for improving our ability to control crime, then the challenge is clear.

NIJ has given us a window of opportunity. Let us make the most of it.

(Applause.)

PROF. LEMPERT: In this morning's session, you have heard a lot about the excitement and the promise of experiments and what experiments have and can accomplish. You will hear more about the promise and excitement of experiments in the sessions that follow, but there are, also, complexities. There are difficulties. There are alternative

research strategies, and you will hear a lot about those in the sessions that follow today and tomorrow. We have two extended sessions in which you will be given an opportunity together to discuss ideas for experiments, how to work out experiments and sort of get into these complexities and difficulties, as well as the promise, firsthand.

For now we will take a break until ten-forty-five.

(Brief recess.)

Discovering What Works: Designing Experiments

DR. DIAMOND: This session is going to focus primarily on ethics and legal problems, but to lead that off we are going to have a talk from Dick Berk of the University of California at Santa Barbara. When I spoke to Dick about how he would like to be introduced, he gave a very specific instruction, and his instruction was that he is not the best criminologist that there is. He is not the best academic that there is. He is not the best practitioner, and he is not the best statistician, but he was the person that we could get. We needed somebody.

(Laughter.)

DR. DIAMOND: He may not be right about any of those responses. He certainly is one of the best statisticians we have doing research in this area, and he has wide experience doing field experiments in the area of law and justice, but I will let him tell you about them.

PROF. BERK: I guess I get to talk about the boring stuff. I am supposed to provide some common concepts and language about experimental design. The rationale, I suppose, is that if we are going to have these lively exchanges, which I agree are essential, we need some common framework from which to operate.

I am going to do this in a summary fashion. I certainly wouldn't advertise this as the complete statement, but I think it will give us some tools that will be useful over the next two days.

I am going to make four basic points. I will make them now, and perhaps then I should sit down. However, they asked me to talk for a little more than that. So, I guess I will have to elaborate.

The first point is that one cannot design experiments by recipe. While there are some common problems that pervade all efforts to do experiments, the solutions to those problems are almost always unique. Each site has its own wants, its own difficulties, and its own constraints that have to be taken into account. So, one cannot design by recipe. It is necessary to think. That turns out, many times, to be difficult, and sometimes even painful.

The second general point is that experimental designs inherently involve trade-offs. There are trade-offs because there are never sufficient resources, money, time to do what you would like to do, but even more fundamentally, the goals that you want to achieve are often in direct contradiction with one another. So, one should not feel frustrated by the process of designing imperfect experiments. There will always be compromises and trade-offs, and that shouldn't lead to frustration.

My third general point is that there will always be, and I need to stress, <u>always be</u>, a gap (and often a very large gap) between what one wants to accomplish in a research design and what one actually does. The delivery of an experiment, just like the delivery of a treatment <u>will</u> be imperfect, and these imperfections must not be ignored. Failures in implementation can sometimes be repaired in part through statistical procedures if data on the failures are collected. And if nothing else,

"truth in advertising" requires that failures in implementation be documented.

My fourth general point is that experiments affect our knowledge in two distinct ways: they may alter what we believe, and they may affect the confidence we have in what we believe. The second point is often overlooked: an experiment can change our minds or can confirm what we believe, but our uncertainty is reduced as well. At the same time, however, uncertainty will always remain. Some naively believe that if we only did a randomized experiment or even a package of randomized experiments, we would know for sure. That is simply not true. There will always be uncertainty. We hope through experiments to reduce that amount of uncertainty, but uncertainty is inevitable.

Those are my four basic points.

Now, lets start by talking about what an experiment is. I will be more brief than I had planned because I think that you have already been exposed today to the basic idea thoroughly and inspirationally.

The key issue I want to stress about an experiment is that the intervention is something we must be able to manipulate. So, for example, if we are interested in the effect of race or sex on earnings, race and sex are not manipulable and, therefore, not something that we can experiment with. Experimentation involves manipulation. There are a lot of important questions that cannot be studied through manipulation of some intervention and, therefore, not proper subjects for experimentation. I am not going to talk about such questions but simply alert you to the fact that there are lots of questions that experiments cannot answer. There remain, however, a very large number of vital questions that they can.

The second point about experiments is that they are, as people have suggested, a procedure or a set of actions to answer a particular kind of question: what works. In slightly more technical language, we are interested in estimating the causal impact of some intervention.

I had planned to talk a bit about what we mean by cause, and then discretion being the better part of valor, I backed off. You have heard already several examples. Pete Rossi gave you an instance. The basic notion is that we are interested in inferring what would have occurred under two conditions; what would have happened had the units, the experimental subjects, been exposed to the treatment compared to what would have happened had they not. The fundamental dilemma in experimental design is that we cannot directly make that comparison. To take an absurd case which makes the point, you cannot both execute someone and give them a life sentence.

In more realistic terms, we can think of lots of interventions we would like to try that cannot be simultaneously delivered, at least not without contaminating one another, and because of this, we cannot directly observe the causal effect of some intervention. We have to infer it, and all the apparatus that I am going to allude to here is basically in the business of causal inference. That is why we have these terrible debates amongst ourselves; we have to infer with imperfect devices under all the constraints I mentioned, about something we cannot directly observe.

Nevertheless, when we experiment properly, as Larry Sherman emphasized, we get a much better fix on that inference than if we fiddle around, to use Pete Rossi's phrase.

Okay, what kinds of things do experimental designs have to address? The first and obvious point is that they have to address measurement.

Let me give you a concrete example. There is a lot of concern in the State of California, and I suppose around the country, about the efficacy of prison classification systems. The idea that someone comes in the front door; we give him or her some formal and standardized diagnostic procedure; we attach a number, and then we assign security levels based on that number. That, by the way, is an advance, many people argue, over the way it was done in the past; someone would come in the front door, and a seat of the pants "clinical judgment" would be made by some intake officer who in turn would make that assignment.

We can debate the strengths and weaknesses of classification schemes, but the point is that we really know very little about their effectiveness. What would we have to do to figure that out? The first thing that we would have to consider is measurement. For starters, we have to measure what we mean by the treatment. Here, the treatment might be security level. Do we assign someone to minimum security, maximum security or something in between? That sounds like a trivial notion. Many experiments start out focusing on a treatment in this trivial kind of way. Likewise, what do we mean by an arrest? Does it mean just hauling someone downtown?

If you think about this for a moment, there is lots going on beneath the label "arrest," or beneath the label "maximum security," that needs to be unpacked.

So, when we measure treatments it is absolutely vital that we measure treatment content far better than we usually do. Commonly, this is ignored. Perhaps in the case of security level what really is going on is something like access to recreation or job training or educational services or drugs or anything else associated with the treatment of security level.

We, also, have to measure an outcome. That, again, sounds trivial. We at least have to figure out which inmates get in more or less trouble, but there are subtle problems here as well. There are lots of different ways that one could measure the amount of trouble somebody gets into in prison, whether they hurt themselves, whether they hurt the guards, on whether they are difficult in other ways. Also, there are lots of complicated processes involved because a series of actions might follow from the security level assigned.

Too often we take the easy way out. We take some official measure, whether or not, for example, the guards report rule infractions, and forget about a variety of other less convenient measures. So, my second point about measurement, measuring outcomes, is that there are lots of them. They happen in sequence, and the easy way out of measuring just one or two, is often insufficient.

A third thing we need to measure when we undertake experiments is attributes of the units, which might make the intervention more or less effective. Does it matter whether or not someone is a first offender? Does it matter whether they are young or old or black or white or a gang member or not? These are attributes of individuals important in determining why some classification schemes work more effectively than others.

A fourth consideration in measurement is implementation of the experiment itself. We have to determine whether or not people are being assigned to different security levels, for example, in the manner we believe they are. We have to figure out whether or not the measures we are using, of the outcome, for example, are effectively being made. In short, we have to worry about the integrity of the design itself. You have to measure what you are interested in studying, but you also have to measure your own research enterprise.

I, also, should mention, although it is a bit of a technical point, that when measures are designed, it is necessary to anticipate the kinds of statistical manipulations that will be applied and, also, think through what sorts of information is going to be useful for policy makers. That means doing the hard work of thinking ahead. Once the data are collected, it is difficult, often impossible, to correct oversights.

For example, we all know that arrests are imperfect measures of infractions, but if public policy is based on arrests, arrests had better be one of your outcome measures. The point is that there are lots of policy concerns which drive measurement, which are too often overlooked and need to be anticipated.

Let me now turn to causal inference. As I said earlier, and as other people have suggested, causal inference is just that. It is inferring from something that you can observe to a concept, in this instance a relationship between what we call a cause and what we call an effect.

Consider another example, to give you some sense of the difficulties of inferring cause properly. There is a program in San Francisco, which is not unique by any means, or even unusual. I just

happen to be familiar with it. Basically, community service is sometimes required as a condition of probation. That is, someone can be sentenced to straight probation or probation plus community service. Community service may involve a variety of things. For white collar types, tutoring underpriviledged children is one instance. For blue collar types, park maintenance is one instance. The point is that there are lots of different activities under the rubric of "community service".

One of the questions being asked both in San Francisco and around the country is whether this program works. Suppose you wanted to do some research to determine the impact, and let us say that for simplicity we are interested in recidivism, rearrest rates. Basically, we want to find out whether community service, in addition to probation, is effective in reducing recidivism; how do we determine that?

The superficial way of proceeding involves going to the records, determining which people were sentenced to probation versus probation, plus community service, then checking the rap sheets to find out who was later arrested, and then simply computing the difference in the average recidivism rates for the two groups.

All of you can immediately think of lots of reasons why that wouldn't be compelling. Clearly the judges and the probation officers take into account a variety of things when they decide who should get community service and who should not, and there is no reason to believe that the two groups are comparable before the intervention. What you take to be the treatment effect may be, in Lester Maddox' terms, a function of a better class of prisoner. People who get community service, for example, may be better risks for one reason or another; they may do better simply because they were less likely to get into trouble anyway. How could you respond to that concern?

The statisticians tell us that we can get lots of measures on the people in the experimental and the control group and adjust the results to take account of pre-existing differences. However, for such statistical adjustments to be effective, you have to know and accurately measure each and every factor related to who got assigned which treatment and to the outcome. For example, it may well be that people with higher education are more likely to get community service, and people with higher education are less likely to recidivate. Unless you can adjust for differences in average educational level between the experimentals and controls, you are going to be misled. So, statistical adjustments, while in principle perfectly feasible, are always suspect.

Are there fallback positions? Yes, there are. The fallback positions involve the researcher or the practitioner intervening in a way that simplifies the world, and permits them to distinguish between the impact of the intervention and the composition of the experimental and control groups.

There are two rather effective ways of doing this. One is to assign to experimental and control conditions by some known rule. Suppose we developed from a prediction instrument an assessment of the risk of recidivism for individuals who were going to get either probation or probation plus community service. We could then take all the people who were high risk and designate them as experimentals (or controls) and all the people who were low risk and designate them as controls (or experimentals).

I know this is going to be counterintuitive, but if I can place all the predicted high-risk people in one group and all the predicted low-risk people in another group, I can make a fair statistical comparison

between the two groups. This follows because I know exactly how the assignment occurred. I made up the rules. In other words, instead of all the hundreds of variables that you and I could think of that are related to both who got the treatment and related to the outcome, there is only one variable, and that one variable is risk of recidivism. Since I was the one who assigned on that basis, and since I have for each individual the measure of risk, I can usually make a very simple statistical adjustment to get unbiased estimates. That is a solution. It is a pretty good solution, although not usually the best solution. The flaws lead me to a third solution which is the most effective way of inferring cause: random assignment. There is no need for me to pursue the logic of that. I think you all understand it, and it has been covered amply by the speakers before me.

An important message I want to get across is that these three techniques -- statistical adjustment, what is called deterministic assignment, and randomization -- can be mixed and matched in response to some of the trade-offs I alluded to earlier, including ethical concerns, to make for a powerful design responsive to practical and ethical considerations. Each one of these three techniques may well have a role to play and they can be mixed and matched.

Let me give you one example now. There is an experiment under way in Los Angeles at Hollywood Presbyterian Hospital. The experiment involves providing crisis counseling to crime victims who come into the emergency room, the hypothesis being that if we counsel people, their post-hospital experience will be better; they will adjust better. To my knowledge, there is only one randomized experiment that has been done before on this question, and as you probably all know, the track record of

counseling and therapy is spotty at best when we have been able to look closely at it.

So, there is ample reason to be skeptical about whether or not crisis counseling for crime victims who come into the emergency room is a good (cost-effective) thing to do. In working with the folks at Hollywood Presbyterian Hospital, we ran into the problem that will be familiar to many of you. There were people who really believed that counseling is terrific, and questioned whether we had the right to withhold the treatment from members of the control group. For them, the ethical considerations seemed overwhelming. We eventually worked out a compromise by combining some of the tools I just mentioned.

Basically what we do is when a crime victim comes into the emergency room, we determine the level of need. I won't bother you with details of the instruments, but based on the kind and intensity of victimization, it is possible to sort people into high and low need (although we actually use three levels). Within these groups, we randomly assign to the experimental and control conditions, but with one trick: for the low-need group, 10 percent get the counseling, and 90 percent get the control condition. For the high-need group, 90 percent get the counseling, and 10 percent get in the control condition. In other words, we alter the probability of getting the "good" treatment by level of need, and it turns out that with very modest manipulations of the data (basically "controling" for the measured level of risk), you are back into a randomized experiment, while still responding to very real ethical concerns.

Now, these kinds of tricks can be employed in various ways. The point I am trying to make is that there are other procedures for mixing different kinds of research designs to respond directly to ethical concerns.

A third component of experimental design that I want to talk briefly about is the role of chance in randomization. When we do a randomized experiment, it's a bit like shuffling a deck of cards and dealing hands. In essence, a hand is dealt to the experimental group and another hand to the control group. While on the average, the hands dealt to the two groups will be alike, by the luck of the draw, we will sometimes deal hands that are quite different. Analogously, randomization does not guarantee that the experimental and the control group will be in the aggregate identical before the treatment is introduced. Randomization assures that averaging over many "hands" (experiments), the experimentals and controls will be, in the aggregate, identical, but in any particular experiment, aggregate comparability may not be found.

Clearly, non-comparable experimental and control groups will produce misleading results. In other words, we are back into the box that we tried to get out of with randomization. However, because we know how the shuffling works, we are able to assess the probability that any particular aggregate difference could have occurred. This allows us to take the rule of chance into account.

Now, chance has a lot to do with the way the data are analyzed to be sure, but there are design implications that are, also, very important.

For example, our ability to take account of chance depends upon having a sufficiently large sample. Think of it this way. It would be fairly easy to draw all cards from a single suit if each hand were only

two cards. It would be much harder if each hand were ten cards. With larger samples, idiosyncrasies tend to disappear.The point I am trying to make is one needs to take into account the role of chance when the experiment is designed, not just during data analysis.

The fourth point I want to briefly talk about is generalizability. Take a concrete example. There is an experiment under way in Alameda County, California, right now in which through legislative action a big chunk of money was given to the district attorney's office in Oakland for a crackdown on drugs. The crackdown, in this case, involves a more rapid processing of drug offenders.

Even if this program is effective, it is not clear what can be said about other locales. But there's more. Would similar programs work for other kinds of offenders such as robbers? Would a similar program work if the treatments were altered a bit?

There is, also, the question of generalizing to somewhat different outcomes such as sentence length or crime reduction, instead of just processing time. In short, generalizability is multidimensional.

There are several generalization strategies that may be employed, and this is not sometimes understood. The first is illustrated by what the National Institute of Justice is doing in response to the Minneapolis Spouse Abuse Experiment. Six replications have been funded in six different cities, using somewhat different treatments, and somewhat different outcomes. The object is to find out whether the deterrent effects found in Minneapolis pop up again. In short, replications are one route to generalization.

A second route to generalization is representative sampling. For many experiments this is difficult, but not impossible. One could imagine, for example, in a prison study taking a random sample of all incoming prisoners, doing an experiment on classification with them and then not having any trouble generalizing to the population of prisoners from which the sample was drawn. Unfortunately, most of the experimental work that has been done to date has not fully capitalized on the possibility of using, as an adjunct to replication, proper sampling.

Third, it is often possible within one's own data set to analyze the data in a manner which facilitates generalizations. In the case of Minneapolis, for example, one can determine whether the estimated deterrent effect was the same for first-time offenders and offenders who had experienced arrest before. If the effects are the same, we can generalize the treatment effect ignoring prior record.

To summarize, there are three strategies for generalizing. Two of the three are almost always ignored in experimentation. Very few replications are actually undertaken, but at least the case for replication is widely known. Sampling and proper statistical analysis are typically neglected, even in theory.

To conclude, I have tried to indicate that there are four dimensions of research design that have to be considered. We have to worry about measurement. We have to worry about causal inference. We have to worry about the role of chance. And, we have to worry about generalizability. In addition, I want to remind you about the points I mentioned at the beginning and emphasize two of those as I finish up. The first is that there is no perfect experiment. What experiments do is they alter our beliefs, and they alter our confidence in what we believe. There is no once-and-for-all answer.

The second general point I want to stress is that the baseline for an experiment is not perfect information but current knowledge. Larry Sherman made that point very eloquently. The question is not whether we know something fully but whether we know more after the experiment than we did before. With current knowledge as a baseline, it seems to me, lots of important experiments could be done that would be impossible if we used perfection as our baseline.

(Applause.)

Legal and Ethical Issues in Criminal Justice Experiments

DR. DIAMOND: Lurking in the background of every experimental plan are concerns about fairness and legality. After all, random assignment can be described as systematic arbitrariness, and arbitrariness, particularly when intentionally imposed by a legal system needs some explaining.

The first speaker for this session is Charles Wellford of the University of Maryland who has both conducted randomized experiments and thought and written a great deal about the ethics of experimentation.

PROF. WELLFORD: Truth in advertising prompts me to make a couple of preliminary remarks. One, I just hope to be able to get through my paper because I found this morning's opening session extremely disturbing on a personal basis. In about 3 or 4 weeks I will sit down with what I have now learned is the leading criminologist in the world to do salary negotiation. While I agree with that characterization, it will just make that process more difficult.

Second, I have not written very much, if anything on ethical issues, although I have spent about 4 years working at the University of Maryland on something called an Institutional Review Board which meets monthly to review research and attempts to grapple with human subjects protection issues. I have found this opportunity to try to put some of that practical work at the university onto paper and into a context instructive for me. Finally, I just want to thank the organizers and the committee for inviting me to participate in this important conference. It somehow makes me think that I am either ethical or legal, conditions which I have never associated with myself in the past.

Prior to the 1960's, there were no agreed-upon policies, procedures or standards to govern the protection of human subjects participating in experimental research. Individual researchers using their own values and whatever guidance might be offered by the principles underlying their disciplines or professions were expected to exercise due caution on behalf of individuals participating in research. This condition existed even though substantial abuses by researchers were documented and brought to the attention of various scientific communities. By far the most troubling of these abuses was the participation, particularly of medical doctors in the systematic maiming and execution of Jews during the Second World War, much of that occurring under the guise of scientific research. Nowhere near as individually or socially damaging but highly publicized in this country was the longitudinal research conducted in Alabama on the effects of venereal disease. In that research a large number of poor black men who had syphilis were tracked to better understand the long-term effects of the disease. Even when proven treatments for the disease were discovered and

made available throughout the medical profession it is contended by many although denied by some of the researchers, that treatment was withheld from these individuals in order to maintain the purity of the research.

These kinds of abuses prompted significant discussion amongst researchers in the fifties and sixties and in fact, were addressed in a general way in the Nuremberg Code, but as I said, no standards or guidelines were developed.

In response to this, the World Medical Association established the first agreed-upon standards in 1964, when it declared that in using new treatment, in which clinical research is being conducted, "if at all possible, consistent with patient psychology, the doctor should obtain the patient's freely-given consent after the patient has been given a full explanation" and that in clinical research with no therapeutic purpose, "nothing whatsoever may be done to a human being without his full consent after he has been fully informed." This declaration of the World Medical Association sparked considerable discussion and controversy. The controversy was directed at the fact that many felt it established such a high standard, such a high threshold for subject participation that research would be irreparably damaged.

In the years since the World Health declaration, particularly in the United States, even more elaborate procedures for protecting human subjects have been developed. In the course of developing these procedures, the moral and ethical principles that underlie human subject participation have been given considerable attention. For today's discussion, I would like to focus our attention on human subject protection as it has been developed primarily by the Department of Health and Human Services in response to portions of Public Law 93-348, the National Research Act.

This Act, the 1979 Report of the National Commission for the Protection of Human Subjects of Medical and Behavioral Research and regulations promulgated by HHS have prompted much debate and formed the foundation for the current efforts to protect human subjects. As was recently observed, these proclamations, laws, reports, regulations have in the last two decades significantly transformed medical and behavioral research. Today in all institutions engaged in such research, it is accepted practice that all research in which human subjects participate be reviewed and approved by a body established to pay particular attention to the issues concerning human subject protection.

In the remarks that follow I want to try to do three things. I would like to first describe the ethical principles that underlie current human subject protection rules and regulations; second, very briefly describe the current human subjects protection regulations as promulgated by the Federal Government and which apply to most of the research funded by federal agencies; and finally, third, indicate what I see, as the implications of these principles and regulations for the conduct of experimental research in criminal justice.

Let me first turn to the ethical principles. One section of Public Law 93-348 created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Commission was charged with the task of identifying the basic ethical principles that should guide the development of procedures to protect human subjects. In 1979, the Commission issued its report, usually referred to as the Belmont Report, in which certain ethical principles were recommended to guide human subject participation. The Belmont Report suggested that there are three basic principles that "are generally accepted in our cultural

tradition," which are particularly relevant to human subject protection. These are respect for persons, beneficence and justice.

The principle, respect for persons, is frequently embodied in the standard mentioned with regard to the World Medical Association that a subject's participation in research be conditioned on informed consent. Specifically the Commission noted that individuals should be treated as autonomous agents and that the wishes of autonomous agents should guide participation in research.

In order for an autonomous agent to function effectively, the principle assumes a rational and often highly intelligent person who is capable of understanding the research and its benefits, is made aware of any potential harms the subjects may experience during participation and who has the freedom to make a non-coerced choice about whether to participate. The principle of respect for persons suggests that under such conditions autonomous individuals should be able to make a decision that is in their best interests and that this decision should be respected.

Related to the concept of the autonomous person informed consent is how this principle operates when there is diminished autonomy. The Commission suggested that anytime an individual is less capable of self-determination than the idealized autonomous person, then special consideration must be given to the concept of informed consent. This is true in the Commission's judgment and in most of the regulations that follow by definition in the case of children. Also, mentioned prominently and importantly for our consideration by the Commission were prisoners and others whose autonomy was or was potentially diminished by actions of the state.

Under this principle, it is quite clear to me that experimentation in criminal justice must pay careful attention to the degree to which subjects, particularly when they are victims or offenders and not necessarily as stringently when they are employees of the criminal justice system, that have diminished autonomy. The powers of the system are so great and can be so subtly applied, and the discretion in the criminal justice system is so complete that in almost all instances it would appear that the presumption should be that subjects in criminal justice experimentation are operating under diminished autonomy and special consideration must be given to their protection. More on that later.

However, as the Commission noted in their analysis of prisoners, this principle is difficult to apply. On the one hand, simply because a person is a prisoner does not mean they should not have the ability to participate, the right to participate in research, if they choose. On the other hand, given the degree of discretion within correctional institutions and the absence of power that prisoners have, it may be that there can be no true condition of informed consent for prisoners.

The Commission in noting this pointed out that in making a decision as to how to apply the principle of respect for persons there should be no blanket exclusion of categories of individual (except for children), and that in the hard cases the decision will often be a matter of "balancing competing claims urged by the principle of respect itself."

The second major ethical principle identified by the Commission is that of beneficence. In explaining this principle, the Commission states that beneficence requires that we protect subjects from harm and that we, also, be concerned with the loss of the substantive benefits that

might be gained if the research is not conducted. Research is not ordinarily done for the benefit of the subjects participating in the research. The primary benefits of research will usually accrue to non-participants. Thus, there are two elements within the principle of beneficence -- do no harm, the basic maxim drawn from the Hippocratic Code, and maximize possible benefits while minimizing possible harms. This, again, sets up a balancing kind of decision that must be made in evaluating the ethical issues in any particular piece of research. Thus, the Commission establishes a general principle but notes that in its application the principle itself will require a balancing of possibly competing interests.

The third ethical principle is justice, and frankly, it is the one that I think is the least troublesome but is the one that criminal justice researchers tend to talk about the most. So, let me deal with it quickly.

In discussing this principle, the Commission raised the issue of who ought to receive the benefits of research and bear its burdens. The Commission urged that in considering human subject participation careful attention be made to the selection of research subjects in order to guarantee that subjects are not systematically recruited from certain classes of citizens, e.g., prisoners, the poor, recipients of public aid, etc., and/or are selected simply because they are available without the ability to avoid participation.

The Commission, also, argued that any benefits that derive from research be made available to the full range of citizens in society. While the latter clearly goes beyond the control of the individual researcher, the former has become a principle that has been translated

directly into standards for research. Attention must be given to the process of subject selection and specific justification given for any procedure that argues for limiting subject participation in some way, especially if that limitation appears to place the burden of participation heavily on those who would qualify as less than fully autonomous.

These three principles which are now accepted as the general criteria on which to make judgments about the nature of rules and regulations to evaluate particular research projects are clearly reflected in the regulations that have been promulgated by the Federal Government to guide federally-funded research. Let me briefly now review those regulations with you. They are included as Tab E in your workbook. You might want to look at that, particularly at Page 104 of Tab E as I go through this section of my remarks.

The regulations developed by HHS were called for by Public Law 93-348, as I noted above. That law required that the Secretary of Health and Human Services, "shall by regulation require that each entity which applies for a grant or contract under this Act for any project or program which involves the conduct of biomedical or behavioral research involving human subjects shall submit in or with its applications for such grant and contract assurances satisfactory to the Secretary that it has established in accordance with regulations which the Secretary shall proscribe a board to be known as an Institutional Review Board to review biomedical and behavioral research involving human subjects conducted or that are sponsored by such entity in order to protect the rights of the human subjects of each research." After much delay, these regulations were promulgated by HHS, as 45CFR Part 46. These regulations, as amended, now guide all research conducted for and by all Executive Branch agencies of

the Federal Government. There are some obvious exceptions to this, security-type agencies; Defense Department has some other exceptions, but basically the agencies you would be dealing with as potential funding sources are at least in theory and hopefully in practice guided by these regulations, also.

While some agencies in the Federal Government have modified or added special conditions to the regulations, generally any research funded by an Executive Branch agency stipulates that these regulations will apply. They would not apply, of course, to the Judicial Branch of the Federal Government which has established its own guidelines which I will mention in a minute.

There are two aspects to human subjects protection as outlined in the regulations, process and standards. First, a few brief words on process. As noted in the Act, the regulations require that each entity conducting research establish a board whose duty it is to review research being conducted by that entity and to certify that the research is in compliance with the regulations. These so-called "institutional review boards" have become an essential part of the scene in all major research settings. The composition of and procedures used by these boards must be approved in advance by HHS and are subject to periodic audits. In a recent review of research within major universities, it was determined, in fact, that these review boards have been established, are functioning and are making decisions consistent with the procedures approved by HHS.

The second aspect refers to the standards promulgated by HHS. The scope of the application of these regulations is greater than what we are considering at this meeting. HHS defines research as any "systematic investigation designed to develop or contribute to generalizable

knowledge." All research is subject to the standards. Furthermore the regulations establish definitions of an important concept for our consideration, the concept of minimal risk.

The regulations define minimal risk in the following way, "risks that are not greater considering probability and magnitude than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations." When a piece of research involves greater than minimal risk, the full range of regulations and conditions in the HHS regulations apply. If there is a piece of research that has less than minimal risk, that is less than the standard that I have just referred to, then we can make some modifications in the research protocol. Basically these regulations establish seven criteria to determine whether the research design pays careful enough attention to the rights of subjects. Those are listed on Page 104. They are actually underlined in the xerox copy. So, I won't go through them in any detail with you, but I would suggest you look at them and note that they directly translate into regulations, the principles that I have just noted above, that is the principles of fairness, the principles of respect and the principles of beneficence.

No review of the issues of experimentation and criminal justice would be complete without mention of the report of the Federal Judicial Center's Advisory Committee on Experimentation and the Law. A summary of that is included in your workbook at Tab F.

Published in 1981, this report summarizes the results of the deliberations of a special advisory committee established to give guidance to the federal judiciary in its consideration of research on innovations in the judiciary. Basically the report accepts the principles with one

major exception in the area of beneficence. In addition the Judicial Center's Report makes very explicit an issue that was touched upon slightly by Dick Berk.

The Judiciary Report correctly notes that random experimentation is among the most rigorous forms of research but, also, the form that can create the most risk for human subjects. Therefore, the report places special emphasis on the process of justifying randomized experiments. The report observes that before authorizing a randomized experiment, "experimentation must be the only practical method of adequately resolving uncertainties," and that the least intrusive experimental design should be selected and that in selecting from alternative designs the design that has the least ethical difficulties associated with it should be included.

Let me now turn to what I consider to be some of the more important implications of this review of the development of ethical principles and regulations to protect human subjects. First, let me simply repeat and emphasize the element drawn from the Report from Federal Judicial Center's advisory committee. In order to justify randomized experimental research on human subjects one must have substantial and compelling reason to believe that the chance in criminal justice practice of procedure that is being considered will have the desired effect. Furthermore, it should be clearly demonstrated that randomized experimentation represents the only means to effectively demonstrate the the form and relationship of the innovation. In order to reach that stage there must have already been a substantial program of research to place us in a position to anticipate the likely cost and benefits of the research both to the subject and the researcher, and also to determine the variables that should be manipulated in the experiment. Therefore, no

random experimentation exposing subjects to more than minimal risk should be undertaken until such a stage of knowledge is reached.

Second, any criminal justice agency anticipating participation in a randomized experiment should be sure that the entity conducting the research has carefully reviewed the implications of the research for human subjects using the HHS or substantially similar standards. If the research entity involved with the criminal justice agency is a university this should take the form of the use of the existing institutional review board as the mechanism for such a review and should include the application of HHS standards. If the agency itself is undertaking the research, then the agency must very carefully consider whether it can establish appropriate procedures to guarantee that effective review has taken place. In my judgement, in most cases it will not be able to do so and it would make far better sense for the criminal justice agency to seek outside review.

Third, in conformity with the ethical principle of respects for persons and the standards of informed consent the human subject review must pay special attention to whether the potential subjects are able to make a non-coerced decision to participate. If there is any indication that the subjects are in the sense discussed earlier non-autonomous then even more special care must be undertaken before deciding to engage in the research. This special care most usually will take the form of establishing as a part of the review panel an advocate for the non-autonomous class of individuals who will be able to review the research protocol and the actual conduct of research. Fourth, of course the agency and the human subject review committee must review the research design to make sure that the principle of justice is met, i.e., that

subject selection is fair, and that the application of benefits will be general. Ordinarily this should not prove to be a major problem in criminal justice experimentation.

Finally, and I think this is the most difficult task of all, the agency must assure that there is a careful consideration of the ratio of risk and benefits for those participating in the research. It is not enough in my judgement to simply observe that while there are some risks to which the subjects are exposed that these risks can be justified on the grounds of the general good that will be secured by the completion of the research. A central theme in the literature on human subjects and in our culture is that we cannot treat subjects only as ends to a general good, particularly in random experimentations and particularly where there is an expectation that our subjects are likely to be non-autonomous.

The risk/benefit ratio must be in terms of the subjects of the research, that is, they should either be at less than minimal risk or derive substantial benefits from participation. If those benefits to the subject do not outweigh the risks that they as individuals are exposed to though participation then the research should not be conducted. A lesser standard, a standard that would simply look to the general benefits for non-subjects and weigh those against the risks to subjects is in my opinion inconsistent with the Belmont Reports ethical standards, the HHS regulations and inconsistent with the kind of treatment that you or I would like to receive as human subjects.

While this may sound like a very stringent standard, I believe it is the only ethically sound principle for us to use in determining whether research involving more than minimal risk should be conducted. Nothing in the literature suggests that such a standard cannot be met, experience

suggests it should be. For example, I recently participated on an adhoc committee to review the replication of the domestic violence research Larry Sherman will be conducting in Milwaukee. One of our concerns was whether the subjects of this research were benefitting in any demonstrative way because clearly they were being placed at some risk. Professor Sherman in response to this had already sought to make sure that every subject was in a better condition than they would have been if the research had not been conducted, i.e., each subject was to be made aware what services were available in the community. Each subject was to be made more fully informed than they had been in the past about the alternatives that they had. In our judgement, without the research the subjects would not have had this information -- would not have had these benefits. Therefore given the minimal risks that they encountered from the research we were convinced as a reviewing committee that the research met the standards of respect, justice and beneficence.

Only by beginning our review of proposed research with an expectation that every subject should benefit not just from the results of the research but from their participation are we in a position to determine the relationship risks to which they were exposed and the benefits they derived from participation. Thank you very much.

(Applause.)

DR. DIAMOND: We are fortunate to have three discussants on this issue who have all either participated in or are presently participating in randomized field experiments and who as a result, have had to deal with both the legal and ethical issues involved in implementing these programs. The first discussant is Judge Bruce Beaudin of the Superior Court of the District of Columbia.

JUDGE BEAUDIN: Thank you, Shari.

Good morning. A couple of preliminary remarks, the first is I want to thank my boss, Chief Judge Ugast for letting me be here and agreeing in advance not to hold me to anything I say to the extent that he will assign me to traffic court for the rest of my term.

The second thing is not so much preliminary as it is to say that there are three events in my life that have very much affected me, both in body and soul. The first was in 1957, when, with a group of finalists in the Curt Gowdy Junior Sportscaster contest, I sat in the dugout of the Boston Red Sox with my idols, Ted Williams, Johnny Pesky, and Bobby Dorr and got to see these otherwise unreachable and untouchable folks. The second was in October 1984, when I raised my hand and took an oath to dispense justice, equal justice, without regard to rich or poor, and joined a bench that was composed of men and women that I had worked with and admired for many years, over 15 or 20 years in the District of Columbia. And believe it or not, the third is March 1987, when I was told that I would become a member of this group. I feel the same honest-to-God awe at being with the people that I have already heard from and will hear from who have taken on the issues that are discussed here and are willing to lay their souls, in effect, on the line for what they believe in, and it is a great company to be in.

In considering the ethics and legality of experimentation, I would divide my points into three areas. There is a role that a judge has, and I would like to talk about that for a minute. There is another role though, that a Chief Judge has. I won't touch that except to say that a Chief Judge in supervising what is right, what is just in any environment, particularly an environment that he or she is responsible

for, has to be satisfied at a different level that experimentation of whatever kind is notgoing to hurt anyone. I think Norval Morris likes to say that the two professions that are the best are medicine and astrology. In medicine, at least you can pretend to do well, and with astrology you know damn well you haven't hurt anybody.

I remember Norval saying that about 20 years ago, and I don't know who he was quoting, but it made a lot of sense. I won't talk about that issue of the Chief Judge. I think, though, that as any criminal justice agency that begins to deal with an experiment of any kind makes a mistake if the Chief Judge of a particular jurisdiction is not consulted.

With regard to the role of the judge who is assigned the case we encounter a different problem. Should the case concern someone who happens to fall into one of the random assignment categories then the judge must consider whether a deprivation of some service exists and what that judge has to do about it. I would like to put that aside for a minute because I have not had enough to do with that yet.

It is a concern that those who are in experimental programs and are randomly assigned to different categories may be deprived of a service. Not as a judge but as an observer, I would say, "What about the two or three who get a service they would not otherwise have gotten?" And I do not think that the final answer is that beneficence assumes that those who participate in the experiment will not not themselves be beneficiaries of services that may exist within that framework.

The role that I think I am most comfortable about discussing is the role of the "proposer" who gets the experiment under way. As to the legal and ethical concerns, well, as the program design begins, if there is not a full commitment, as a manager, to the success of the conduct of

the experiment, (let alone the outcomes of the experiment) and the time invested with the people that design the research and the people with whom you will work is not intense at the front end, then I see nothing but trouble for that experiment.

Why do experiments work in some places and not others? Once the notion of the experiment is conceived, the packaging and salesmanship of that experiment is probably as critical a piece of whether the experiment will have a chance as is the work that actually goes into the experiment. For example, if someone decides that it would be a fine thing to test every criminal accused that is processed through a court system for drug use, how does one convince a prosecutor whose sworn duty it is to protect the safety of the community that, in advance, he or she should know that one out of three of the people identified as drug users will receive no services, will not even receive surveillance? That prosecutor can see that a potential threat to the community can't be taken off the street even though he or she continues to use narcotics because of the random assignment. How does one convince a chief judge that his judges will have to deal with complaints by civil liberties union lawyers or public defender lawyers that a defendant similarly situated has been denied a service, and what is to be done about it?

How does a pre-trial services agency or any other agency implementing such a program promise the subject of that experiment that he or she will receive no detrimental consequence from his or her participation in the experiment?

What hammers can you hold, ethically, to ensure that your subjects cooperate?

In designing the program that Chips spoke about earlier, we had to think of those things at the front end and say, "Can we globally decide each of those issues?" and the answer was "no." We cannot globally decide those issues. We can anticipate them. We cannot decide them. So, what do we do? How do we say, "No harm to the status quo, no threat to a decision maker or policy maker"? My answer to the prosecutor who said, "I cannot agree to have you conduct this experiment and randomly assign people to a control group that gets no testing and surveillance and one that has services and one that has intensive services, I cannot agree to that, because what if an armed robber falls in the random assignment where nothing is done?" was, "What have you got now? You cannot do anything with anybody now. How would you like to have one out of three?" That was enough. Inertia being what it is and recognizing how inertia works in the criminal justice system, my next proposal to the prosecutor was that any individual case that he wanted to take out of the experiment we would remove. In eight months, not one reclassification was requested.

To the judge who said, "Just a minute, I want this particular defendant despite whatever program or experiment you may have going, to be placed in a program of my choosing," my answer was, "Fine, Judge, we will do that. Anytime you want to do that, we will take that case out of our experiment, and then deal with your needs and concerns." That was fine because I had 51 judges and knew damn well that 51 of them were not going to follow their cases that closely, in fact, could not follow the cases that closely. Only a few asked that exceptions be made.

So, as a practitioner, as a proposer, as an experiment leader, I had to make decisions that I was not real happy with, when I compared them with that oath I took in October 1984, and yet, I never had any doubt in

my mind that the experiment was going to be fair and that any of the concerns that anybody had could be answered on an individual basis without destroying that experiment.

The last piece of the role of a proposer that I think needs to be mentioned, and it is probably well understood by everyone in this room, is that the level of acceptance of the leader of the experiment has a great deal to do with what those agencies that have to deal with this experiment are willing to take, and there is no doubt in my mind that the level of credibility and acceptance of the proposer of the experiment will have a great deal of effect on the acceptance of the system in trying whatever that experiment is.

There are about, I think, 31 or 32 other things that I wanted to say. My father said to me when I was a kid, "When you have to say something, stand up so that you can be seen, speak up so that you can be heard, and shut up, so you will be appreciated."

I close this particular part of my discussion by saying that I would appreciate it if those researchers who design experiments would do something a little better than present me with the conclusion that Michael Jackson and Bill Buckner are similar, in that each wears a single glove for no apparent reason.

(Laughter and applause.)

DR. DIAMOND: Our next speaker is Del Elliott of the University of Colorado who from the academic side has done experiments in the criminal justice system and of course, has had to respond to ethical and legal concerns.

PROF. ELLIOTT: I would like to talk about some of the practical problems involved in doing experiments, specifically being caught between the requirements and legal concerns of the agencies that are involved in the experiment and faculty members concerns about having their research program passed by a human subjects review committee which often very zealously pursues the formal HHS restrictions and the procedures (which have been laid out for you earlier by Charles Wellford).

Let me state from the beginning, however, that I am 100 percent committed to the spirit of the human subjects guidelines and am committed to compliance with those procedures. I think that is healthy. I think it is necessary, and in my own personal experience it does not seriously jeopardize the conduct of the research. It does frequently raise some very interesting questions and debates within the university community over what is appropriate and what is not appropriate.

I think it is important to note that university requirements typically are such that even though you are doing research which is not funded by HHS, that the university requires that you comply with these regulations because HHS has said to the universities, "If all research on your campus does not comply with these guidelines, then we will withhold funds from those projects which are funded by HHS." That is, they essentially say that you will apply these guidelines to <u>all</u> research on that campus or you will not receive any funds from HHS. This is an effective threat and as a result the human subjects committees on most campuses look at all research on that campus, whether it is funded or unfunded research, as long as it is sponsored by that university. So, those guidelines and procedures are very broad, if you are in a university context.

I guess I should also report to you that we have done four or five experiments, experiments which did involve some tricky legal issues, and to this point in time, we have not been involved in any legal suits or any challenges to random assignment procedures which have been used in all of those studies.

The fact of the matter is that under some circumstances, it is very easy to argue for random assignment because it provides a means of introducing equity into circumstances where there are limited resources for some treatment or some disposition.

We encountered a situation in Oakland a number of years ago in which they had a work study program which funded 200 summer jobs for those who met poverty guidelines in that city. For 3 years running, they had over 800 applicants for those positions, and had engaged in some kind of decision-making process trying to select which applicants would be awarded those positions. They encountered a great deal of hostility in the community over their decision making criteria.

We wanted to do an experiment evaluating the effectiveness of the work study program and suggested to the city that if, in fact, we were to randomly select from those applicants who were eligible for these work positions those who would get the positions, this procedure would provide a way of achieving equity in the distribution of those positions, and at the same time facilitate an experimental design.

We, in fact, did that experiment in Oakland, and it was the first year in which they had no complaints from the various interest groups in the community about how those awards were distributed. So, there are some circumstances in which what we call random assignment is, in fact, a means

for achieving equity in the distribution of scarce resources, and if you are in that kind of a situation, that is a ready-made argument for random assignment.

In our attempt to replicate the Minneapolis domestic violence experiment in Omaha, we encountered a number of human subject issues, and I thought if I ran through some of those for you, it would give you some idea of the practical kinds of concerns involved in conducting an experiment. The first concern that was expressed by the city of Omaha had to do with a concern over false arrests. As a matter of fact, they required us to explore the possibility of purchasing false arrest insurance for the city of Omaha and for the police department as a condition for doing the research in that community. If you have ever looked into false arrest insurance, it is very expensive. After some negotiation, both with the city attorney's office there and with our own legal counsel, it was determined that if we were very careful to specify that in all cases there had to be evidence of probable cause, that there was, in fact, no need to provide false arrest insurance. Still this was a major concern on the part of the city of Omaha with respect to the introduction of this experiment.

On the other hand, at the university, the human subjects committee never raised this issue. It is not surprising, I guess, that their concern had to do with the denial of legal protection to victims who might otherwise be protected by this legal sanction (i.e., arrest), as a result of the random assignment.

We were put in a very difficult position there because the Human Subjects Committee at the University of Colorado assumed the outcome we were trying to investigate, i.e , they assumped that arrest <u>was</u> an

effective deterrent. Since this was a replication of the Minneapolis study, there are, of course, some reasonable grounds for making that assumption. However, in trying to do a replication of that study, we had to start off with the position that we did not yet know whether arrest constituted a deterrent against subsequent violence. The Human Subjects Committee said, "We know that arrest is a deterrent, and you have to justify for us that you are not reducing the probability of arrest and thereby putting subjects at increased risk." In the City of Omaha those people who were to be eligible for this study were those involved in misdemeanor violence cases. In the past these cases had not been subject to arrest. As a result, the net effect of the study was to increase the probability of arrest in that community by one-third for those who were in the study. On the basis of this arguement, we squeaked by the Human Subjects Committee asserting that we would not reduce the overall probability of arrest; we would, in fact, increase it. If, in fact, the arrests were thought to have the opposite effect, that is to increase the risk of subsequent violence, we would have had difficulty in obtaining human subject clearance.

Another legal issue which we have faced a number of times concerns the fact that one of the kinds of information that we typically want to examine involves self-reported involvement in criminal behavior which if released or subpoenaed by the court, could provide incriminating information against that particular subject.

Now, when you are working in the justice system, and attempting to assertain the true level of involvement in criminal behavior as a part of the experiment you must be able to provide protection to the respondent as a basis for obtaining that information. There is currently available

through the government a procedure which is called a Certificate of Confidentiality which is provided for in federal legislation. Both the current drug legislation and the delinquency legislation allows the investigator to obtain a certificate in which the respondents in a funded research study are guaranteed protection of subpoena from any state, federal or local court. Most researchers who are involved in this kind of research in the justice system would be well advised to obtain that protection for their respondents. Otherwise you are in a difficult situation in which your data may be subpoenaed and involved in some kind of litigation involving one of your respondents.

We recently had such an experience, and it was a very difficult one for us. In a National Youth Survey we have been following a group of adolescents since they were 11 years old. They are now adults. I had a call from a mother just recently whose daughter had confessed to her that she had been sexually molested by her father on repeated instances. Two months after that event, the daughter was killed in an automobile accident. The mother subsequently called us to see if the respondent had reported these events to us, and if so, whether we could give her or her attorney information we had obtained in interviews.

The protection of confidentiality certification we have precluded our being able to give that mother any information with respect to the sexual molestation charge. We sought counsel from the university attorney and were told that that guaranty did require that we not give any information to those persons.

We got into another difficult issue in the Omaha domestic violence study associated with the possibility of an interviewer's encountering information from the respondent about child abuse. Because

there are special regulations which relate to reporting situations of child abuse, the human subjects committee said, "What happens if when you are interviewing a victim of domestic violence, she says that she intervened in a particular fight to protect her children or because there was some victimization of her children going on in that home? You are then subject to federal and state guidelines to report that information, which puts you in violation of the confidentiality guarantees which you give to the subject." We had to go through a very lengthy investigation as to how we could, in fact, deal with this problem, i.e., to protect the confidentiality of the interview and at the same time be required to report in the event of a child abuse situation in that home.

We were able to work that out. It was a very difficult situation which required extensive training to our interviewers about what to do under certain circumstances, e.g., if they witnessed an abuse situation or if the respondent divulged an abuse situation to them in the interview. I don't have time to describe that procedure, but there was one that was worked out satisfactorily to the human subjects committee.

Finally I would like to note that the data collection procedure itself may be an intrusive element in the experiment. We have encountered a number of situations in which in the process of interviewing victims of domestic violence, the interview itself raises some very emotional responses, i.e., recalls very traumatic and difficult events for the subject. The data collection process itself sometimes precipitates a great deal of trauma for subjects. Fortunately, that is not a frequent occurrence in our experience, but it is one which happens. I think our commitment to human subjects issues requires under those circumstances that we be prepared to assist respondents under those circumstances and at

the same time avoid contamination of the study. We provide our interviewers with referral information for subjects who are experiencing trauma as a direct result of the interviewing procedures, train them in how and when to use these materials.

Those involved in experimetns in the justice system should expect to encounter issues that raise ethical and legal questions, and in many respects, these questions will be unique and have to be dealt with on an individual basis. Dick Berk noted earlier that in many cases these are very, very unique situations, unique to the community and unique to the particular kind of intervention or treatment that you are investigating.

So, addressing these issues takes a lot of ingenuity on the part of the researcher in combination with institutional review boards and the agencies who are sponsoring experimental interventions. But the problems are not insurmountable and rarely constitute the primary obstical to experimental research.

(Applause.)

DR. DIAMOND: This session is going to go a little bit longer in order to give you in the audience an opportunity to ask some questions. The next speaker is going to be Mike McCann who is district attorney in Milwaukee, and he will address some of the legal issues that we so far have not given a great deal of attention to. Then we will take some questions from the audience, and Dick Berk will close.

MR. MCCANN: It is very much a pleasure to be here, and one appreciates that there must be some academics involved in the arrangements. You will see that Judge Beaudin is seated at the floor level and I on the platform. That is just the reverse of what usually prevails.

(Laughter.)

MR. MCCANN: The effort that is under way here to stimulate random experiments in the criminal justice system is vitally important. None can doubt the critical need within the criminal justice system to insure that our precious limited resources are allocated in the most rational and meaningful way. Everyone is familiar with the jail overcrowding situation throughout the United States. Are we using our space in the wisest fashion? Who can safely be released and who not released? What will random experiments teach us in that area? Prisons also are overcrowded throughout the United States. There are in excess of 30 state prison systems laboring under federal court orders capping inmate populations. What experiments can be conducted to identify the most rational and reasonable sentencing decisions so that the precious resource of prison space is used in the most prudent fashion?

District attorneys are burdened with thousands of cases inundating their offices. Which cases should receive attention, and which may be safely diverted? Judges are quantitatively the most precious resource. There are thousands of policemen in the typical metropolitan center, hundreds of assistant district attorneys and only a dozen criminal judges. Can randomized experiments with diversion and disposition by administrative examiners point the way to better utilization of precious judicial time? Police costs stress municipal budgets. Several years ago in our community homeowners pushed the common council to adopt an ordinance that subjected to money forfeitures persons who didn't appropriately clean up after their defecating dogs. Can you imagine a policeman staking out a private lawn to catch some dog that might be defecating there. It was obvious to all of us in the local criminal

justice system that it would be irrational for the police chief to dedicate any resources to enforcing that ordinance. Most questions of police resource allocation, however, are much, much tougher.

How much is committed to burglary containment? How much is committed to addressing the robbery problem? How many stakeouts should there be and so on. These are difficult questions which research to date does not adequately answer and that randomized experiments hopefully will. The criminal justice system is an adversaries system. On one side is the prosecutor and on the other the defense attorney. You must anticipate that whenever you enter into randomized experiments, there will be challenges. The Fourteenth Amendment guarantees equal protection of law. Inevitably implicit within the concept of randomize experimentation is the reality that similarly situated persons will be treated differently. If the defense attorney perceives that this will inure to the legal detriment of his client or that his client is not treated as favorably as another defendant, the experiment may be subjected to constitutional scrutiny on the basis of denial of equal protection of law.

Prosecutors typically are presented on television as somewhat hamfisted individuals, insensitive to civil rights and the last persons to appreciate that the accused may be innocent. In fact, most district attorneys are considerate of individual rights and are deeply concerned about justice. They will be concerned about the legality and ethics of any randomized experiment. In dealing with district attorneys, you will need to explain what a randomized experiment is because many of us are schooled in liberal arts and law and are not well informed in scientific studies. You cannot assume that the district attorney will appreciate the differences between randomized experiments and other experiments or tests

of programs. In our office, when we considered becoming involved in a randomized prosecution experiment, the reservations that were initially raised were ethical reservations, not so much legal reservations.

The district attorney often makes his policy decisions not on the basis of careful study of what the past shows because there have been few such studies. The district attorney to some extent, within his broad discretion, relies on intuition and not infrequently anecdotal experience. The keen hope that sound randomized experiments will provide a pool of knowledge to better inform district attorneys and other criminal justice professionals in their decisions lends substantial ethical justification for the conduct of such experiments.

The issue of informed consent will arise in many criminal justice randomized experiments. Let us consider a randomized selection among spousal battery cases with some alleged offenders being prosecuted and some alternatively referred to counseling. Ought the victims in each case be informed and provided an opportunity to object? Should the defendants be advised and offered an opportunity to opt out in favor of regular case processing? Clearly the interests of both the defendant and victim may be seen in different lights.

There is a downside risk to the prosecutor who on a randomized draw doesn't prosecute someone. Assume that person returns to the streets and commits a heinous offense that is widely publicized. Further, assume it's election time and one of the opposing candidates picks up that issue and says that the offender should have been prosecuted, and the darn fool incumbent district attorney relied on a computer to decide who or who would not be prosecuted.

A key ethical justification noted by Professor Sherman is the fact of existing, prevalent ignorance. The reality is that many of us in the criminal justice system do not know the impacts of our decisions.

I anticipate that as you approach professionals in the criminal justice system proposing randomized experiments they will raise various legal issues. There have been a number of cases, although the volume is not high, which address the issue of equal protection vis-a-vis various governmental experimental programs.

Within the human services field, the cases are rather consistent. Typical of those cases is Aguayo versus Richardson 473 Federal Second 1109, a 1973 case. The case involved what we would call today a workfare program. In designated areas within the State of New York certain AFDC recipients, if ablebodied, were required to report at a work site. If they failed to report, they were to be suspended for 30 days from the AFDC program. The experiment involving only certain parts of the state did not effect AFDC recipients in other parts of the state. A legal challenge was raised on the basis that this was violative of the equal protection requirement of the United States Constitution. The court rejected that argument citing with approval the rationale behind the experiment.

> "A purpose to determine whether and how improvements can be made in the welfare system is as legitimate or appropriate as anything can be. This purpose is suitably furthered by controlled experiment, a method long used in medical science which has its application in the social sciences, as well."

Quoting another case, the court commented that to stay experimentation in things social and economic is a grave responsibility. It is important to point out that the court in Aguayo held that this right to receive welfare benefits without the imposition of reporting to a job site was not a fundamental right and as a result would not be subjected to the same strict equal protection scrutiny that certain other fundamental rights would receive. A number of other cases in the human services field reached similar conclusions.

United States versus Thompson, an 1971 federal district court case, at 452 Fed. Second 1339, involves the criminal justice field. Two federal statutes were under equal protection challenge. First, the Bail Reform Act, passed in 1966, had liberalized the criteria for pretrial release of defendants in all federal courts. Second, the Court Reform act of 1970 applying only to the District of Columbia, imposed more stringent requirements on the pretrial release of accused persons. The prosecution sought to have the defendant who was arrested in the District of Columbia held under the more stringent Court Reform Act on the grounds that the act was a proper experiment to determine if the later act served society better than the earlier act. The court struck down the Court Reform Act holding that the Eighth Amendment right to bail was fundamental and that defendants in the District of Columbia could not be used as "human guinea pigs - particularly when basic human rights are involves." The Thompson case sounds a warning concerning experiments in which the most fundamental rights are involved.

On the other hand, in Chandler versus Florida, a United States Supreme Court case reported in 1981 at 449 U.S. 560, the court upheld the right of Florida to experiment with the use of television in the criminal

courts. In People versus Superior Court for County of Monterey, a California appellate court case decided in 1972 at 29 Cal. App. 3rd 397, the court upheld the right of California to vary among counties in its manner of treating drunkards holding that "a state must be allowed to experiment within its borders to determine what is the best way to deal with the problem in inebriates." The appellate court noted, however, that fundamental rights were not at issue.

None of the above cases involved strictly randomized local experiments. They are helpful, however, in highlighting problems that should be anticipated.

The Hippocratic Oath, "First, do no injury," has been quoted. Harm is difficult to avoid in the criminal justice system, whether you charge a person or not. The Belmont Report has been cited for three issues, first, respect the persons: second, beneficence; and third, justice.

This echoes the prophet Micah in the Old Testament urging man first, to act justly, secondly, to love tenderly, and finally, to walk humbly before your God. We have to walk humbly, as we undertake randomized experiments, because we are toying, in a sense, with the lives of human beings. The need for information is critical and obvious, but as we experiment we must take great care to respect the individual liberties of our citizens and to protect them against the danger of improvident decisions resulting in serious injury or even death.

(Applause.)

DR. DIAMOND: You have all been very patient, and I am sure that you have questions for either Professor Wellford or the other panelists, and we will take them now before Professor Berk gets the last moments in.

They are all hungry. Aha. Yes?

DR. AUSTIN: I am Joe Austin, with the National Council on Crime and Delinquency. Are there any studies that we know of that have been successfully sued, and the researchers lost in terms of random assignments?

DR. DIAMOND: I think that was probably a question for you, Mike.

MR. MCCANN: I am not aware of any such suits, but you have to recognize that these random experiments have not been widely undertaken within the criminal justice system. You are talking about third party potential liability, not simply the victim suing the defendant. Within the victimization movement there has been an expansion of third party lawsuits. District attorneys under a case called Imbler versus Pachtman have a very broad immunity from civil damages liability even for unconscionable decisions. I think in a good faith experiment that generally, the district attorney, using his discretion, would be immunized from civil litigation.

The police are not quite as adequately protected. I know when we considered an experiment in Milwaukee, our chief of police secured an opinion from the city attorney. I think you might find police more liable to damage judgements. Judges are virtually totally immunized for their decisions. The experimenter may have less protection than any of the others depending on the degree of active involvement.

PROF. BERK: Just as a footnote to that, it is possible for researchers, academics and others to get what is called malpractice insurance for all kinds of terrible things that can happen to you, including suits of this kind.

DR. AUSTIN: Does NIJ pay for that?

(Laughter.)

DR. DIAMOND: You have to put it in your grant proposal. Bob?

DR. BORUCH: Bob Boruch, Northwestern. In partial answer to the last question, there is a very fine article by an attorney-scholar. His name is Marshall Breger who is actually currently with the White House, perhaps not for long, on the idea of randomization in the law and especially with respect to constitutional law. One of the cases he mentions, it might be California Welfare Rights versus Richardson or another Atlanta case, let me see, it was another health insurance case in which the idea of a copayment scheme for health insurance for poor people was being tested. The experiment was designed not as a randomized experiment but as a so-called "quasi-experiment." The objective was to understand the effect of copayment schemes on physician utilization. Some objections were raised to the experiment on grounds that the Fourteenth Amendment was being violated in some sense. When the case was brought to the court, the court issued an injunction against the quasi-experiment after hearing testimony that the non-randomized experiment could not produce estimates of the effectiveness of the copayment scheme, that is to say, just as with Thompson, with some concern about direct violation, using randomization, there is another side that says that if you don't do a randomized experiment, you won't be producing estimates that are credible, and you will have violated the law in doing so. So, it is going to get even more complicated.

DR. DIAMOND: There was a question back there?

MR. SHAPARD: Yes, I am John Shapard from the Federal Judicial Center. What I have is a little more observation than a question, and it has to do with my dispute with the application of the Belmont Report in

its essential thrust, at least, to experimentation in the criminal justice system.

My comment was merely that Professor Elliott's reference to the difficulty in satisfying his institutional review board about a replication of the domestic violence experiment. It struck me that what the institutional review board, what satisfied it in those circumstances was a protocol that would ensure greater harm to the persons who were most obviously the subjects of the experiment, that is the persons to be arrested, and, also, I take it that they had no concern about obtaining informed consent from those subjects in that experiment?

PROF. ELLIOTT: Informed consent was a given. So, it was certainly involved.

MR. SHAPARD: The persons who were arrested consented to arrest?

PROF. ELLIOTT: Since the experiment itself did not involve any data collection until after the police action and then only from the victims, no consents were required from the offenders. There were two issues involved. One has to do with the follow-up to police action, and the interviewing process with victims which required informed consent from victims. We were not required to obtain informed consent for the police taking action because that was an issue which the police department assumed the responsibility for, not the study. That would go on whether there was a study or whether there wasn't a study.

DR. DIAMOND: Okay, Professor Berk will give a very short response to the things we have talked about so far.

PROF. BERK: Just a couple of quick observations, things you can think about in response to these ethical concerns. First, treatments can be postponed so that ultimately the control group gets the treatment.

Second, it is possible to compare two treatments, both of which are interesting in and of themselves and don't in advance imply that somebody is getting a better or worse deal. Third, it is possible to anticipate the consequences of treatments and then offer reparations for harm done. Fourth, it is possible to buy access. To give you a quick example using AIDS testing in California, one of the questions that has come up is whether we can use AIDS test results from a particular individual to warn other people who may be at risk from him or her. Clearly, there is a real concern about confidentiality. One proposal is to buy the confidentiality of anyone who tests positive. That is, we might offer to buy a person's confidentiality for, say, $200.

Fifth, there is the possibility of capitalizing on naturally occurring randomization. For example, I am involved in a study of the impact of the Vietnam war draft. It turns out that in the late sixties and early seventies a number of us were subjected to a lottery for the draft. There is a randomized experiment and can be analyzed as such.

Sixth, there are very often administrative procedures, which approximate randomized experiments. If you can convince the administrators that these procedures are effectively haphazard, true randomization may become a viable and acceptable alternative.

In other words, there are lots of design strategies, that respond to ethical concerns. In addition, I want to emphasize that the status quo <u>is a treatment</u>. To worry about the impact of an intervention without at the same time worrying about damage that can follow from the status quo is short-sighted.

DR. DIAMOND: That concludes our session. I have an announcement about where we are going to have lunch.

(Administrative announcement.)

(Thereupon, at 12:40 p.m., a recess was taken.)

Saturday, March 14

Randomized Experiments: Anticipating Problems and Finding Solutions

DR. BORUCH: Why don't we get started. As many of you know, doing high-quality research in general and randomized experiments in particular is akin to trying to repair a watch while falling from an airplane at low altitude wearing shorts that are too tight. These things are demanding, and one of the lessons we have learned is that we ought to pay attention to the difficulties, the dilemmas, the frustrations and the like, and we ought to exploit them, capitalize on them, learn from what trouble we have.

The panel this morning has brought with them an alarming array of problems, as well as some potential solutions, and we are pleased that they are willing and able to be here to discuss them. By way of format, I think each speaker will probably get about 20 minutes, and rather than delay questions until much later in the morning, we will open the floor to questions immediately after each speaker. That is probably more in the interests of those of us with feeble memories and spontaneity than anything else.

Our first speaker today is Mary Toborg who is President of Toborg Associates in Washington. She is an expert in research on pre-trial release and is currently involved in a urine-testing survey in Washington, contributing in a small way to what Chips Stewart yesterday reminded us may be a kind of growth industry.

Mary?

MS. TOBORG: Thank you, Bob. In keeping with the theme of the session, anticipating problems and finding solutions, I thought I would tell you about some of my experiences and those of my colleagues at Toborg Associates in trying to implement randomized experiments in the real world of operating criminal justice agencies where things don't always work out the way the textbooks suggest that they should or the way that you would like them to.

I will discuss some recurring problems that we have run into in trying to conduct experiments and the ways we have tried to deal with those problems and then offer some general suggestions about ways that practitioners and researchers can work together to accommodate each other's needs.

My work, as Bob indicated, has been been mainly in courts research, not police or corrections, and there are some special problems in working with courts, particularly if you are working with the pre-trial release stage, which is where our randomized experiments have been conducted. First of all, you are dealing with defendants who have been charged with crimes but are presumed innocent of them. So, you have a number of constraints, such as issues related to voluntariness that don't arise when you are dealing with convicted offenders.

Second, there are many actors and agencies involved. It is not uncommon in a large urban jurisdiction to have 50 or more judges on your court who often don't act consistently, and at the pre-trial stage, you, also, have the involvement of prosecuting attorneys, defense attorneys, pre-trial release agencies, jail staff and others. So, you, also, have some constraints caused by the large number of people involved and the need to get some sort of agreement on consistent procedures.

It seems to me that in the last few years there has been a sharp increase in the level of interest by practitioners about research. When I did my first randomized experiments in the pre-trial release field in the late 1970's, a major problem was finding sites that would consider being part of such an undertaking. We did eventually locate four places that were willing to conduct randomized experiments to look at the effect of expanding pre-trial release operations, but when we first went to these places and to others, we were greeted by a real sense of suspicion about what this was all going to be about.

In contrast in the case of the randomized experiment we are now doing in Indianapolis with the support of Steve Goldsmith's office and Mike Robak, the Municipal Court Administrator, and others there, the research actually grew out of discussions with the practitioners. People working on pre-trial issues in Indianapolis told us that they wanted to buy some electronic monitors and see if placing defendants on them, in essence, under house arrest, could reduce jail crowding, and they came to us and asked us how they could best find this out. We suggested a randomized experiment. We got research funding from the National Institute of Justice, and we have been working cooperatively with the people in Indianapolis to implement the project.

We have had a similarly good experience with the work that we are now doing in Washington, DC, where we have a randomized experiment, also, funded by the National Institute of Justice.

In this case, we are looking at the effect of setting up a urine testing program for juveniles at both the pre-adjudication and the probation stages of processing. Now, that project grew out of our earlier work looking at the impact of urine testing for adult arrestees and

defendants, and in both of those programs there was a real spirit, I think of give and take between the practitioners running the program and the researchers trying to help set up a new, operational, working program in a way that would facilitate research.

I think it is largely due to the efforts of Chief Judge Ugast who is here, Judge Beaudin, the former director of the DC Pre-Trial Services Agency, Jay Carver, the current director, Al Schuman, the head of the Social Services Division, Larry Polansky, the Court Administrator, and a lot of other people who are at this conference and, also, back in Washington that these projects got off the ground.

From what others at this conference have told me, my experience is really fairly typical: the climate is right now in a lot of places for researchers and practitioners to work together on randomized experiments and on a lot of other things; and this is really quite different from the situation just a few years ago. I think this interaction of researchers and practitioners makes it a lot easier to do research, and it, also, makes the research a lot more policy relevant because it is so much better grounded in the realities of the operating system.

Turning for a minute to some specific problems we have run into with randomized experiments, I want to focus on two. One is finding a control group, and the other is keeping it once you have found it. First, finding a control group. This concept seems to cause different types of problems for practitioners. On one hand, there is concern that there may be unfair denial of service to members of the control group, but on the other hand, there is concern that someone in the control group may commit a serious crime and cause embarrassing publicity about why the person was in the control group rather than subjected to some greater restraint. How

can you deal with those concerns? There are several techniques that I think help. One, it helps to have a situation where something new is being added to the system. For example, urine testing of juveniles in Washington, DC, is a new activity. Electronic monitoring in Indianapolis for a pre-trial population is a new activity. In those instances you are not withholding something that was previously available. You are making something new available to part of the population.

Also, if you are adding something new, you may have genuine capacity limitations or the need for a phased startup that limits the availability of the new service and really provides ideal conditions for setting up a randomized experiment.

Second, you can limit the size of the control group and keep it as small as possible to still get good research results.

Yesterday, in fact, there was a suggestion that you can vary the size of the control group with the risk level of the population. So, you have a small control group for high-risk cases, and a larger control group for low-risk cases.

You can, also, allow exceptions in unusual cases. This really raises the comfort level of practitioners about experiments, and in reality, unusual cases are probably going to be handled in unusual ways anyway. So, you might as well accept that from the start.

The researcher's fear, of course, is that there will be too many exceptions, but often practitioners, at least in our experience, want the possibility more than the actuality. As Judge Beaudin was saying yesterday, in the DC Urine Testing Project, we allowed the prosecutor to say that certain defendants would never be placed in the control group, but the prosecutor didn't exercise that discretion. He merely wanted to have it.

I am going to turn now to a different problem. Once you have a control group or treatment group, for that matter, how do you keep it throughout the experiment? Our DC Urine Testing Program with adult defendants provides an example of the problems that can arise here.

When this program was first being planned, we really didn't think it would be possible to implement a randomized experiment with a true control group, but as we discussed this more with the practitioners, it did seem to be feasible, and so, one part of the study did include a randomized experiment in which selected drug-using defendants who were released before trial were randomly assigned to three groups. One group participated in a new program of periodic urine testing before trial. One group was referred to drug abuse treatment, and a third group was a control group where there were no special drug-related conditions of pre-trial release.

We were concerned about the randomization process because there are more than 50 judges and hearing commissioners on the DC Superior Court, and they are under a lot of pressure to move through the daily calendar, so we didn't really think it was very feasible to expect them to follow randomization procedures. So, we worked out a way that the randomization would be done by the Pre-Trial Services Agency. All the judge had to do was order the defendant to report to the agency for placement, and the agency then did the randomization to the three groups, and this worked beautifully. The defendants were, in fact, randomly assigned; the process was controlled; the groups were comparable.

The problems arose after that, and they arose in part because the pre-trial period in Washington is often quite a long one. It can be 6 or 8 months or even longer, and during that time a defendant may have several

appearances before a judge. It is not the case that you release defendants, and they come back for trial and that is that. There are a number of appearances in between. After the urine testing program started, and the judges became aware of it and were using it, they started ordering defendants who had not been initially assigned to the urine testing program to show up for and participate in it.

As I say, this occurred several months after the initial random assignment had been done. Ironically, this action by the judges in a way shows the high value that they placed on the program, and one of the questions that our study wanted to address was how the rest of the system reacts when you implement a new urine testing program.

Now, a second problem arose because some of the defendants in both the urine testing and the control groups opted to enter treatment during the course of the pre-trial period. We had made a provision that any defendant who wanted to enter treatment at the time of the initial random assignment could do that. We simply excluded those defendants from the experiment. We had thought that most defendants who wanted to enter treatment would want to do so at the beginning and would tell us that and be excluded, but over a 6-or-8-month period or longer, a lot of defendants may change their minds, and many of them did.

I might just say for those of you who are not familiar with Washington that the city has a citywide drug abuse treatment program that anyone can enter. So, we could not have stopped defendants from voluntarily seeking treatment without, in fact, denying them a service that had previously been available.

Needless to say, these problems, which I indicated occurred after the inital random assignment caused havoc with our original plans for analysis of the experimental part of this study.

Fortunately, the overall study had a number of parts. It included analysis of whether drug test results at the time of arrest are good indicators of pre-trial release risk. It looked at the extent to which participation in a pre-trial urine testing program serves as a risk-signaling mechanism, and it, also, includes implementation analysis, looking at operational and legal issues involved in introducing a systematic program of urine testing in a large urban city.

The results of the study are really quite interesting. I won't go into them here, but I will say that there is a series of draft reports that will come out over the next few weeks, talking about different aspects of this study, and I would urge those of you who are interested in this particular problem to read those and to give us your comments about them.

There are several things to be learned from our Washington experience. One is that when you are dealing with all the start-up problems that are involved in setting up a new complex operational program that also involves a randomized experiment, it is easy to focus so much on the startup problems that you don't anticipate problems that might happen down the road.

For example, if we had anticipated in Washington that the judges would order defendants into the Urine Testing Program in the middle of the pre-trial period, we might have been able to work with them and get an agreement that they would not do that for the duration of the experiment. In fact, we had been concerned about exactly the opposite problem, namely,

that the judges would not use the drug test results. Before the urine testing program started, judges in Washington as a group had rarely paid very much attention to whether defendants complied with their pre-trial release conditions. So, it was something of a surprise to find the extent to which the judges were, in fact, using the information from the urine testing program.

Finally, I want to make a few general comments about ways to try to avoid problems with randomized experiments. The first and most critical thing, I think, is the need for researchers and practitioners to work together from the beginning. It is important that the researchers understand the nuances of the local system, and I think that is one of the best ways to anticipate problems and to find solutions. As it was put yesterday, there is no recipe for a randomized experiment. It has to be tailored in each case to the local system and to the local problem.

I think it is, also, important to get the key actors in the system involved from the beginning, the chief judge, the prosecuting attorney, in our case, the pre-trial release agency director. If there are problems down the road, those are the people who can help solve them, and it is important that they be involved from the beginning. In fact, you may want to get a written memorandum of understanding signed by all affected agencies so that everyone knows what is going on. This was done in Washington, for example, with the seven agencies involved in the juvenile urine testing program. While it is important for researchers and practitioners to work together at the beginning, I think it is just as important for them to work together throughout the project. You cannot set up an experiment and go away. You have got to have frequent communication and watch over it and solve the problems as they come up.

Finally, I think you need a spirit of give and take by researchers and practitioners both. The researchers, I think, need to understand what it is possible for practitioners to do and not ask them to do the impossible. As a practitioner friend of mine said, "The practitioners should be partners in the research, not see themselves as victims of it." I think it is in this search for a working partnership between researchers and practitioners that we really have the best hope of anticipating problems and of finding solutions.

Thank you.

(Applause.)

MS. TOBORG: Are there questions?

JUDGE ROTH: Judge Roth, from Portland, Oregon. In your pre-trial testing, urine testing, is that done at a certain time after so many days? You said that it is about 6 or 8 months that your docket is behind in Washington, DC. What do you do if they are supposed to go for urine testing, and they do not show? Do you issue a bench warrant?

MS. TOBORG: Not at that time. There were several levels of sanctions in the program. The first were administrative sanctions that the program would apply after there had been two positive tests in a row or three over 3 months. At that point the person would be required to come in for more frequent urine testing.

Testing started on a once a week basis, and then went to twice a week; only if the person failed at that time would a violation report be sent to the court that recommended that a contempt hearing be held to see if the person should be held in contempt for violating release conditions. At that point, the judges had the option of holding a hearing or not, first of all, and then secondly, of holding a person in contempt

of court of not. The judges were very creative in Washington about how they dealt with this problem, because we have an overcrowded jail like everybody else, and you wouldn't want to take people and routinely have 3-4-or-5-month contempt sentences. So, what we often found the judges would do is hold the person in contempt, sentence him to 60 or 90 days in jail but suspend all but 2 or 3 days of the sentence. The judges felt this gave them an extra lever, if you will, over the individual, if he violated the condition again.

Are there other questions? Al?

MR. SCHUMAN: I am Al Schuman from Washington, DC. We worked with Mary on that project, and I would just like to emphasize that one of the reasons I think that the project is working very well is the attitude that Mary brought into this negotiating process because she was very open, and she was willing to look at the practitioner's problems and some of the realities that our day-to-day probation officers had to deal with, with the tremendous drug epidemic in DC, and if all researchers could learn from that and come in with that kind of attitude, that partnership attitude, I think much more research could take place around the country.

MS. TOBORG: Thank you, Al. That was, also, true on the other side in Washington and Indianapolis, where we are also working. They are just excellent jurisdictions in terms of practitioners wanting to know what makes a difference and how they can find out and what they need to, perhaps, do a little differently in order for us all to learn more. So, it has been a real two-way street.

MR. NOVIK: My name is Jack Novik. I am with the Criminal Justice Agency in New York. Mary, you have talked about the interaction between researchers and practitioners, and it is clearly important that

the practitioners understand the reasons for the research and become involved in the work of the researchers. I wonder whether you have any thoughts about the opposite problem of the researcher losing a sense of academic disinterestedness and becoming part of the policy-making apparatus that goes on in the decision-making system in the city, either in DC or Indianapolis and really becoming instead of a researcher independent and distant from policy making, actually getting involved in promoting and becoming an advocate for the research issue involved in that particular case?

MS. TOBORG: It is an interesting point, and I think it is another area where there has been a change over the last few years. I think that it used to be traditional that agencies would try things and then after they tried them, an evaluator or researcher would come in and look at them and make some comments about whether they thought it was good or bad. Now, I think, there is more of a spirit of working together from the beginning to try to get a sense of what works, and there probably is some loss of objectivity on the part of the researcher. On the other hand, there is a lot more information about what the realities of the problem are, and while I don't think that it is necessarily wise for researchers to become advocates of certain things, I do think that you have a responsibility to make recommendations to the people who are responsible for operating the system about things they may want to think about and change and possibly do better.

Jim?

DR. AUSTIN: Jim Austin from NCCD. Mary, do you have a summary, with all this urine testing going on, what is going to be the evidence that suggests that it is good or bad? What is the purpose and what would be the demonstration that it should be continued or discontinued?

MS. TOBORG: That is a tough question. A couple of things come out of the current research and the earlier research on drug-crime relationships, and that is you do find that drug users are at much higher risk of subsequent criminality. So, I think there is some reason to be concerned about identifying drug users and try to do something to reduce the drug use which is at least related in a lot of studies to crime.

Secondly, one thing we found in Washington is when you put people into a program of urine testing before trial, they very quickly sort themselves into two groups. There is the group that complies with the program and the group that doesn't, and the group that doesn't does that very quickly. They do it within 2 or 3 weeks, and they have very high rates of pretrial rearrest and, also, failure to appear, and that is true when you control for all the other things that we think affect those two outcomes.

So, what we have concluded in Washington is that at least pre-trial urine testing is serving as a signaling mechanism where you can take a group of people that you think because of their drug use are at higher risk of rearrest, release them subject to a urine testing condition, and then look at their behavior and allow them to identify for you the people within that group who are at the highest risk that you should worry about and the people who are at very low risk, and you should simply leave alone.

DR. AUSTIN: Is there any thought to doing a test that would accomplish that without the urine testing, to identify high-risk people without having to do a urine test?

MS. TOBORG: Yes. One of the things we have talked about is what other types of signaling mechanisms you could set up. There have been some efforts, as you know, with regard to supervised release programs to do that sort of thing. My impression is that those programs really haven't worked to reduce pre-trial rearrest very much. Now, maybe what we need are some more types of things related to that. In fact, one of the things that we have thought about is: let us say that you set up a urine testing program, and you have the signaling mechanism going on, and you find some people who are good release risks, and you leave them in the program and leave them alone; what do you do about the high-risk people? You are not going to lock them all up, and you couldn't do that if you wanted to. So, maybe with that group what you need to do is try to work out some signaling mechanisms for them, maybe some different types of drug abuse treatments that are available most places, maybe some different types of intensive supervision programs. I am not sure what those should be, but I think if you move toward a system where you try to screen the population for the low-risks and leave them alone and then focus your more expensive options on the people who are higher risks, you may start to move toward a more reasonable system.

DR. DENNIS: My name is Mike Dennis, and what I was interested in finding out was before you released them, did you do a urinalysis to determine whether or not they had drugs or alcohol in their system prior to arrest?

MS. TOBORG: Yes, in Washington everyone except people arrested on really minor charges like traffic charges has a urinalysis done while they are in the lockup - basically, it is after the time of arrest. It is not for alcohol, but for five drugs (opiates, cocaine, PCP, amphetamines and methadone).

DR. DENNIS: Are you able to compare whether or not those tests have any predictive validity in terms of later rearrest?

MS. TOBORG: Yes, that is one of the hardest things we are doing, and we are still working on it. The problem is that it is a complicated analysis because not everyone gets released before trial, and so, you have got some real selection bias there with the people who were detained. But, yes, that is one of the things we are working on: to try to see how well the initial urine test results classify defendants with regard to subsequent risk.

Thank you.

DR. BORUCH: Thank you, Mary.

Our next speaker is Richard Lovely who is assistant professor of sociology at John Jay College. He is interested for the moment in the problem of generalizing from some of the lessons he has learned from his work in restitution.

PROF. LOVELY: Good morning. I have a tale today to tell you that deals with some of the things we have heard about before, some things which may surprise you, themes such as soul and passion. Soul was a theme that Judge Beaudin introduced to us. He indicated that people who get involved in this business of doing experiments put their souls on the line. I want to tell you I am a person who did put his soul on the line, and it was trampled.

(Laughter.)

PROF. LOVELY: I welcome this chance. In a way, this opportunity is sort of a thing where someone called me up on the phone and said, "It has been a few years now; would you like to talk about it?"

So, here I am to talk about things such as soul, and even a little bit of passion, and I want to keynote my remarks for you practitioners in particular, to let you know that academics who get involved in this business in doing experiments can be passionate creatures. We are also idealistic, and the reasons we get involved in research are not mercenary. There may be some of us out there. I am sure there are perhaps some criminal justice program pimps, as there have always been poverty pimps, but I don't think the people who get involved in experiments are of that kind. Experiments are just too much hassle, and they require too much thought to bother, if you are just interested in making a quick buck. The program I am going to tell you about was a noble attempt that failed. It is a program you may have never heard of, and it is surprising that you haven't. The handout you got about it was on your table at the first session: the Connecticut Restitution Program.

If you will, imagine with me a program that would really try to solve all the problems that criminal justice program development and evaluation have suffered from the earliest days and which really still persist. Imagine a program that would merge the interests of practitioners and researchers from the very beginning, so that we didn't end up having to use a post hoc Band-aid Approach to trying to figure out what was going on. Imagine a program that tried to use multiple sites around the country to see how program variation might affect the delivery of outcomes in criminal justice; and imagine if we could have a research design which would give us data from all those different sites that we could then pool into a larger data base to do even more sophisticated tests; and imagine if we could collect that data through the use of a controlled experiment with random assignment to control and experiment

groups. It sounds pretty good, doesn't it? That is exactly what we tried to do. The experiment in restitution was an effort, in fact, that may well be what we could call one of LEAA's swan songs in trying to make good with the world and to accomplish its purpose.

I was involved in that effort as a local site evaluator in Connecticut. The principal investigator of this project was the Criminal Justice Research Center in Albany, New York. If you have read the handout, you will know that this program had some unhappy endings. I am a distinguished member of the criminal justice fraternity,I am probably the only person in this room who was a member of a program that was ever actually terminated by the Federal Government for failure to meet program outcome goals.

(Laughter.)

PROF. LOVELY: In a way that was a heartening experience, that the federal government had that kind of mustard. It only did so by virtue of the fact that GAO kicked LEAA in the tail to take a look at this program and to really see what was going on. I will say that I think that the effort was an honest attempt to do as much as possible to try to find out what restitution was all about and to answer some questions, and there were some very passionate and energetic people involved in this project.

Let me go through some of the basic ideas of what we tried to do and some of the problems that came up. First, I will tell you some of the frustrations that we suffered and then try to generalize from those frustrations to some lessons that we can perhaps apply to all experimental research situations.

I came here with some mixed feelings because I wasn't sure I ever wanted to do an experiment again. In fact, Seymour Sarason, a psychologist at Yale who has written any number of books about program development, observes that the problem with people who get involved in starting any kind of social programs is that once they do it, it is very difficult to get them to try to ever do it again. So, there is tremendous attrition. I will let you know when I finish whether I am interested in trying an experiment again.

Some of the problems that we faced. First of all, with all due respect to the so-called "lay people" in the audience and to the researchers, never overestimate the capacity of the judge, prosecutor, public defender or whatever to understand the peculiar logic and the jargon of social research.

Let me share with you an anecdote which shows that this can lead to tremendous frustration. We were involved in protracted negotiations with the Research Center, the Federal Government in Washington and the State of Connecticut to try to come up with a workable research design. There was considerable contention over the various aspects of the design, and we were in meetings with the head of the Connecticut Justice Commission which was at the time the great sugar daddy of Connecticut criminal justice.

This person was supposed to be acting as a mediator involved in the disputes, and after a 2-hour session in which we were carefully trying to lay out the alternatives to the experiment, he said, "You know, I don't understand the problem and why this is all such a big deal. Why can't you just take all the people who get restitution and figure out what they are doing and then take all the people who don't get restitution and look at

them and see what the difference is? Why do you have to go through all this stuff of randomly assigning people?" I was somewhat flabbergasted to think that we had sat through 2 hours of this meeting, and this gentleman hadn't the slightest idea of what the issues were. He was still totally naive as to what the agenda was as far as research goes. So, don't overestimate the pedagogical problem of teaching this in a quick and dirty way. It is, I think, formidable, and we can exaggerate how successful we are and think people know what we are talking about when they really don't.

Second, one of the things we came up with in this program, and we hence heard about it here before, is the marvelous ruse of using scarce resources to justify a random assignment to control or experimental groups. The ruse goes simply: "We don't have enough resources to service everybody. So, it is no big deal that some people end up in the control group. We cannot do it to everybody, and the fairest way to do it is through lot." This sounds wonderful. The problem comes in if you give the party, and then nobody comes which is exactly what happened in the State of Connecticut. We had a hard time maintaining this rational for it was a ruse from the very beginning. We had to have a rationale that was agreeable to the practitioner community, and that was the one we came up with. But for a long time they referred very few cases. So we had the staff waiting, ready to go, and no cases came. Then, of course, when a few cases did start coming in, we were in the position of having to turn down a case for the sake of the control group when the program virtually had no business at all. So, be very careful in using that ruse of scarce resources as far as justification for control groups. Later, I will have a general proposition that stems from this tactic.

Another problem is sticking to the program design. Now, in this program the request for proposal to Washington was responded to by the Chief Court Administrator of Connecticut. Of course, his staff were the people who actually came up with it, and they designed it as a program that the court would be administering. The idea of the program was that the judge at conviction would request that the case be reviewed by the Restitution Service to determine whether or not restitution was feasible: how much the victim had lost and what the capabilities of the defendant might be in paying back.

It was designed based on the textbook myth of the way the court works, that the judge is the person who makes the decisions in these cases, and as presented the Research Center was beyond reproach. It was so well done that nobody at Albany, nor in LEAA in Washington bothered to ask the question, but is that really how it works? And this follows Mary's argument about your needing to really be in touch with the nuances of the system. I don't even think this is a nuance anymore but it slipped by anyway.

The general procedure in Connecticut as in most states is that most convictions and sentences are determined by agreed recommendations between the prosecutor and the defense attorney. The judge doesn't have that much to do with it. Some judges don't always accept them, but normally they rubberstamp them and they go through, and there is rarely a problem. A few judges are cantankerous enough that they have their own minds, but for the most part, that is the general pattern.

So, the program as it was originally designed didn't really fit. What happened was absolutely wonderful from the perspective of someone developing a program. While the program was designed to be geared to a

limited range of cases to be referred after conviction, we had cases come from every conceivable direction and stage of the process. Private defense attorneys referred cases. Prosecutors referred cases. Even the first case ever referred to the program was not one which was supposed to be referred. Referrals were supposed to be made after conviction and the treatment was going to be whether the person got restitution or not. The first referral was a case of a convict who was getting ready for a parole hearing. His defense attorney decided it would sound good to tell the parole board that this guy was going to pay restitution to his victims. Naively we accepted the case before we realized what kind of situation we were in. That continually happened, and the irony was, of course, that because of the obsessive need to try to control the program, as far as the practitioner was concerned these opportunities to explore new avenues were thwarted. The contractual obligations we had to LEAA didn't allow us really to explore those opportunities, and of course, they did have devastating effects on our experimental design.

Another problem we faced was that we required the cooperation of the judges of the State of Connecticut in doing this experiment. In fact, we required far more than that. We clearly required the cooperation of the states attorneys who are the prosecutors in Connecticut, as well, and before the process started, of course, we did get agreements from people. Yes, this is a wonderful program, and we will cooperate and do our share. The chief court administrator signed off on the grant and assured that the judicial department was enthusiastic about it and would do its part.

When officials such as chief court administrators sign off on grants, however, I am rather convinced they don't necessarily read them. That is a job for the staff, and later on when it comes back to haunt them

that they signed to do all these things, they also don't necessarily understand the implications of what they have signed. I think it is very important to have absolute certainty that the chief decision makers understand what they are signing. The other rub is, of course, that a chief court administrator cannot dictate to an independent judge sitting on the bench what that judge is really going to do. So, it was not a very meaningful agreement in any case. The court administrator did make some efforts to try to give judges instructions on how they were to proceed, but it was not to much avail, since judges were effectively independent in deciding whether they wanted to cooperate with the program. Not surprisingly, most of the cases which were assigned to the control group were sentenced to restitution anyway.

Another interesting thing that happened, though, through no sense of malice by the program personnel, was a good bit of cheating on the experiment. The way we worked the random selection procedure was to compare the docket number of a case that got referred to a table of random numbers to determined if the program would serve as the case or not; again, based on the idea that we couldn't handle all the cases that we were to receive.

Some of the conscientious people in the program who were handling the referrals from the court, would peak ahead to see what kinds of docket numbers were coming up and then look this up on the table of random numbers. They did this because they were interested in marketing the program, and were trying to resolve one of the real dilemmas that we faced. At the same time that we were asking people to go out and sell the program - yes, isn't restitution wonderful, and shouldn't you refer cases

to this new program - they were faced with the specter of having a given case that they had hustled come back to the court with the program saying, "We don't want to work on this after all." So they wanted to try to save face by looking ahead, and to see if a case would be an experimental or control. They would hustle cases that they knew were going to be experimental cases, and they neglected those that weren't.

Now they were supposed to actually go into the court and look at the docket and see what kinds of potential business the program would have. Thus, they did not do this out of malice. I don't think that they ever really thought that it was going to dramatically affect the experiment. Again, it is important for people to understand the logic of such an experiment in order for them to be able to figure out whether what they do is going to negatively affect it, I think sometimes such harmful action is very naive, and it is not malicious sabotage of the experiment at all. They are just trying to do their job.

Let me speak quickly to something that Mary mentioned, and that is the problem of objectivity with respect to the researcher and the program itself. I was, indeed, a man in the middle. I was the liaison between the research center in Albany and the program people in Connecticut. For a long time I struggled to try to devise an experiment to make it work. I will confess to you, and we have heard this before, in particular in the police conference yesterday, that sometimes if people don't believe in the experiment they won't exactly do their job, I confess to you now after some years that after a time I didn't believe in the experiment anymore. It became just so much nonsense. There was no point to it. It was purely a perfunctory exercise, and we stopped religiously collecting the data that we were supposed to collect. It just was

senseless. See, here you have a problem of somebody who is involved in this program not because of mercenary reasons; I was there because I was an idealist trying to find out, and once I realized, well, we are not going to find out what I wanted to know, why bother? Why bother to go through with it, despite the contractual obligations we might have had? I confess that happened to me. It doesn't surprise me that it would happen to someone who doesn't even have a research orientation.

In conclusion, some quick notions about what we can generalize from my experiences in Connecticut to your problems in trying to devise experiments.

First of all, pick something that people care about. Pick an interesting problem. I always had the sense that in picking restitution, we picked a strawman. I mean who cares? What can really go wrong with restitution? "Nobody doesn't like restitution." In a survey of reactions to the progam of judges and prosecutors, I asked, "Is this really something we ought to do research about?" They said, "No, I don't care about that." We really need to either motivate their interest in the problem or it has to be obvious to them why we want to know.

It seems to me, that no matter what we might learn about restitution it is not going to change our ideas about what we ought to do about it that much. The worst thing that could happen with restitution, I suppose, is that some probationer is going to steal to pay off the restitution. That was about the worst thing we could think of that might happen as a result of it. So pick a problem that is compelling.

The other thing we have heard a little bit about in this conference is organizational change. Tony Bouza in particular pointed to that. Before you try to do an experiment become familiar with the literature in this area. You know there is a vast literature on how you change an organization. It is not as though we don't know anything about that. If you do certain things, you are going to find out that other things happen. We can predict what they will be. In systems, such as the criminal justice milieu, there are natural rhythms to change, to diffusion, which you should know intuitively. You cannot force things to happen before their time.

Let the natural entrepreneurs in a system do their thing. We know that it is easier for a judge to convince another judge to do something than it is for a researcher to come in. So, let the judges buy into it, and let that diffusion process work. That is even true with selling the idea of experiments, and that is my next point. We don't need 3000 different jurisdictions in the United States to do experiments. If we only have 100 to 200 where we have people who have bought in, who become true believers that if you want to know something this is the way to do it, that is all we need. As for the rest of them, if they don't come in to buy the idea, fine. Find people who believe and are willing to make sacrifices and be happy with that.

Another thing is don't bother to try to do an experiment on something before you have mastered the treatment. In the case of Connecticut, the rub was we were trying to develop a program, a new idea that had never been tried before at the same time as we were trying to implement an experiment. They were totally incompatible problems. They were also interactive problems, as you can perhaps figure out from the things I have told you.

If the treatment is problematic, if the case load of the program is problematic, you are going to be beating your head against the wall because there are continually going to be things that occur over which you have no control and you have no leverage to fix.

So, that is about it. But would I do it again? Surprisingly this conference has revived my spirits. I wasn't sure that I ever wanted to. It is a crazy thing to do. If you get researchers involved in doing experiments, they are likely to work very hard, and this is the rub. If the program fails, well, we can still learn something, and there is something for somebody to take away from it, but if the experiment fails, the researcher gets a totally empty suitcase. I mean there is nothing there, absolutely nothing, and you have never heard anything about the national evaluation of restitution programs because it was a total bust. There were many, many people years of energy and effort put into that project. There has been relatively little professional mileage that the researchers involved in that program obtained from it. It just disappeared into a black hole. I am sure many of the other people who were involved in the program, as I, never even wanted to think about it again, let alone write articles and publish books about it. So, there is a great risk to researchers in going into these things, particularly when we do it for idealistic reasons. Yet, at this point, I will just say this to you as practitioners: if you are brave enough to really want to know what is going on, what you are doing, well, I guess I am crazy enough to try it again.

(Applause.)

PROF. LOVELY: Questions, please?

DR. BORUCH: This is Bob Boruch. One of the chronic problems with attempts to do large-scale big science, applied social science has to do with understanding the stage at which one ought to terminate when things are not going well. I guess I am sort of curious about the stages at which you had an opportunity to terminate and the extent to which that was well-thought out beforehand?

PROF. LOVELY: That is a fascinating issue. There was actually an administrative law hearing about this case. We had a law judge come down. The government attempted to terminate the program, and Connecticut appealed, and we even hired lawyers. One of the key issues that the Feds would always raise was, "Why didn't you pull out, if you knew it wasn't working?" I think that is not a very viable option for people. In the first place, there are some sort of quirky things here. LEAA announced that Connecticut had won this grant before Connecticut even knew about it. There was a press release in the paper. So, that doesn't give you much of a chance to pull out. Another thing is to keep in mind that people who develop programs and even people who want to do research, I think we tend to be boundless optimists. We know we can make it work, damn it, and if we just keep working at it, we will figure out the problems. The other rub is that you end up in this case with a program that had 12 to 15 employees, I forget exactly, and these people's jobs were at stake. You end up with a program director who as nobly and as idealistically as he might have thought about the research at the beginning, ultimately his main concern toward the end was just keeping those people employed. They were counting on him. So, there was no easy mechanism by which to say, "Okay, we quit." That was the thing that later on we were constantly asked, "Why didn't you quit?" What I would say at

this point is that if you know the trouble is brewing, you are going to know it very early, and at the earliest stages you should pull the plug on it. It may be that today the environment is such that you could do that. This was the time, remember, when programs were never terminated for anything, and it really was the time when you got your money and did pretty much what you wanted. LEAA was at the time, desperately trying to change that culture, to force criminal justice programs to tighten up.

I dare say at this point that as an experimenter, I really would say to a practitioner, "Look, if you are not going to play our game, I will just take my ball and go somewhere else," assuming that the ball belongs to me and not them. Even if it is their ball, I am still not going to play with them.

So, my attitude about that today is probably a bit different than it would have been previously, and hopefully the culture has changed. Another thing that I think is probably beneficial today is that criminal justice programs really aren't an industry today the way they were at that time, and I think that the lower funding is probably very healthy in that we don't have perfunctory program development, and we don't have perfunctory research evaluation. What you have to ensure is that it is spontaneous on both parts. I think that this idea of the practitioner approaching the researcher with the question, "Hey, we really want to know about this," is the healthiest of all the possibilities.

Yes, sir?

DR. AUSTIN: Jim Austin, from NCCD. Just a comment, I know NIJ used to do these field experiments, and they had a very good methodology, I thought, where they would spend a lot of money up front, looking at potential sites, and then they would select like 3, 4 or 10 possible

sites, and then they would make it very clear that the money would be pulled if they didn't follow the specs of the design, and that seems to work pretty well.

PROF. LOVELY: One of the fatal flaws of this program as it was designed was that they attempted to do something like that, but they had the research request for proposals go out at the same time as the program request for proposals. So, there were parallel processes, rather than integrated processes, and the latter certainly would have been far healthier.

I tend to think that one of the things you have to appreciate though is that in designing a program there is a tremendous amount of groundwork involved and it is presumptuous of funding agencies to expect people to put out an effort to design a program and then just be told, "No thanks," without any kind of incentive.

PROF. LOVELY: Yes, sir?

DR. DENNIS: My name is Mike Dennis. Another thing that I know NIJ has started moving towards that is a little better than trying to have the right to pull the money out is to do multiple stage funding where you fund different stages of the process. So, there is more of a motivation that they have to complete the stage appropriately before they go on to the next one.

PROF. LOVELY: If there is an ear of NIJ here and the funding possibilities are there, I really think another fatal problem with the restitution program was that it was just too shortsighted.

The follow-up stage, for example, was based on a prayer that the national evaluator in Albany had that the states would continue to cooperate with them after the formal terms of the grant were over. That

brings to mind another issue about the restitution experiment, and that is that there was a lot of wrangling about what the grant contract said. To my mind this is much like in the business world. If it gets to the point where people begin talking about what the contract says, you are already dead in the water. If the spirit of the experiment and the whole enterprise comes to where people have to start watching the letters of the contract you are dead. You really have to have that enthusiasm, the spontaneous enthusiasm to maintain a healthy experiment, and when people start calling each other on contractual provisions, it is over.

Other questions?

Well, thank you very much.

(Applause.)

DR. BORUCH: Thank you, Richard.

Our next speaker is Stephen Goldsmith who is district attorney in Indianapolis. He has been with Indianapolis for nine years. He maintains that he is going to talk about experiments which he believes no one else will do.

MR. GOLDSMITH: The topic that is given to me "frustrations and opportunities" is a little bit elusive, but let me suggest some of the reasons that we are involved in experimentation in the criminal justice area, some of the frustrations and then some of the specific projects.

The main reason to go into experimentation is to take a program or test whether the project works and then expand upon it or discard it. One reason I am particularly delighted to be at this conference is because if you work as we all do in a parameter of limited expanding resources the efficient use of policing and prosecutorial dollars is of prime importance, not only to those of us who are practitioners but to the

public generally. In my community of Indianapolis, we spend $85 million a year on police, prosecutors and courts, and the efficient use of those resources is a primary reason to go into experimentation, and we can take the results of the experimentation and redirect the resources.

Now, those are the opportunities which is one topic. The frustrations are a little bit more difficult, I think, to put one's finger on. An initial frustration derives from choosing a standard of success. The difficulty comes from running an experiment which proves an intermediate parameter which has no long-term effect. For example, somebody said that all good managers should read all the Tom Peters' books. So, I read Passion for Excellence which states that everybody ought to have a mission and a credo. So, I went to my office, and said, "Okay, let us have a mission. Let us get the 12 supervisors in the office and find out what it is you think we do." It took us a day of arguing just in the prosecutor's office to figure out what we were about.

That is not irrelevant to this issue of experimentation because what is the standard of success we are trying to prove? For example, trying to prove that convictions are up strikes me as a meaningless experiment. It has no long-term effect to the person on the street. I can create an experiment to increase convictions to 100 percent by just decreasing the number of people I prosecute to one per year and ensure 100 percent conviction rate. So, the standard of success is a little bit difficult.

I would suggest crime reduction generally to those of us who are police or prosecutors as the standard we want to accomplish. Now, some of you here may think that doing something good to the offender is, also, a worthy result, but just in terms of my little universe, I will take crime reduction, and if there is some spinoff result to the offender, good.

Selling the standard to the public is the second area of frustration. Everybody talks about the political and media ramifications. I, frankly, don't think that those are very serious problems. I think that the risk from political or media criticism is very slim. If one announces a program, and the purpose of the experiment is to test whether resources are appropriately used in the community and takes particular caution not to let serious offenders participate in the experiment, then the risks are really pretty minimal. Will a person commit another burglary while participating in the experiment? Most of us are in communities where one additional burglary is certainly not a newsworthy event and probably not changing the status quo creates more burglaries than running the experiment to begin with. I would submit to you that almost any experiment can be run that doesn't involve murderers and rapists or people who were previously convicted as murderers and rapists, and the test can well stand the criticisms and embarrassment because it is not four or five burglars or 10 or 15 bicycle thieves; it is one murderer, one child molester or one rapist who would not otherwise have been out but for the experiment which raises the risk and makes for the newsworthy story.

Therefore, if the experiment is well thought out and designed to utilize community resources and reduce risk, the public issue is not a particularly significant one.

What are the other frustrations? I think one of the main frustrations of a practitioner in experimentation is that almost all of them are related to some sort of time study or longitudinal study which means that it takes so long to get the results that those of us who are in the system when we start the experiment may not be there when they come

in, and we cannot change our policies, and therefore who cares? I, of course, care because I care about my community generally, but some people are less farsighted, may not have quite that long a view.

It is very difficult to determine whether an experimental program is successful in a short enough period of time that somebody can do something with it. Any sort of longitudinal experiment that uses rearrest as the standard for success risks failure because most of our systems are so permeated with holes that "scoring" rearrest may evaluate police proficiency as much as the inefficiency of the person who went through the program.

A couple of other problems worth mentioning, it seems to me as frustrations are the ethical ones. Most of us, I believe, who are practitioners do not view the "ethical" problems as obstacles.As a defendant goes through the system, and is adjudicated guilty, so long as what happens to him is fair and within legislative limits the ethical restrictions are surrmountable

One other area of frustration, of course, are the judges. Some judges are pretty well convinced that they know what is best regardless of what the research is. So, I find it easier to run experiments that do not relate specifically to changing judicial sentencing but relate to pre-trial or prosecutorial decisions or police decisions or otherwise.

The last frustration I want to mention is the difficulty of getting from those of you who are out here with PHD's and statistical backgrounds some sort of unequivocal answer.

(Laughter.)

MR. GOLDSMITH: Popper, a statistician, says that the whole purpose for experiment is to determine what is not yet disconfirmed. He views his job as trying to disconfirm a thesis. It is never confirmed, it is just not yet disconfirmed. The lack of unequivocal response is somewhat difficult.

Let me take a few examples quickly and then see if you have any questions.

Let us talk about our spousal abuse experiment that is being run now. I, of course, as everyone else in the world heard about Tony and Larry Sherman's successes in terms of the Minneapolis spousal abuse experiment. So, I said, "Okay, we will do that," and we changed the law to allow police to make more arrests. "What is the next step forward?" I am a prosecutor and I wanted to find out does it have an outcome effect whether they are prosecuted and then treated or whether they are just treated, in other words whether pre-prosecution diversion coupled with treatment. It relates to whether we prosecute thousands of cases of spouse abuse a year. Therefore the current Indianapolis experiment on domestic violence really is a matrix of arrest, non-arrest, conviction, non-conviction. I am very much interested in the results, and they may have a dramatic effect on how we allocate resources in Indianapolis.

The study has a couple of interesting ethics issues. How do you randomize whether somebody is convicted or not. They are arrested; they come into the system, and we go, "Okay, you get convicted; you don't get convicted?" Rather, we changed our policy from not prosecuting to conviction to prosecuting to conviction except now, as a result of the domestic violence research study in Indianapolis, the people who participate in the experiment get a break.

Also, I would say that one of the things I think that made the Minneapolis experiment work, if you will, in terms of result, and it may make ours work is that domestic violence in terms of outcome can be more easily measured. If the woman calls that address in the first place you probably have a higher occurrence of police intervention in whether the act reoccurs than if you let a burglar out and wait for him to be rearrested because of the opportunity of rearrest of the burglar is less than the opportunity to have the spouse call when she is abused again. So, I think the spousal abuse prosecution project in Indianapolis is of interest.

We are doing electronic monitoring for pre-trial adults, post-trial adults and pre-trial juveniles, and different aspects of the monitoring have different implications.

As a result of all this cooperative process our electronic monitoring test does have more serious offenders in it than it originally did as I was persuaded to take a few more political risks which I do not think in the long run will be particularly significant, but will make the results of the experiment more significant and which may affect the allocation of dollars in Indianapolis in terms of jail beds, jail space and otherwise. I would say that in terms of electronic monitoring, it is interesting to run it as an alternative to pretrial detention, and the results will be important to determine to what extent the people on electronic monitors are recidivating and showing up for court, as contrasted to those who are locked up in jail which is a pretty good control group.

We are, also, using electronic monitoring as an alternative to home detention, and anyway, it seems to me that electronic monitoring has to be better than home detention or probation, but that is more of a test, and the experiment is the purpose of the test.

Let me give you a couple of other quick examples of things that we are not doing but which we should be doing. I am Chairman of our state Drunk Driving Task Force, and we have carefully followed and watched some research in the University of North Carolina Traffic Safety Institute on whether drunk drivers who were diverted to treatment recidivated more frequently than those who were just convicted. They are granting federal drunk driving money in Indiana and the state task force originally took the position not to give drunk driving dollars to any prosecutor who diverted first offenders, and there are a couple of prosecutors who think that that is shortsighted. So, now, we are in the process in the city of Evansville of saying, "All right, if you will let us run a randomized experiment, and we will give you money; and we will run the experiment and determine the results and therefore see whether this policy position that we won't fund first-time diversion is, in fact applicable in the State of Indiana," and that has some fairly powerful implications in terms of expenditure of dollars, results and recidivism in the drunk driving area.

Similarly, we are doing some urine testing in terms of pre-trial probationers, parolees. I think that with respect to the question about what do you look for, we ought to look to the extent to which urine test participants recidivate, commit crimes, show up for court as contrasted to those who either have money bonds, or, 10 percent.

The more difficult area relates to the 10,000 people a year we refer for treatment in lieu of prosecution. We have no idea, not the faintest idea which of those programs work, which of those programs are worthless, which of those programs cause recidivism, which reduce recidivism, which of the juvenile facilities are better than conviction or worse than conviction. Now we are trying to set up a local program evaluation unit to do some experiments and some research on which of the programs to which we refer kids work.

Adding all those things together the opportunities of this conference are particularly exhilarating because we are now in the situation where sophisticated policy makers, police, prosecutors, probationers and pre-trial are in a position where the public will accept and, in fact, eventually, I think will demand that we look at what we do and how we spend our dollars and see if it works. Thus from experiments that are tied to innovation experiments that are tied to current programs like treatment we might be able to go to our taxpayers and say, "You are wasting $1 million sending these kids or these adults to that facility," more and more you will see practitioners asking researchers to participate in studying what they do so long as researchers can guarantee that they will have some unequivocal answer at the completion of the research.

Thank you very much.

(Applause.)

MR. GOLDSMITH: Does anybody have any questions?

MR. MAC DONALD: Malcolm MacDonald with the American Probation and Parole Association. On the issue of marketing a research experiment by saying that you are not dealing with murderers and rapists, actually you know, you go on into aggravated assaults and all the other crimes, and

then you end up convincing your funding authorities, your legislative bodies that you should be working with Boy Scouts and virgins. The issue is somehow that dialogue has to change, I think, because there are experiments that can be done with violent people in the community. We end up with violent people in the community. It is a matter of fact, be it bad evidence, a poorly prosecuted case, whatever the case may be they get out on parole. So, we deal with these people, and to some degree we are selling a bill of goods to legislatures by saying that we are not going to work with these violent folks because they are there.

MR. GOLDSMITH: I have two responses. Probation offers perhaps the greatest opportunity for experimenting because most of what the public funds does not work, i.e. there is a great amount of money in probation without in most communities any thoughtfulness about how that money should be directed. Thus there are great opportunities for experimentation in probation.

Secondly, we can deal with more serious offenders. What we should look for here is a formula that says, "Let us pick the least expensive response which results in a positive outcome." That is to say, "Do we want to arrest or do we want summons, and what is the outcome?" Do we want to prosecute to conviction or do we want to divert, and what is that outcome? We have a conviction. Do we want to sentence to jail or do we want to sentence to treatment? What is the outcome of that? And each time we take those things and make each little cell and make sure we have enough numbers in each little cell, and we choose the least-expensive response that has an effect, and if it has no effect, then we ought to drop back to the less expensive treatment.

MR. MAC DONALD: For example, we would need to break down the categories of offenders, rather than say that a certain research design or program is not for murderers, we say that they are not for certain types of murderers. In Texas, our previous Speaker of the House was murdered by his wife, and you know, the woman never spent a day. So, there are types of murderers. There are child abusers that we can work with in the community, and there are fixated pedophiles that to my knowledge, there is no technology to turn the people's value systems around.

MR. GOLDSMITH: Actually, I agree. I think the bigger problem, the bigger risk for those of us who are practitioners is taking a person into a program who has a prior violent history, such that when he goes out, is unsuccessful in the experiment, you can be tagged with, "You should have known." I am more concerned about putting a burglar into a monitoring system who has a previous rape because of the previous rape as an indicator than being an appropriate subject of public criticism for allowing him out on the street.

DR. AUSTIN: This is Jim Austin again. It seems that in terms of doing certain kinds of experiments we have this problem of the practitioners being willing to do certain kinds of experiments but not others. For example, prosecutors are very willing to experiment with increasing their ability to prosecute, and to convict more people. Do you think it is feasible for prosecutors to consider doing an experiment where they, for example, are willing to experiment with reducing time served for certain kinds of people who are convicted, with the possibility that you may find, you may discover that you could reduce time served by X amount of months without having a negative impact on public safety? Is that something that is possible?

MR. GOLDSMITH: Here I am in all these jail reduction programs after I spent nine years trying to get the jail overcrowded. It is a little bit incongruous. I think though to redefine the issue, the answer comes out the same, that is to say that most of our communities have jail and prison capacity problems. The issue then becomes -- of course, you could keep expanding the numbers that are there -- how to keep within the capacity with the people who ought to be there and experiment enough so that you can get the people out who could take advantage of some other program. For example, with juvenile offenders who would have otherwise gone to boys school, we started a wilderness training program, and we should be experimenting with randomly sending people to boys school or to the wilderness program, evaluating the responses, and if you find they work, then you can reduce the size of the boys school. It is just politically unconscionable that you will find a prosecutor who would want to experiment to prove that he could reduce sentences, but I think you could do it in terms of redirecting the resources that are available in ways that are more efficient.

Thank you very much.

(Applause.)

DR. BORUCH: Thank you, Steve. We do have 15 minutes more, and the moderator is, also, listed as a speaker. So, I will take advantage of this opportunity. Pete Rossi mentioned yesterday that I would doubtless bore you with some ancient or medieval history on the topic. Partly because Tony Bouza has already invoked the Renaissance, I feel compelled to see if I could stand on their shoulders.

Let me begin with one of Saint Paul's letters to the Thessalonians. Those of you who read the New Testament zealously might remember that he advised us to try all things and to hold fast to that

which is good. He did not, however, provide much in the way of instruction about how to try all things or even to understand what is good when you have got it.

The research of interest here is partly a kind of actualization of that advice. We have been talking in one way or another about how to discover what is good, at least in some senses. Mike Dennis and I are actually trying to fill in some of the blanks in Saint Paul's instructions at Northwestern by focusing on the problems in the execution of large-scale, medium-scale and small randomized field tests and their solutions. We are doing it in part for criminal justice, but also for medicine because there are difficulties there as well as in employment and training, in transportation and other areas.

The premise is that where we have an opportunity to fail we will fail, at least at times. The data seem to support that. If you look carefully at, for example, the bibliography on randomized tests that we produced a few years ago, which contains about 300 experiments in the social sector run between 1950 and 1979, one finds that about 1/5 of them are, to put it kindly, imperfect. That is to say, they were not executed as designed for one reason or another.

Let me give five short lessons which stem from early experience and which reiterate some of what you have heard already.

The first lesson is that randomized experiments have to be <u>sold</u> before, during and after the experiment. This has to be done repeatedly. By way of appealing to ancient or at least 19th century wisdom on this, let me quote Claude Bernard who wrote the first introduction to experimental medicine published in the western world: "Man is naturally

metaphysical and arrogant, and thus, capable of believing that the ideal creations of his mind which express his feelings are identical with reality. From this it follows that the experimental method is not natural to him."

Well, between 1865 when that passage was published and 1955, when the Salk vaccine trials were run, the idea of randomized experiments and formal experimentation in the interest of understanding what works relative to scientific standards of evidence was not very popular in the medical arena or in any other social arena.

Since the fifties in medicine and since the midsixties in the social sector, the climate has changed a great deal. Mary Toborg suggested some of the nature of these changes. It is still the case, however, that whenever the idea of formal randomized tests is introduced into any new arena, into any new audience, the immediate reactions are, (1), it is illegal; (2), it is immoral, and (3), it is not much fun.

We know now far more than we did 10 years ago. We know that experiments can be legal, are legal or at least don't violate legal sensibilities terribly. Mike McCann's paper yesterday raised those points. There are, to be sure, exceptions and limitations. But by and large when the question has come up about legality the courts have taken the view that experimentation in the interests of understanding what works, so long as human rights are not violated, is legitimate and compatible with what is sensible. The Federal Judicial Center's monograph on ethics and the law should be far more visible than it is, not only in the legal community but also in the community of those people who do tests in employment and training for ex-addicts, minority females, and single parents.

On the issue of "it is not much fun," that bears on incentives for experimentation, which is a very ambiguous area. It is clear, for example, that money is an incentive. If you pay people to experiment, they will experiment often, to the extent that they are able. The money has to be front money in the sense of actually paying for time. It might, also, carry the promise of future money: "If you do a good job now, we will give you more later." But it is clear that money is not sufficient. People have taken the money and run in legal experiments, in employment experiments, and elsewhere.

That there are other incentives that have to be appealed to is also clear. The question, as Larry Sherman implied, as Steve Goldsmith implied, as Richard Lovely implied has to be an important one. Those middle ground questions are those which are non-trivial and important enough to generate some interest outside and sufficient interest within an organization to invest the time in randomized trials. As Tony Bouza expressed yesterday, asking these questions and trying to find answers can be intellectually refreshing and it can lead to a kind of vitality which is important in any institution.

The second lesson concerns the fact that the target group is always smaller than advertised. A variant on this comes from the medical community. As soon as the contract to do the experiment is awarded, as soon as the check is signed, the size of the control group drops in half. Although this sounds a bit cynical, it was true for about 5 or 6 years in the medical clinical trials arena. Part of the problem had to do with the exaggeration about problems, the number of homeless, for example, or the number of people with particular kinds of disease. Some of it had to do

with a particularly exploitative perspective, that is "take the money and run." Since I cherish mercy, I would prefer not to identify particular projects. But by way of solutions, the third reason for this particular problem, however, is legitimate. It is often the situation that we haven't done enough scouting in the area in which we try to impose an experiment to know how many cases to expect.

For the domestic violence experiments Larry Sherman's initial forcasting study is an example of this phenomenon. Many domestic violence calls to the police turn out not to be domestic violence situations and in the fraction that are, only a fraction of the individuals who should be arrested are actually present. One winds up with a situation in which out of 20 possible cases, only 3 or 4 become legitimate cases for the experiment. By way of solutions to this problem, in the past we have had to extend the time required for an experiment, and more recently we have tried to encourage high-quality pipeline studies to forecast this situation. Ultimately, I think the state of the art in doing pipeline studies will improve and insufficient cases will be less of a problem.

The third problem is that it is not only the new program that needs to be tested, but the experimental procedures themselves need to be tested. Between 1 out of 5 and 1 out of 10 well-designed experiments fail. The reasons for failure have to do with the novelty of the setting, poor scouting, insufficient time, and maybe, insufficient money.

One recommendation is to pilot test everything or at least write off the first 6 months of the experiment as a de facto pilot test in the interests of developing a better sense of understanding the situation with the hopes of doing a better job in the long run.

On a larger scale, for example, Jerry Corsi did some fine and underadvertised experiments in administrative law using telephone hearings instead of in-person hearings in the state of New Mexico. He wanted to understand the extent to which telephone hearings produce the same results in appeals cases, on food stamps, unemployment insurance and the like, as in-person hearings.

He pilot-tested the entire program in one city, (Denver), first in order to understand the extent to which he could get agreement from lawyers, from civil liberties groups and the like about conditions for experimentation, but also to assess the extent to which he could actually randomize after people had promised that they would participate. Those pilot tests in Denver ultimately led to a far better experiment. Similar things can be said about early experiments at the Vera Institute being replicated much later and about much larger experiments conducted by the Department of Labor. The bottom line is: pilot test procedures as well as programs.

The next lesson concerns randomization. Randomization is crucial in a field experiment and to the extent that it can be subverted, it will be subverted. This is a variation on old Navy rules which most of you know better than I do. Ten percent don't get the word, and another 10 percent will fail to understand it or fail to buy into it, even if they do get the word.

Again, randomization and the randomized assignment process is not natural. Monitoring it is essential. Understanding how to get it off the ground is also essential. That is the reason for the pilot test.

Another suggestion, directed at some of the problems that Richard mentioned, is to have large-scale, important experiments routinely generate a 10 to 30 percent sample very early on in the study and run through the entire analysis of the data based solely on a 10 to 30 percent sample in an effort to ensure that things are working as advertised, and to correct problems midstream.

The next basic truth is that no program is ever delivered as advertised. Even if it is, we often don't know what exactly is delivered. The only person who seems to know is the offender, who is generally unwilling to talk about it. Good field experiments ought to measure this and control it. The restitution experiment ought to be written up and publicized, in part, because it is an important example of failure to understand the delivered treatment.

In the long run, we ought to know far more than we do about what constitutes ambiguous kinds of treatments such as counseling, intensive parole, or non-intensive parole. What those types of treatments mean in reality isn't at all clear. What we understand about them sounds more like the oral history of primitive tribes than it does of a literate society. If one does experiments with these treatments, one ought to have a better sense of what exactly goes on.

Let me finish by appealing for the last minute to 13th century rabbinic history. We have talked about high-quality research, including randomized trials and some non-randomized ones. David Farrington outlined some perspectives on that yesterday. We heard from David Chambers and Bill Bowers on time series analysis. What we haven't talked about is the extent to which the quality of the investment ought to match the kind of question that is being addressed.

The rabbinic story goes as follows: In the 13th century the rabbis developed an argument over how one ought to draw a sample of olives in order to judge the value of an olive crop. The reason for the need to make a value judgment had to do with tithing to the synagogue. Should one go out to the nearest basket of olives, grab a handful, take a look at that and then declare the value of the crop? Or should one hire an able sampling statistician, contract with an able survey research group to design a scientifically valid sample, and come up with an estimate of the value of the crop? That kind of issue, understanding what one ought to do is important today. It is as important in surveys as it is in designing experiments and quasi-experiments to estimate the value of potential programs.

In that particular instance, the rabbis argued for about 25 years over the issue. As an aside, I might mention that this story actually comes from a book entitled Probability and Statistical Inference in Ancient and Medieval Jewish Literature. It is a slender volume and doesn't sell well, nonetheless it has lots of interesting contemporary problems. As I said, they argued this vigorously. The resolution went as follows: In this and in other matters the investment that you make in trying to address the question, the quality of the information that you seek, has to depend heavily on the origins of the questions. For example, if the demand for the information is rabbinic in its origin, if it is a matter of low-level bureaucracy, in their own terms, one can afford to go to the nearest basket and pick out a handful of olives and estimate the value of the crop.

If, on the other hand, the demand for the information is Biblical in its origin; if God wants it; if it is of some import, you jolly well ought to go out and design yourself exceedingly good samples to come up with an estimate of the value.

That rule of thumb can probably be extended well into this arena, also. It won't always be possible to do very high quality work. We should exploit those opportunities when they occur, but we also ought to recognize that the nature of the investment that we make has to depend on the importance of the question, on how defensible the answer must be, and on how much we have to invest in answering the question.

Thank you.

(Applause.)

DR. BORUCH: Any further questions, comments? Yes?

PROF. BRANDL: I am John Brandl from Minnesota. Bob, I would like to pick up on the first of your lessons that randomized experiments have to be sold before, during and after, and my question will be ultimately about the management and replication of experiments, but that point that you made has been coming up over and over again, and I have just started jotting them down, and I will mention a few instances.

Professor Lovely tells us that if you fall into a funk, it is hard to run an experiment. Steve Goldsmith said that judges cannot be counted on in his view, to get enthusiastic about an experiment; so, don't involve them. Professor Sherman told us about Kansas City police sabotaging an experiment because they didn't like the idea. Judge Beaudin talked of the difficulty of convincing prosecutors to permit drug users to go back on the streets, and Tony Bouza, my friend from Minneapolis went so far as to say that the most important thing about an experiment is the

Hawthorne effect which is a profound parenthetical remark of his, and he resolved that by going out and buying beer for the experimental team.

Now, I think the point of this is that for many experiments the enthusiasm of the people who are running the experiment is not just something to think about or work about, it is part of the treatment. It is not parenthetical; it is of inherent importance. Now, I am more familiar with experiments in other areas. I have been involved in some in income maintenance and especially in education, and in education at any rate we have got a whole menu now of things we have learned in the experiments, but people aren't picking up on them. They are not using them which is not surprising, given this list of things we have just been talking about here.

My question is this. Are these aspects of management, the motivation, the enthusiasm, the involvement of the people who are carrying out the experiment, are they systematically measured and considered part of the treatment in the various experiments we are talking about here? If not, I believe the experience from other areas suggests both that these experiments won't be replicated and that if they are replicated, the results will be disappointing.

DR. BORUCH: In answer to the question, if the enthusiasm or the caliber or the character of whoever is doing the evaluation is being observed, I know of no good systematic case study of that, and the answer is no. I know of none. There may be some out there. It is an important point, but it is one that affects almost any of the sciences, and it has affected it historically. I mean you can go back to Iven Caldoon and look at what he says about personality criticism of historians or Claude Bernard in medicine. It depends on the doc and so on or the doc as

investigator. Probably as a research policy matter the bottom line has got to be replication. We know Minneapolis was unusual in some respects. I mean they had an unusual chief, an unusual team of researchers. It was an unusual department in some respects, I mean good people in ways that are different from other cities, and an unusual target population. How many cities have Indians going off to the reservation instead of sticking around to be arrested after assaulting their wives?

So, I guess as a research policy matter the replications will help us understand the extent to which the treatment, say, arrest, or the counseling is robust against what should be very low-level influences from the researcher.

I would hope that the experiments do so, but the second stage after that has to do with the extent to which having done the replications are the innovations actually picked up? Well, for domestic violence cases it turns out that people pick that up probably a little prematurely, and I mean the evidence is there, but it is still a bit equivocal.

I don't share Stephen Goldsmith's belief in unequivocal information, at least in this sector, but partly because the replications of the Minneapolis experiment are cooperative and individually they are going to be important. They will, also, be important in aggregate, partly to understand the distribution of the effect of arrest because it is going to range from large to small and partly to understand and assure that researchers' imperfections are spread out, researchers' influences are spread out. Those data will be analyzed in aggregate, as well as individually, partly to address that kind of question.

Yes, Mike?

DR. DENNIS: Mike Dennis, again. One of the things, if you look in the abstracts you will note that the DWI rehabilitation experiment and the TARP experiment, Rossi's TARP experiment, both involve cases where there had been some earlier research that showed that this program seemed to work, but that there had been dedicated research staffs in the kind of situation you are talking about, and they then attempted to replicate them on a large scale, using existing criminal justice officials. They were doing it across the entire system and with the people who really had to do it, and generally speaking when you do that, you do get a dilution of the effect.

Now, if you are out there looking around for what works, you try to test first the most likely case to find it. If you don't find it there, you don't worry about whether or not it is the system that is keeping it from working. Once you have found it, then you see if it works in the system as is or if you can tinker with the system to make it work, but first you have got to find something that works before you try testing it in the weaker system-wide situation.

DR. BORUCH: We are due for a break, and I have two announcements, possibly three. We would like everyone to complete the evaluation form that was sent with your material and return it to the registration desk outside this room.

If you have managed to lose your form, there are extra forms available. You may fill out two forms, if you like. After the breakout groups are finished at noon, lunch will be served in the same place as yesterday, that is to say at the Gaulois Restaurant.

Finally, there is a registration table outside and a signup sheet for concept papers for the State Justice Institute. Dave Tevelin will talk about that later this afternoon. It is an important new enterprise for those of you who are interested.

AFTERNOON SESSION

Saturday, March 14

Interpreting and Utilizing Results from Field Experiments

PROF. REISS: I am going to be very brief so that we can, though we are starting a few minutes late, try to keep on schedule.

The purpose of the first part of this session is to have an overview of how you interpret results of experiments and particularly what are the implications in terms of utilizing the results of experiments.

We are going to have two presentations, and we will have questions from the floor after each one.

First, Sally Hillsman from the Vera Institute of Justice in New York City is going to talk about the pre-trial diversion experiment. Then following that, Rick Lempert is going to talk about -- not an experiment he did, but an experiment he is taking a look at -- the Minneapolis domestic violence experiment. We will have questions after each one.

Sally, do you want to begin?

DR. HILLSMAN: Before talking a little bit about interpreting and utilizing the findings of this field experiment, I thought I should probably tell you at least a little bit about the history of this experiment because it occurred in the latter part of the 1970's which is practically ancient history these days, and so, it is not one that is perhaps freshest in all of your minds.

In 1975, the National Institute of Law Enforcement and Criminal Justice, NILE, which was actually the predecessor of NIJ, funded the Vera Institute to conduct an experimental evaluation of one of the first adult pre-trial diversion programs in the United States, the Court Employment Project in New York City.

It was begun by Vera in 1967, and CEP was acknowledged to be the prototype for the very large number of diversion programs that spread across the country during the next decade. Indeed, Malcolm Feeley has referred to the 1970's as the decade of diversion.

Vera's experimental evaluation of CEP, which by this time was an independent organization funded by the City of New York took 3 years and cost $600,000.

Despite the time and expense, the National Institute was extremely anxious to have the evaluation carried out. First, pre-trial diversion was not merely a passing reform idea. It was a process that had been wholeheartedly embraced by the criminal justice community. Diversion programs had expanded rapidly in the late 1960's under Department of Labor funding, and then in 1971, LEAA adopted pre-trial diversion as a central major court reform concept. It required all states receiving bloc grants to include these programs in their plans and, it also funded many of the programs directly through its discretionary grants.

As the decade progressed, millions of federal dollars and increasingly substantial amounts of state and federal money were spent on pre-trial diversion programs, especially those targeting youthful adult offenders.

Despite this stunning success, the programs were controversial. By the mid-seventies diversion faced legal and ethical challenges because of its potential use as a pre-adjudication sentence imposed without due

process or informed consent. Individual programs, also, faced empirical challenges from non-experimental evaluations which raised questions about the extent to which they were accomplishing their aims.

The CEP evaluation was planned by NIJ, Vera and the project itself, the Court Employment Project, and carried out in the late 1970's at a time when the "alternatives movement," which was largely initiated by pre-trial diversion, was beginning to expand into other areas, as well: court-based mediation, community justice centers, community service, restitution and so forth.

The ability of researchers to answer as unambiguously as possible whether the most established of these alternatives, pre-trial diversion, was, in fact, displacing prosecution, conviction and punishment or whether it was a questionable addition to existing criminal justice processes was an issue of considerable relevance to many of the newer reform efforts that were being launched at that time. It not only had ethical implications for policy makers, but important practical and cost ones, as well.

By the mid-1970's, a great deal of federal and local money had also been spent on unsuccessful attempts to implement experimental evaluation of pre-trial diversion programs. It was widely accepted that the pressing policy issues could not be answered without a true field experiment.

Meanwhile pre-trial diversion programs proliferated. Obviously, therefore, the Justice Department wanted the policy issues resolved so that the decisions about support for diversion could be better informed; but they also had a policy agenda on the research side. I might add here that Joel Garner at the National Institute was at that time the sort of

major person there pushing for experimental designs; he was the major figure there who really helped us implement this experiment successfully and he was extremely helpful, not only in keeping pushing us so that we filled all of our responsibilities to produce a true experiment, but, also, substantively aiding us in thinking about the design and its interpretation.

NIJ wanted to demonstrate that field experiments involving case processing could be carried out successfully in courts. Experiments were beginning to be seen as useful and desirable research tools in the criminal justice system, and it was, also, increasingly recognized -- something we have talked a lot about here -- that there are some kinds of questions that can be only answered by experimental designs.

In the mid-1970's pre-trial diversion was identified by NIJ and others around the country as just such a case.

The design of the evaluation reflected these purposes. It was first and foremost a field experiment -- random assignment to the diversion program or to normal court processing. Because the screening of defendants for diversion by prosecutors, pre-trial programs, judges, defense counsel, and defendants themselves was generally a very highly selective and complex process, researchers and policy makers came to agree that there were no naturally-occurring groups, and that none could be statistically created, with which to compare diverted cases in order to demonstrate what would have happened to them in the absence of the diversion program. An experimental design, therefore, was the only option, if the results of the evaluation were to be interpreted unambiguously.

I would like to add here that we have talked a lot about the implementation issues, and so forth, in experiments and that we had all of the problems that everybody has talked about. Somehow we managed to muddle through it all and did, indeed, reasonably successfully put in place and complete the experiment. But besides the random assignment, there were two other aspects of the CEP research that were extremely important to the issues I have been asked to address this afternoon: those are how adequately the data produced by the experiment can be interpreted, and how fully the findings can be utilized by policy makers.

The first of these involves issues about the extensiveness of the data analysis that was undertaken, and the second involves our attempt to provide some kind of conceptual explanation for the findings.

Let me turn to the first of these. Like many criminal justice reforms and interventions, pre-trial diversion programs espouse multiple, complex goals and rationales. This meant that the CEP study had to address an extremely elaborate and detailed set of dependent variables, along which the randomly-assigned groups would be compared.

Of paramount importance, however, was the comparison of experimentals and controls with respect to case disposition, in order to address the central claim that diversion was an alternative -- a process to displace full prosecution, criminal conviction and harsh punishment and the stigma thought to be associated with these for youth.

There were other equally complex issues addressed by the experiment, but I would like to focus here only on our interpretation of the case disposition results. If we want to talk about the rest later, we certainly can.

Pre-trial diversion's ability to displace full criminal prosecution and punitive sanction was the issue of most immediate importance to policy makers, but it was, also, the analysis of these data that most influenced the use eventually made of the experiment's findings.

CEP was designed to provide youthful adults facing prosecution on felony charges with an alternative to traditional prosecution and punishment. Specifically, CEP sought to obtain a dismissal of the felony charges by diverting defendants to a program of supervision and social services.

Data from the field experiment showed that of all the defendants who were diverted to CEP and who completed the program successfully, 55 percent of the experimentals, did have the pending charges dismissed by the prosecutor. However, 41 percent of those who did not complete the diversion program successfully, also, had the pending charges dismissed.

Overall, therefore, 72 percent of the diverted experimental group had their charges dismissed, a percentage which compared favorably with the 46 percent dismissal rate for the control group.

The initial analysis of the experimental data therefore, indicated that diversion to CEP had some impact on the proportion of cases dismissed. The difference between the 72 percent dismissal rate for the experimentals and the 46 percent rate for the controls was statistically significant.

However, researchers and program administrators then, together, walked through a very detailed and far more refined analysis of experimental and control comparisons, a mutal process that resulted in a quite different interpretation of the experiment's overall findings.

Extensive additional examination of the data suggested to all of us that the initial statistically significant findings could not be viewed as meaningful in any important substantive or policy ways.

For example, the control group's experiences showed that if there had been no CEP -- no formal pre-trial diversion -- almost one-half the experimental defendants would not have been or could not have been prosecuted.

In addition, although the remaining control cases were prosecuted, most of their cases were disposed leniently. One out of four controls was convicted of a violation (for example, disorderly conduct which is not a criminal offense in New York) and an equal proportion were adjudicated youthful offenders, which is a non-criminal finding of guilt.

In short, while all those in the field experiment who were selected by criminal justice decision makers for pre-trial diversion had been charged by the prosecutors with felonies, less than 7 percent of those not diverted, the controls, were convicted of a criminal charge, most at the misdemeanor level. This is compared to 2 percent of the diverted experimentals. The overall extent of criminal conviction for the felony population targeted for diversion, therefore, was not very great. Although the difference between the two groups was statistically significant, it was not viewed as evidence of a meaningful displacement outcome, either by CEP itself or by many local policy makers.

I would like to turn now to the second aspect of the CEP experiment which is particularly relevant to how these findings were interpreted and then ultimately used. NIJ, Vera and the program itself recognized at the outset that if the results were to be useful for policy purposes, the fact that the experiment involved a single program -- only

one research site -- had to be taken into account. Even powerful, refined experimental results for the Court Employment Project could provide little policy guidance to programs elsewhere if the design was unable to provide any explanation of the findings that could be used to review the activities of programs not subject to the experiment. Without this, the field experiment would merely provide an evaluation of one program whose failed efforts would have little or no relevance to understanding pre-trial diversion generally.

Indeed, unless the findings could be placed in some framework, even CEP itself would be hardpressed to know how to utilize the findings. How should the program change? Was the root of its problem the program's failure to implement diversion properly, the failure of the criminal justice system to divert appropriately, or was it more a fundamental problem of the failure of the pre-trial diversion concept itself?

To provide some explanatory framework, therefore, the research generated and assessed substantial amounts of data that were collected outside the experimental design. We compiled extensive collateral data from research interviews with defendants, program personnel, prosecutors, defense attorneys, judges, as well as from an analysis of cases rejected from the program and from the experiment, and from an analysis of official court statistics covering the decade prior to the field experiment.

To illustrate their use in interpreting the experimental data, let me focus again on the dispositional findings, although as I said, there were a variety of other sets of findings that we, also, addressed.

The diversion literature of the sixties and seventies had tended to evoke the image of criminal courts as prosecuting and convicting most cases brought before them, even the less serious ones, but certainly the felony cases.

During that same decade, however, empirical studies of courts were becoming far more frequent, and the evidence they presented began to undermine this picture of American courts. The CEP research compiled data on New York City, as well as other urban jurisdictions, which showed that a substantial proportion of arrests were voided, dismissed or rendered roughly equivalent dispositions.

The statistical data, also, showed that this was not a recent phenomenon. By the time of the CEP field experiment, for example, about 40 percent of all cases in the New York City Criminal Court were dismissed, including over one-third of all those commenced by felony arrest.

Therefore, to understand the results of the CEP experiment one had to understand something not widely recognized at the time, namely, that in the normal course of case disposition, not only was a defendant's chance of dismissal high, but there was also a substantial chance that another disposition would result, one that was as a formal matter something less than full criminal conviction.

In most jurisdictions, many dispositions carry some of the characteristics of diversion. It was rather difficult, therefore, for formal diversion programs like CEP to displace punishment and stigma in the dispositional process for a significant number of cases.

Nevertheless, not all felony arrests in New York or elsewhere are dismissed in the normal course of adjudication. Indeed, most are not. The CEP research, therefore, had to explore more than general dispositional patterns if we were to understand why these convicted cases were not the ones that were being diverted.

In order to understand who gets diverted and why, it was necessary to examine what influences the exercise of prosecutorial discretion and what influences the agreement of the defense.

To do this, the CEP research studied cases rejected for diversion. The data revealed that the primary reason prosecuting attorneys refused to divert otherwise eligible felony cases was that they viewed them as "too serious," for diversion. However, what prosecutors meant by "too serious" was that the cases were prosecutable as misdemeanors or violations.

In short, the prosecutors were actively blocking the diversion of many convictable cases while they acquiesced in diverting the less convictable ones. A senior prosecutor in charge of diversion screening said in an interview that of the cases diverted to CEP pre-trial about one-half would fall into two categories, and this is a quote, "those where there is a technical problem with the case and those that would have been adjourned in contemplation of a dismissal without the program but where the assistant district attorney felt that the defendant needed services."

This self-report parallels precisely the findings of the field experiment: almost one-half of the control group was dismissed.

In further interviews designed to explore why these decisions were made, prosecutors who selected cases for diversion into CEP reported that as a matter of course, their goal in diverting cases pre-trial was to create some form of supervision for defendants who would not receive any supervision as a result of normal adjudication.

Some defendants, generally those knowledgeable about the court system were able to prevent this by withholding their consent to diversion.

For less experienced defendants, however, research interviews revealed that diversion appeared to them to be a break. Many reported anxiety that harsh punishment and especially jail sentence was the likely alternative to accepting diversion. Furthermore, interviews with their legal counsel suggested that defense attorneys were not always advising their clients that the dispositional benefits of diversion were marginal, even when they were fairly certain they could obtain similar lenient results without it.

Responding to their own structually-defined interests, defense attorneys were glad to see their clients diverted because the process was rapid and maximized the likelihood of a dismissal with the least expenditure of their own all-too-scarce resources. Thus, the results of the CEP field experiment, and the collateral data collected, suggested the following interpretation of pre-trial diversion as a process to both practitioners in New York and to the researchers.

Simply providing decision makers with an opportunity to use an alternative process or disposition is not likely to be a sufficient incentive for them to do so, or for them to use such new additional resources as those providing them would prefer.

For traditional patterns and goals to change either powerful incentives must be introduced to maximize the likelihood that new goals will develop or innovators must be able to exercise independent control over key dimensions of the decision process.

In the case of pre-trial diversion generally, not just in New York, prosecutors necessarily control this process through their exercise of discretion. There was nothing in the field experiment's findings that suggested the availability of diversion services or supervision could

ensure or even encourage, prosecutors' use of this pre-trial alternative primarily in cases they could otherwise convict with a sanction.

The Court Employment Project's management had been actively involved in this process of interpreting the findings of the field experiment at every stage of the data analysis, and its reactions to the results of this process were dramatic. It entirely abandoned pre-trial diversion, withdrawing the social services and supervision resources from the system, at least for this purpose, and undertook a new total redesign of this program.

It began to focus on cases post-conviction in the upper, rather than the lower court, and it fashioned a deferred sentence program designed to displace detention and imprisonment for young offenders, mostly robbers.

The experience of the CEP diversion experiment, incidentally, was confirmed by two other field experiments that followed shortly afterward in quite different jurisdictions, one of which was done by Jim Austin of NCDD. This experience sensitized criminal justice policy makers in New York City, and more generally in the state, to the difficulties involved in fashioning alternative court programs, especially those whose role was to displace imprisonment. The CEP field experiment did not deter policy makers from trying to create new kinds of alternatives but rather it encouraged them to establish far higher standards of judgment for assessing whether efforts to forge alternatives were likely to succeed.

Elsewhere in the country as well, experimental studies of criminal justice reforms, of which this early NIJ-sponsored evaluation was one of the first, have encouraged greater caution in the alternatives field, particularly about launching efforts to displace existing

processes, especially punitive sanctions such as imprisonment, without experimental verification of their outcomes.

In the area of alternatives, the field experiment is now clearly viewed as not only feasible and desirable but, also, as the standard against which other methods of assessing impact are judged by researchers and, also, by policy makers.

Does anyone have any questions?

(Applause.)

PROF. REISS: Questions? If not, we will move on with the second presentation.

PROF. LEMPERT: I would like to raise some of the issues that arise in interpreting results from field experiments. In doing so, I shall use the Minneapolis spouse assault program - of which we have heard so much - as an example. I should say by way of disclaimer that Larry Sherman or Dick Berk should really be presenting this talk. I am not an insider and am limited to what I have read in published sources. I should also note that the experiment was too rich and time is too short for me to address all aspects of it - so I shall just highlight certain results and use them to pursue with you the logic of how one interprets experimental data and decides what policy conclusions are justified.

In generalizing from field experiments there are three questions that especially concern us. The first concerns the quality of the experiment itself - was the experiment properly designed and was the design executed as intended. The social science jargon for this is "internal validity." This is the question that Tony Bouza, the Minneapolis police chief and the Minneapolis City Council had to consider once results of the spouse assault experiment were in. They wanted to

know whether the experimental results really demonstrated that arresting spouse abusers in Minneapolis would reduce subsequent assaults. This turned in large measure on whether the experiment was properly designed and conducted.

There are many ways in which research designs can fail either in their conception or as applied. If the Minneapolis experiment failed in any of these ways, it might be a mistake to rely on it. Then again, it might not; most research efforts fall short of perfection but we can learn from imperfection. Police chiefs and city council members in cities other than Minneapolis had an additional question that they had to confront before they relied on the Minneapolis experiment as a guide to policy. They were not only concerned with the internal validity of the experiment but also had to wonder whether what worked in Minneapolis would work in their city, this is the "external validity" question. Just because something works in one setting doesn't mean it will work in others.

The third question which confronts anyone who relies on a field experiment for policy guidance is closely related to the second in that it also involves issues of generalization. The experiment may be well conducted and its results may be likely to apply across a variety of settings, but we must still ask what have we really learned; what precisely are the results we obtained? It is too easy to generalize in ways that go beyond precisely what the experiment tells us.

For example, if spouse abuse complaints diminish following arrest; we might easily generalize this result to the conclusion that married women will suffer less abuse if their husbands are arrested. But this is not precisely what the data tell us, and the conclusion may not follow. Suppose, for example, the experimental subjects include many

live-in boyfriends. If arrest breaks up informal relationships leading to zero rearrest rates but increases violence slightly where relationships do not break up predominately among the married, it may be that an overall decrease in subsequent assaults hides an increase in assaults by married men. Or it may be that arrest deters most men - boyfriends or spouses - from subsequent assaults but incites a few men to greater violence. What we think we have learned - that arrest deters spouse assaults - is part of a more complex story. Since we did not measure the degree of violence in the subset of arrested men who offended again we may miss an increase in the seriousness of their violence attributable to the arrest experience.

The policy maker must, of course, be concerned with this dimension. How should one tradeoff saving many women from battering with subjecting a few women to very serious beatings or death? One must look closely at the data reported to learn what possibilities were considered. If the experimental design did not seek to evaluate a particular treatment or to measure a possible outcome the experiment will not directly address the implications of the treatment or the incidence of the outcome. Sound theory or common sense may, however, allow the policy maker to generalize from the experimental results to other matters of interest.

Now let us return to the Minneapolis experiment and see how we and the experimenters can go about answering our three questions. To begin at the very beginning, we must first have an issue that concerns us and we must, as Dick Berk said, be uncertain about what we know. It is no accident that the Minneapolis experiment was not done in 1971-72 but instead waited until 1981-82. This is not only because we may have been less concerned about spouse assaults in 1971 than in 1981 - many people, especially the police who had to deal with them - were very concerned. It

is also because in 1971 we thought we knew the answer to the problem - counseling. Thus many police forces sought to train officers in counseling or established special spouse assault units. We came, however, to realize that counseling was not the answer and to wonder what the answer might be. Hence the experiment.

The experiment was very simple (social scientists call such simplicity "elegance") in design. Participating officers would treat abusing spouses involved in misdemeanor (not felony) assaults in one of three ways: by arrest, by separating them from their homes for 8 hours, or by advising the couple with an attempt to mediate the dispute. These treatments were randomly assigned because each officer was, except in certain specified situations such as if her were attacked, supposed to treat the assailant in the way specified in a color coded form book with forms placed in a random order (this means, incidentally that an officer might find that in four consecutive cases he was to advise; random distributions often have this counterintuitive type of clustering). It is in the administering of the treatments that we have an important source of possible slippage - a threat to internal validity. If the officers ignored the forms, treatments would not be allocated in a random fashion and and an analysis that assumed random assignment would be misleading.

Sure enough the experimental design was compromised. While 99% of the subjects who were supposed to be arrested were arrested; only 78% of those supposed to be advised received advice and only 73% of those who should have been separated for 8 hours were separated. In almost all cases those who didn't receive the scheduled treatment were arrested. It turns out that this was fortunate; if there was to be any failure in the administration of treatments, better this direction than any other. The

reason is that arrest proved to be the most effective treatment. Yet one would expect the arrest group to contain a disproportionate number of bad guys or poor risks - i.e. those whose behavior was so serious, such as those who attempted to assault the police, that the police ignored their contrary instructions and arrested them. These people swelled the arrest group, and one would expect them to present worse than average risks of recidivism. It the treatments had been delivered as administered, one would expect the arrested group to have looked even better. On the other hand, had the arrested group done no better than the other groups the failure of randomization, that is, an internal validity problem, would have precluded the ready conclusion that arrest was no more effective than the other treatments, although a variety of statistical techniques might have helped us evaluate this possibility. Indeed, despite the favorable direction of the slippage the Minneapolis study experimenters, Larry Sherman and Dick Berk, engaged in some further statistical analysis to ensure us that the implementation failure did not affect the results, but the details need not concern us.

An important point emerges from this discussion. Experiments, or any research design for that matter, amy be informative although they fall short of perfection. We must evaluate the failures of design or implementation and look closely at their implications. Common sense and knowledge of the problem are at least as important in this enterprise as sophisticated statistical knowledge.

Now let us consider the results of the Minneapolis experiment. The basic result, taken from police records of reported repeat violence for the 314 cases involved in the experiment, is that 10% of those arrested became official recidivists, as did 19% of those advised, and 24%

of those separated. The difference between the separated and arrested group (24% and 10%) is statistically significant, which means that by generally accepted scientific conventions we can be reasonably confident that the difference is not a chance result of our random assignment of treatments. The difference between the advise and arrest conditions is not statistically significant, which does not necessarily mean it is not a real or reliable difference, but it does mean that by the same scientific conventions we are not sufficiently confident that the difference reflects the effects of the two treatments to claim that we have spotted a real treatment effect.

There are several things to note about these figures. First, note the difference between the best and the worst treatment if 14%, (24%-10%) and that 10% of the arrested men later commit spouse assaults. Thus arrest is not a perfect treatment, and separation may not be a bad treatment. We lack the untreated control group needed to make the latter judgment. Importantly, if we arrest in every case and these figures hold, we will be delivering the most severe of the three treatments, arrest, to many men who don't need it. In particular, if our only goal is to prevent future spouse assaults and we arrest in every case, in only 14 out of every hundred cases will it matter. That is 10 out of every hundred men arrested will go on to beat their spouses again despite being arrested, and 74 out of every hundred would not have beaten their spouses had they been merely separated. When, as in Minneapolis, arrests do not lead to convictions this disjunction might not be of too great a concern to the policy maker. But suppose arrested men were routinely imprisoned for 3 months or were put on supervised probation. We should recognize - if these statistics still held after conviction - that 84 cents of every

dollar we spent in arresting, convicting and sanctioning these men was wasted if our sole concern was to protect women from future assaults. The experiment can give us an idea of trade-offs involved, but its results cannot make the policy decision for us.

The third thing to note about these figures is that they are open to a different interpretation. It may be that arrestees are unlikely to again (i.e. within six months) be the subject of police reports because their wives although beaten again do not again call the police. These women may be too terrified to call - perhaps because the prior arrest was followed by a worse beating - or they may be unwilling to call - perhaps because the family suffered, maybe through lost wages from the man's first arrest. Thus arrest might actually be the worst rather than the best of the three treatments.

The experimenters anticipated this ambiguity, and included in their original design interviews with the victims. Of the women who had originally called the police, 161 responded to a full series of 12 follow up interviews over a period of 6 months. Nineteen percent of the arrested group reported being assaulted again, 33% of the separated group and 37% of those advised. The difference between the advised and the arrested group is statistically significant although it was not in the police report data, and the difference between the separated and arrested group - again unlike the official reports - is not. There are problems with these data: the low completion rates for the twelve interviews, and the reversal of the relative effectiveness of the two non-arrest treatments and of their statistical significance is troubling; but I shall not dwell on these problems. The fact that these data confirm the relative effectiveness of the arrest treatment is most important. Not surprisingly

the proportion of men who were repeat offenders is higher than it appeared from the official reports. The "unnecessary arrest rate" (this time involving the arrest-advice comparison) remains at 86%.

Two more general points also emerge from these data. First, a very serious threat to our ability to make sense of the official data was anticipated by the experimenters - thus allowing us to have greater confidence in the efficacy of arrest. Had the ambiguity in official data not been anticipated, we might have done the experiment and not known what to make of the results. Second, if the victim interviews are representative of the entire sample, our fear that treatments affect future reporting may be well founded. While about the same proportion - 51 to 52% of all those who saw their spouses arrested or received advice and were later beaten later again called the police; 72% of those who saw their husband/lovers taken away but not arrested were if beaten, again willing to call the police. Thus if the victim data reflect the experience of all women in the sample, in opting for an arrest rather than a separation policy, we may be opting for a 20% fall out in the willingness of battered women to again call the police.

I will deal with external validity problems only very briefly. I am sure it is obvious why what works in Minneapolis may not work or may not work in the same way or to the same degree in other cities. These reasons include the different ethnic makeup of different cities, (16% of the accused abusers in Minneapolis were native Americans); the quality and reputation of the Minneapolis police force, the weather, and perhaps most importantly the treatment that follows an arrest. In Minneapolis 43% of those arrested were released from jail within one day, 86% within one week, and virtually no one was convicted on an assault charge. The costs

and effects of arrest in a city where arrest was often followed by conviction might be quite different. The general point is that when we intervene in one segment of the law enforcement system, as when we change police policy, we must consider how other aspects of the law enforcement system, like the prosecutor's office, will be acting. Even if there are substantial similarities in two cities and similarities in the unit (e.g., the police force) that is to implement the policy change, effects in the two cities may be different if the prosecutors office, the courts or the corrections's systems in the two cities differ in important respects.

Now let me talk about the third question: precisely what did the Minneapolis experiment tell us; what did it measure and what have we learned. Time forces me to be very selective. First, the advantages of arrest over the advice/mediation treatment compares arrest with a rather unclear and probably inconsistent type of advice. It may be that some kinds of advice or mediation programs are as or more effective than arrest even though the varied, unmonitored police advising given to victims in the Minneapolis experiment was not. Second, note that the experiment as described thus far tells us what works- not how it works. Now knowing how something which works is not always necessary for wise policy implementation - the observation that vaccination with cowpox prevent small pox preceded by many years knowledge about why there was a preventive effect - but usually knowing how something works is important to wise policy making.

The data collected in this study provide some examples of the importance of knowing not just whether but also why a treatment works. For example, 32% of those Minneapolis men who were arrested and released took a week or more to return to their spouses or never returned, while

only 10% of those who were separated for eight hours were not back with their spouses/lovers within a week. If arrest works largely by breaking up relationships - and I am not saying the Minneapolis data show this to be the case - we may doubt whether arresting when the women doesn't want it - perhaps because she fears it will break up the relationship - is a good idea. At a minimum it is paternalistic - deciding for the victim what is in her interest.

Second, and most importantly, figures based on victim interviews show that arrest coupled with police attention to the victim leads to a 9% recidivism rate while arrests by police who do not listen to the victims leads to a 26% recidivism rate. While the authors do not provide measures of statistical significance in the published reports, it is likely that arrest without attention to the victim is not significantly more effective than either of the other two treatments. In other words, there were actually four conditions in the Minneapolis experiment; arrest only, arrest with attention, advice and separation.

The only effective treatment was arrest with attention; the apparent effectiveness of arrest reflects a lumping together of the first two treatments under this label. What the authors are measuring when they speak of arrest is thus not a pure arrest effect - but the effect of a treatment that combines arrest with attention to the victim. If the numerous police departments that adopted a "routinely arrest" policy following the Minneapolis experiment did it not knowing that listening to the victim was as important as arresting - and if they did this in response to media reports they might never have known or lost sight of this fact - and if the police in these departments are not in the habit of listening to the victim, they may have adopted a policy which will have harmed some men and helped no one.

Indeed a "routine arrest" policy may reduce listening to the victim. For example, a department which arrested occasionally might have listened closely to the victim to decide whether to arrest. The implementation of a "routinely arrest" policy may have made listening to the victim, appear less important to the police. The tragedy will be all the greater because by attending more closely to the data, the police departments that borrowed from Minneapolis might have devised a treatment that really worked. The general lesson is that we must attend carefully to the concepts employed in experimentation, and we should try to "unpack" them to the extent feasible.

Where does this leave one on the issue of how or whether to, implement a spouse abuse arrest program? From a scientific standpoint there remains doubt about the efficacy of arrest, particularly of arrests that are not accompanied by careful attention to the victim, and particularly arrests in cities other than Minneapolis, There is also substantial uncertainty with respect to the costs of a "routinely arrest" policy in terms of unnecessary arrests and the public expenditures, lost jobs and broken relationships they may entail. We also don't know whether a small portion of arrested men are more violent and hence more dangerous to women because they were arrested.

There is, in short, a lot that remains to be learned - and much of it is important. Yet we have learned a lot, and this brings me to a point that Larry Sherman and Dick Berk have made before me. The status quo of limited or no arrest is also a treatment. It is a treatment that seems less wise than it did before NIJ sponsored the Sherman-Berk experiment. Wisdom today appears to call for a greater frequency of arrest than it once did, a greater frequency coupled with careful

attention to the complainant when an arrest is made. It does not necessarily call for increased prosecution of those arrested, but we do not know that this would be without effect.

Some may wonder how I can favor the greater use of arrest given what I've said about such potential effects of arrest as the stimulation of brutal violence in a small number of cases or the discouragement of repeat calls to the police. The first possibility was not ruled out by the experiment and the second is suggested by some of the data the experimenters report. With respect to the possible brutalization effects of arrest, it is important to note that the mere fact that an experiment does not rule out a possible, plausible outcome does not mean the outcome is likely. While the brutalization hypothesis is plausible, I believe it is more plausible to suppose that in diminishing the absolute number of repeat assaults we are also diminishing the number that would escalate into particularly violent attacks. In similar fashion one must often fill in gaps in experimentally acquired knowledge by reference to scientific theories or common sense. I do believe that arrest - as opposed to separation, for example - will discourage repeat calls to the police, but I regard this as tolerable in that the level of repeat violence is also depressed. Moreover, it might be possible to couple arrest with the kind of attention to victims that encourages repeat calls if assaultive behavior should reoccur.

To be very specific, if I were a police chief, the results of the Minneapolis experiment would have led me to a policy of presumptive arrest in misdemeanor spouse assault cases. However, I would instruct officers to listen carefully to the woman before and after making the arrest. I would probably not encourage arrest where the spouse did not want it, for

this responsiveness to the complainant might lead her to feel empowered. I would also prefer separation to arrest, where the man had a job that might be threatened by an arrest, for it appears that male unemployment is an important factor contributing to spouse abuse. An arrest which results in a job loss is, I believe, likely to enhance prospects for future familial violence more than it suppresses it. Finally, I don't know what I would do if the jurisdiction I was in routinely punished and prosecuted arrested spouse abusers. The Minneapolis experiment tells us nothing about the relative costs and benefits in this situation, and I do not believe I know enough to make even a good guess.

In addition, there is one other thing I would do. I would monitor closely whatever changes in policy I as a chief ordered. Measuring the effects of policy changes should not stop when the experiment which stimulates the change concludes. While post-experimental changes are unlikely to be instituted experimentally, a lot may be learned from them regardless. Indeed, if the 150 departments that responded to the reports of the Minneapolis experiment by implementing a "routinely arrest" policy monitored spouse abuse we are likely to learn more about the deterrent effects of arrest from the combined experience of these departments than we can from the Minneapolis experiment. If the Minneapolis results generalize and if 150 police departments have already changed their spouse assault arrest policies in response to the experiment, a failure to see a marked diminution in repeat spouse abuse complaints during the past year will require some explanation.

Thus, I ultimately conclude where researchers often seem to conclude, wisdom may counsel certain policy changes, but it also counsels seeking greater wisdom. We should strive to use experimental techniques

to measure the effects of criminal innovations, and when innovations have been generally implemented, we should, for a few years after the implementation at least, use quasi-experimental and other techniques to monitor the effects of implementation. Both the early experimentation and the later evaluation require teamwork between researchers, practitioners and granting agencies of the kind the Minneapolis experiment exemplifies. It is this kind of teamwork we hope this workshop will stimulate. Thank you.

(Applause.)

PROF. REISS: We now have three speakers who are going to assess the role of field experiments in criminal justice, and while we were all given a note about the goal of this, I think that speakers have taken as their topic to somehow create here for us a forum in which both the policy makers and research sponsors can exchange some views on field experiments, and so, we are going to begin first by hearing from Mac MacDonald who is President of the American Probation and Parole Association.

MR. MAC DONALD: Thank you, Al.

You are on the way to success. I don't wear a watch. So, you can call the time. Last week I had the pleasure to speak to a meeting of the Regional Criminal Justice Planners on behalf of the Bureau of Justice Assistance, and I was introduced similarly. "Keep it real short." I am getting used to that introduction.

My message is a little bit longer in probation and parole, but that day will come. How many of you have seen the movie The Mission?

(There was a show of hands.)

MR. MAC DONALD: A handful of people. How many people hate people who tell you how a movie goes before you have seen it?

(There was a show of hands.)

MR. MAC DONALD: If you have seen THE MISSION you are aware that there are three main characters. One of the characters is very committed to an ideal and does not survive as a living person because of that commitment. Another character is committed to an ideal but is very practical in attempting to see that ideal maintained. This person also dies. The third character seems to compromise. He seems to blend the practical with the ideal. It is this blending which seems to survive and give at lease some continuation to the ideal.

Our discussions have presented the ideals of research and the practicality of operationalizing criminal justice. I suggest that both researchers and practitioners have to blend their values in order to see our mutual value of improved criminal justice service delivery survive.

An example of this blending occurs in research conducted by the Rand Corporation. This research addresses the effectiveness of probation in Los Angeles and Oakland. When released it caused a dialogue throughout the country, media, and in particular throughout probation. In California this research was used by probation administrators in instructing their funding authorities and in attempting to secure additional resources. After all, it is reasonable to think a poorly funded system will lack effectiveness. In other states probation administrators responded practically by conducting similar research to see if their systems had similar problems and if so, if additional resources would solve the problems.

Approximately a year after "Granting Felons Probation" was released, an additional research monograph was developed by the Rand Corporation. It was called "Prison versus Probation in California." It

seems to indicate that ex-prisoners with the same profile as some probationers have a higher recidivism rate. This research was supportive of probation and useful to probation administrators even though it did not receive national attention.

Neither research monograph was treated as gospel truth. What this research did for probation was the creation of a national dialogue and new thinking by probation administrators. We examined our effectiveness under a new light. We rethought the allocation of resources for probation as an alternative to incarceration programs. We analyzed our goals to see what progress we were making to achieve them.

I am pleased to say that today as we get into research on electronic monitoring supported by the National Institute of Justice, we are at the front end of looking at an intervention tool available to a probation officer in supervising offenders. The outcome of this research should assist probation administrators in their decision making process concerning the implementation of this technology. It will provide information concerning appropriate target populations, workload considerations, technology capability, and other managerial issues.

The Bureau of Justice Assistance is currently funding research on intensive supervision (ISP) in California, Oregon, and Wisconsin. Joan Petersilia of the Rand Corporation is working on this project with judges who have agreed to random selection of offenders for diversion into ISP or placement into institutions. The outcome of this research will be valuable to probation administrators in refining and expanding ISP's as the political emphasis on alternatives to prison increases.

Both the electronic monitoring and the intensive supervision research are examples of quality research which can be blended with practical application--the ingredients for success in achieving our ultimate goal of an improved criminal justice system.

Going beyond past and current research, I would like to suggest that there are three themes which should be addressed in all future probation research in order to increase its practicality for probation administrators. These are: a) offender population criminal and personality profile; b) management and objectives; and c) officer intervention strategy. We may think of these three themes comprising the corners of a triangle.

First, let me discuss the theme of offender profile. Throughout the country probation officer assessment of the offender is becoming more sophisticated. Through research conducted in Wisconsin, case classification instruments were developed and are now being adopted nationwide. Probation officers through risk and needs assessment instruments, are profiling the offenders to determine how best to manage the risk they pose to society and themselves. Through the strategies for case supervision or the client management classification system, probation officers are categorizing five personality types requiring different goals. officer/probationer relationships, use of auxiliary resources, supervision techniques, and treatment: selective intervention--situational; selective intervention--treatment; casework/control; environmental structuring; and limit setting.

This assessment process is not bounded by offense type. Categorizations by offense type ignore risk and needs and differentiated supervision strategies, and ultimately reduce effective supervision.

The second corner of the triangle is management and objectives. How was the program managed? Was it supported by the administration? Were the objectives fully understood by the line officers delivering the service? I know administrators and line officers mandated to implement alternatives to incarceration yet they believe these offenders should be in prison. Such contradictions would be surfaced in research when a program is being analyzed for its effectiveness.

The third corner of the triangle is the officer intervention. I will use electronic monitoring as an example. Electronic monitoring technology is a tool for the probation officer. If it is to be demonstrated as effective in diverting persons from incarceration, research would have to address questions regarding officer intervention with the offender on electronic monitoring. Did the officer instruct the offender on how to use his "leisure" or "incapacitation" time at home productively? Did the officer merely conduct observation and surveillance uncomplemented by counseling? Depending upon the intervention strategies used by the officer, very different outcomes can occur and be attributed incorrectly to the technology.

If an intensive supervision program is being researched for its effectiveness, the component parts of the program would have to be studied in order to determine why there was success. Did the community services restitution make the impact? Were the number of face to face contacts significant in controlling behavior? Did referral to treatment take place and was it effective? ISP programs throughout the nation vary in service delivery strategies. I assume no research will indicate ISP works, but rather certain components of ISP programs seem to produce desired outcomes.

The triangle is complete--offender profile, management and objectives, and officer intervention. Research which examines all of these and their interaction will be most practical for probation administrators and have the best potential for generalization and replication.

Future topics for research in community corrections which I suggest to you include: residential community correctional services; boot camps for probationers; case classification instrument refinements; surveillance functions; alternatives to externally imposed control; education about thought patterns and value systems; and unique populations such as Hispanic offenders and child abusers.

I recognize that even in fields of research which appear to lend themselves to more objectivity than ours, such as accounting, research outcomes are implemented only when the information is politically acceptable. At the American Probation and Parole Association, the Research Committee is mandated to be APPA's liaison with the research community and work towards having the research outcomes accepted by practitioners. This liaison and your involvement with APPA should foster the political process which can translate the research ideals into practical public policy.

Thank you.

(Applause.)

PROF. REISS: I find that I sometimes neglect the duties which are cast upon a moderator, and one of them is to make some announcements. I suddenly realized that sitting in front of me was a piece of paper which had two announcements on it which I should have made at the beginning of this, but in any case, one of the announcements is to remind you, again

that these plenary sessions and that some of the discussion groups have been taped or are being taped and that the tapes are for sale and that there are order forms outside. So, if you are leaving before the end of this session and want to order, be sure to do that. We hope you will stay on, however, until the end of the session.

Secondly, there will be a meeting of the working group at the podium here at the close of this meeting.

We turn now to the second of our speakers. In introducing him, I simply want to say that every once in a while there is a new kid on the block, and Dave Tevelin is a new kid on the block, but he is the rich kid on the block or at least I understand he is reasonably rich, and he may even become affluent. He welcomes coming to our block, since he came from a rather tough neighborhood. The tough neighborhood was called the United States Sentencing Commission, and he is probably happy to have moved into this new arena seeking asylum.

Dave?

MR. TEVELIN: Thanks for taking my first two jokes, Al. I appreciate that, but I am happy to be here, believe me, in more ways than one. I don't think I am quite as happy as Judge Beaudin who equated his appearance here with being in the dugout of the Boston Red Sox with his boyhood idols. I had the privilege of serving as security in the Oldtimers dugout, at the Oldtimers game in Washington the last three years, and this is a notch or two below that. So, please bear with me though, I am happy to be here.

I am the new kid on the block. I am holding on to my wallet here, just in case anybody wants to find out how rich I am temporarily, but I will tell you exactly what we have in store for you in the community

of researchers and judges that are interested in the quality of justice in the state courts. I think a better way, maybe, to characterize it is we are not just the new kid on the block; in sprinter's terms, we are the new kids out of the blocks because we really have hit the ground running, and have put together a program guideline which was distributed here. It appeared in the Federal Register on Monday, and what I would like to do is talk a little bit about it generally to familiarize you with what the Institute is all about and then delve into the area of experiments specifically.

We are a non-profit organization that was chartered and funded by Congress in 1984. I think the closest legal model to us in terms of a structure would be something like the Legal Services Corporation or the Corporation for Public Broadcasting. We are to conduct our activities like a non-profit organization and in fact, are eligible for 501(C)(3) treatment. So, we could go out and even get private monies in the future.

Congress appropriated us approximately $6.7 million for Fiscal Year 1987, to hand out in the form of grants, contracts and cooperative agreements. It is no-year money which in the parlance of the grant game means we don't have to spend it by September 30, 1987, and that has permitted us the luxury of coming up with we hope, what will be an orderly, fair, rational grants process. We do not have to just dump the money on the street to show some sort of accomplishment for Congress.

We are asking for almost $13 million next year, $12 million of which will be dedicated to grants and programs. We have a very receptive Congress. The Administration is slightly less receptive to our existence, but I think that this is probably an opportune moment for me to at least mention that we don't see ourselves, and I think Chips feels the same way;

we don't see ourselves as competitors of the National Institute of Justice. We see ourselves as complementing, cooperating, trying to work together to present programs that will benefit the courts generally. I would like to think that it is a colleague relationship, and we have already engaged in a lot of communication and are seeking to explore things we can do together.

We have some aspects of our authority that the National Institute of Justice doesn't have, and NIJ can do some things that we cannot do and has already established some very firm liaisons in the research community and with others, and we want to take advantage of that. So, I think there is a reason to be mutually receptive to each other, and we intend to do that. Let me summarize what the statute lets us do.

We have a statute that authorizes us to fund programs in 15 broad areas, and I will not list all 15, but I will list those areas that the Board has designated as quote, unquote, special interest areas. These are areas that will be given a slight preference in the funding process, and one thing I want to caution you about right up front is that this is not the type of program that LEAA was which is where the bureaucrats in Washington are going to sit there and tell the state courts what they need. We have designated things as special interest programs but we have not given them "Priority status," in that these are the only things we are going to fund. If the concept papers and the applications that come in to us demonstrate that the courts are interested in other areas, we will give them the same treatment as we have designated to these other areas as special interest and give those a preference in the application process, also.

So, with that being said, let me just run through the list very quickly of what are identified as special interest areas in the guidelines. The first one is the education and training of judges and other court personnel, the use of technology to improve court management, reduction of expense and delay in litigation, alternate dispute resolution, improved jury management, fairer treatment of victims and witnesses in both criminal and civil cases, improved handling of domestic violence cases, improved procedures for imposing, collecting and enforcing fines and other financial penalties or obligations, the development of innovative measures to encourage and enhance judicial careers other than increasing salary -- we didn't think we needed a grant program for that -- and research to help develop creative ideas and procedures to improve the administration of justice in the state courts and at the same time reduce the workload of the Federal courts.

This would be examining things like habeas corpus review by the federal courts of state convictions, diversity jurisdictions, multistate litigation, things like that, and we are interested in looking at some good seed money kinds of innovative creative thinking to lead to larger projects in that area down the road.

We are going to have two rounds of funding for this fiscal year. The first one starts with the submission of concept papers April 17. That will turn around and eventually lead to applications being due July 24, from those whom the board invites applications from, and we would hope to award money for round one in August or September.

Round two will start shortly after that. There will be a substantial amount of money reserved for the second round, the idea being we don't want everybody just to think what they can think of in five weeks and be afraid that the money is going to be gone.

There is another good reason for that delay, and it is one of the special things in the statute that I want to mention. There is a provision in the Act that if a state or local court applies for money from us they have to have the application endorsed by the state supreme court or its judicial council. So, that is an extra hoop that the local courts have to go through, and when you add that to this very tight time frame for round one, it it apparent that they are not going to make it. So, we do anticipate getting a lot of applications from the state courts in round two and perhaps more from the judicial organizations themselves and researchers, universities in round one, but we will see.

A few other provisions that you really should be aware of: One is that the statute accords priority to the State courts and to the local courts as recipients, as well as to the national judicial organizations that serve them and are served by them.

So, for instance, the National Center for State Courts, the Council of Family and Juvenile Judges, the Judicial College could be included in that priority category.

Then there is a second category for non-profits, law schools and other private agencies, and this is in the words of the statute: "If the objectives of the program can be better served thereby." That is a term of art that we will elaborate on more fully as we see what the applications and the process look like. We have not really come to grips with that issue too closely yet. Then the third category of priorities is any other governmental agency, if the objective of the program cannot be achieved by either of the first two groups, basically.

It is a 50 percent match program for the State and local courts, and we are giving a preference to hard match, that is cash match, but it is not an ironclad preference. For non-profits and other organizations,

there is no match requirement, and even in the 50 percent case it can be waived by the State supreme court and a vote of the majority of our board. So, it is possible to get in for less match.

With respect to experiments, this has been a tremendous learning experience for me. I am not a social scientist. I know many of you couldn't tell from talking to me, and I am much less a director. It has only been 6 weeks.

What I am by training is a lawyer, and I have had some familiarity with research projects carried on by NIJ and other agencies in the Justice Department because I was in the General Counsel's Office in the agency that served NIJ for a number of years. What I am going to tell you, and I am lucky in this regard, is that these are my personal views. In listening to what has happened here in the last two days, I really am not in a position to commit my Board which in all candor has not come to grips with the issue of experiments in criminal justice and how that fits into our funding scheme. But I think I can give you some insights into how they are going to look at this whole area and tell you about some of the reefs that are out there that any of these kinds of proposals are going to have to cross.

The first thing that I ought to bring to your attention is that methodology is not this Board's primary concern. This Board is much less concerned with how we get to this point than what is the point we are getting to. They are looking for action-oriented projects. They are looking for research that has a quick, visible payoff that will result in change not only in the jurisdiction under study or where the project is but that can be replicated nationally and that respond to those needs that the state courts articulate that appear to have some national

significance. Any of the programs that we are going to fund, the things that are defined as special interest projects are, by definition things that respond to the most serious needs articulated by the State courts and that have the capability of being transferred to other jurisdictions. There is a limited amount of money to go around, and we want to get the most bang for the buck. So, I can tell you for sure of the Board's interest in that area.

I do not perceive at this point a Board bias or a likely Board bias either for or against experiments, but I would caution in line with what I just said that with very limited money available at this time and the risk of failure that was noted by Professor Boruch and others here that as much as one in five have a poor implementation design and go down the tubes for that reason, when you are trying to persuade a group of lawyers and judges that this is the way to spend the very limited money you have left, that is going to be a hurdle to cross.

They are looking for innovative things to do, but they are also looking for proven programs that have not been transferred to other jurisdictions either because of lack of knowledge and the other jurisdiction's lack of money available for technical assistance. I should have mentioned, by the way, that the last special interest area is technical assistance transfer programs in any one of those previous 10. So, the Board has put a heavy emphasis on this idea of replicability and marketing effective programs, but given those qualifiers, I think that if an applicant can demonstrate that the experiment is well designed and has strong support of the court in that jurisdiction that the board is going to give it a very, very serious look.

There are people on this Board who are practitioners, and there are, also, people who have designed projects, are legal scholars and researchers in their own right, or at least have it in their background and can appreciate what a well-designed experiment looks like. They, also want to be real careful that whatever they do is for the State courts, not to the State courts; they want that court to buy into the project, and if there is any suspicion that the court is an unwilling victim, an unwilling participant, it is not going to sail with this Board.

Now, there are certain special interest areas that I listed that appear to me to have some possibility for experimental research, more than others, and I will just list them, and I am sure in your own imagination you can come up with more than I can, but the testing of alternative dispute resolution approaches and techniques, jury selection procedures; delay reduction techniques; fines, imposition, collection and enforcement; and others that were identified in the courts workshop: What kind of drug treatment should be given to certain offenders; which kind of offenders who are child abusers would benefit from treatment which would benefit from incapacitation?

So, I think it is a fertile field. It is an open area. This Board is not going to be looking at experiments for experimentation's sake. It is going to be looking at experiments that will produce the kinds of programs that will have some impact of the kind I discussed, and they will be particularly interested in concept papers and applications that present proposals for experiments that will either demonstrate the success or the failures of programs in those key areas of special interest. So, I would just direct your attention to those areas and really try to encourage you to put your applications in on that process.

Any questions about what we are and what we are going to do?

Write your Congressman. Thank you.

(Applause.)

PROF. REISS: Our last speaker needs no introduction to you, since he opened this forum, but in presenting him here this afternoon, I entered into a contract with him. Some of you will remember the comic strip that Al Capp wrote for years called Little Abner, and one of the characters in that comic strip was called Marrying Sam, and Marrying Sam, among other things that he did very well was to perform wedding ceremonies and ceremonies of various kinds, and he had several prices, however, for them. One of them was the 50 cent wedding, and they went up, and the top was the $2 wedding, and the $2 wedding was one in which everybody, including Marrying Sam ended up in a trance at the end, and I have asked Chips to give us the $2 wedding this afternoon in 50 cents worth of time.

MR. STEWART: Thank you, Al. I did feel a little bit like Marrying Sam, the great matchmaker, because I was around throughout the entire workshop. There were so many people who got so excited, I felt the passion running quickly in this hotel corridors here, and I tried to match up the likely candidates with the likely recipients, and I think we have done a good job.

I have heard an awful lot of people who have said, "You know, I haven't thought much about this experiment, but it sounds like a great idea, and I have found a researcher that I sort of like. What do you think?" I would sort of give them the 50-cent wedding, and they are still out there talking in the hall, I hope or I hope that is the case.

A couple of quick comments that I would like to make. First off, I am delighted that I am sharing this panel with Malcolm MacDonald from the American Parole and Probation Association because I think that there is an awful lot of very important experimentation that could be and ought to be done, and we are very pleased that they have their own research committee.

I want to say, also, that we encourage everybody to get out and start doing some research, but we have to say that it is not easy to do, and it has to be done in a very effective way, and we are all looking for the magic bullet. I am very pleased that the practitioners have taken the bit and are beginning to do, I think, some very effective work.

I would like to say that when Malcolm, however, says that we have to find more alternatives to prison, I want to caution all of you out there that we may have plenty of alternatives to prison right now. Seventy percent of the people who are convicted of felony crimes usually are addressed to Malcolm's organization, and go through the county jail and are back out on the streets, back in the community under supervision. Only 30 percent usually make it to our state prisons and of that time much of the time that they spend in there continues to dwindle more because there is more pressure to find more alternatives. The question is that politically we have a statement that we have to make about danger, relative risk. I think that is the thing that Malcolm and I and others have joined hands on, working together to try to find the best way around, and I think that this is the kind of thing between us all that we are going to do quite well.

Let me, also, say that the Bureau of Justice Assistance has some money to manage, $225 million for anti-drug abuse, and we are very fortunate that Dave Tevelin is here representing the State Justice Institute with several millions of dollars to help improve state court management and the effectiveness of the courts, and there are other organizations that are growing up. The National Institute of Justice is sort of the father and the premier of which Bureau of Justice Assistance and BJS and others have grown out of, and we are pleased to welcome our family as it continues to increase. We are delighted that we are able to work with them in so many different areas, and we continue to be the premier and the principal agency of research in the Department of Justice. We are engaged in some high-risk experimental stuff that other people probably would not take on, and that is the reason we are, also, very grateful to the National Academy of Sciences for their very fine work that is continuing to look at the tough questions that we all face.

Let me say, also, that I am particularly grateful to all of you who are still here. It is an amazing tribute, either to your dedication or to the numbness that we are all facing. My dad was a great orator and speaker to which I probably could never aspire, but he always told a joke, and he said that as the speakers continued to impart great wisdom to everybody in the audience that the audience began to leave with their bags and things like that, and there was one fellow still sitting at the table, and the speaker finally stopped, and he said, "I want to thank you very much for staying to listen to my final remarks, and I am so grateful to you." He said, "Don't thank me. I am the next speaker."

(Laughter.)

MR. STEWART: In terms of my gratitude though, it goes to the practitioners and the researchers, you see, because without you there wouldn't be any experiments. There wouldn't be a reason for a conference like this, and there wouldn't be a reason for an NIJ. You see, because the NIJ, I was introduced the other day as a bag man, you know, and everybody thought that was great, except me, right? The National Institute of Justice really comes up with some good ideas. They nurture them. They decide about a plan. They work hard to try to get that and try to find out those marriages between researchers and practitioners. I want to say that we are not dollars with no place to go. You see, if we want to throw money at a problem, we have got to find quality people who are willing to first off develop and devote their lives to designing experiments, research, and surveys. We, also, have to look for practitioners who are willing and have the courage to take a risk to try to find out whether something works or not. It may not work, and people hate poor performance, but in the area of research you have got to be able to address that.

Larry Sherman in his opening remarks talked about the 500,000 compounds that were looked at in the treatment of cancer, and they were all excited because they found 40 that might help or have some relationship to treating some forms of cancer. That is the kind of enthusiasm we need in criminal justice. The fact is that I want stuff that works every time. I have been jumping up and down on everybody to come up with this great idea, and fortunately, we win once in a while, and when we do, it saves lives and changes the way the policy makers look at and see their roles.

It, also, does the same thing for research because when we have a finding, it also, changes the researchers. I think that is very important to take a look at, and so my gratitude goes out to you, the researchers and the practitioners because without you, the dollars in the National Institute of Justice would be wasted. They would be essentially dollars with no place to go because like a rich person who can spend a check, they have got to have something to buy that is worth something, and your time, your dedication, and your lives are worth something, and they have made the dollars, the few dollars at NIJ go a far place and to really pay off. So, our gratitude goes directly to all of you.

In the past day and one-half we have met to discuss the role of field experiments in criminal justice policy making, and I think we have accomplished with great enthusiasm that goal. The discussions have been lively, and the debate sometimes intense. That is exactly the way I planned it, sometimes frustrating. I didn't plan it that way, but the intense debate I wanted. The perspective of each of us in this room and at this podium, I think differs to some degree on the value, the need and the contribution of randomized field experiments. The diversity of our viewpoints has made these discussions all the more profitable.

I would like to take this opportunity to thank Dr. Rick Lempert, Jeff Roth, Dr. Roth and Christy Visher from the National Academy of Sciences for dedication to a fine job and a tremendous conference, and Rick, and Jeff and Christy, we certainly have all appreciated all of the fine work that you have done and the careful planning that has gone into making what everybody says, "Geez, I cannot believe this is a successful conference." It is, and thank you very much. We really appreciate the work you have done.

Now, in reviewing the events of the past two days, there are four points that I think are worth emphasizing and then I'll open it up for discussion.

First, I think that we should remember that the availability of large numbers of field experiments is a recent phenomenon as is the use of these experiments in policy making. That parallels somewhat the vigor with which NIJ has come onto the scene and tried to engage in the confluence of both of these areas.

Second, we should recognize the difficulties inherent in the researcher-practitioner collaboration in implementing field experiments and the legal and ethical limitations to experimentation.

The third point is that we could benefit by expanding the range of opportunities for conducting field experiments. Lastly, I would like to suggest that experimentation plays a crucial role in determining whether or not the criminal justice policy making remains a craft, relying almost entirely on rules of thumb or whether it advances as a profession based on the rules of evidence.

In our discussion in the past two days, we may have overlooked one rather startling fact. The number of field experiments in criminal justice has grown astronomically. While we all sit here with new enthusiasm, we are really looking back that in the sixties the experiment was a rare and tenuous event.

One of the reasons is because nobody had any idea what caused crime, how to measure it and what the different variables were and how to set up tests of measurement procedures. So, I mean we had to do a lot of background to build an infrastructure that would let us do that. In the seventies, that age of great enlightenment after that turbulence of the

sixties, there were a few more experiments, but those were really anomalies. In the eighties, an era of tight resources, Experiments have changed the way policy makers begin to look at their problems. They also have changed the way that legislatures look at the problems, and they certainly have changed the way that criminal justice is looking at its problems.

I think that in the eighties every area of criminal justice policy making can claim scores of experiments completed and many more in progress, and for that I think you should all be justifiably proud and pleased, and I am particularly proud of the staff at the National Institute of Justice. They have made our job, all of us in this room, a job that has been worth doing and very stimulating because it has been them and the long hours that they have worked to be able to provide an Institute for all of you to come to. I hope that in the next dozen or so years that they continue to do extraordinarily well because they have made a difference in all of our lives, but as recently as last fall when NIJ started to work with the National Academy of Sciences on this workshop, few of us realized the true popularity of the field experiment.

Experiments described for you in the materials for this conference are really only a small sample of what has really been done. We would have had one of those giant books for you to carry back, and so, we relied upon scholarship to begin to cull out what we believed, what the scholars believed were probably of the most impact.

There remain a large number of field experiments that we still know little or nothing about, either because they never were published or they never were completed. Al Reiss was talking in his police experiments group earlier this morning. He talked about this idea of police

departments intervening on a crime in progress, and we got all of the strategy about rushing police to the scene on certain streets and setting out dragnets and so on and so forth. He said that it didn't cost any money because police do that anyway. So, we tried it out. He said, "You know what happened? It didn'[t make any difference. We never caught anybody." He said, "Yet we've done this time and time again. We believe this axiomatically, that this is the only way to do it, and yet we tried it, and it didn't work, and it didn't work, but it didn't get published either, and it had to wait for this conference for an idea like this to surface."

I think that is very important because there has been a chance to question these ideas.

Now, this increase in the number of field experiments has been associated with an increase in the use of research by policy makers. The most dramatic example, of course, is the Minneapolis domestic violence experiment which has profoundly influenced the police response to spousal assault. Our personal thanks, of course, goes to Larry Sherman and to Tony Bouza who was the man who took the chance, and he said that he takes chances every day. I certainly believe that, but he is one of those rare breeds that is willing to take a chance with rigor and back it up with documentation. Three years ago only 10 percent of the police departments in America in large urban jurisdictions had a policy favoring arrest as the appropriate response to spouse assault. To give a sense of the power of an experiment and the thirst for it out there, now up to 44 percent of the major police departments have changed their policies.

Now, much concern has been mentioned about that by the researchers because they say, "Wait a minute. We may not know what we have tested. We are not exactly sure what has worked." But the key, I think, is that the police and the policy makers are looking at something that works, something that works.

We may not think it works 100 percent, and we might think it works 70 percent, but we are glad to go before the city council. We are glad to go before the civil liabilities groups and the groups that say that you don't do anything with some shred of evidence that gives us some confidence that the policy that we select makes a difference on future violence, and I think that is why there has been that kind of enthusiasm.

I, also, think that other research, such as the court employment experiment in New York City that Sally just described and the numerous experimental studies on rehabilitation programs have had a substantial influence on policy making. I clearly think that rehabilitation is a goal that we have not attained in any shape or form but one which we have to continue to pursue because I think our greatest resource is our people. We need to try to recover those people in the best way so that it doesn't harm others, but I think that this substantial influence on policy making is very real.

Yet we are not at the level of the medical profession where new drugs, new medical procedures must be tested with randomized clinical trials before it is legal to introduce them into the market, but much progress has been made in the use and understanding of field experiments by the criminal justice policy makers. The growth in the use and the number of experiments is likely to continue. The interest that all of you have expressed by taking some time out of your busy schedules to attend this workshop is one clear indication of this.

Your participation on the weekend, particularly here in New Orleans speaks highly of your interest and commitment to considering field experiments, and it, also, testifies to your willingness to endure. The support and the involvement of the national organizations like American Probation and Parole, the State Justice Institute, the Police Foundation and others is another reason I believe that field experiments responsibly done will continue to expand in use.

NIJ is committed to encouraging the use of field experiments in all of its research programs, and last year one of my staff counted 20 field experiments that we funded in the month of September alone. This year our research program emphasizes our interest in field experiments and challenges the research and policy-making communities to find new and imaginative ways to work together on field experiments.

Let me just mention just quickly about our program plan that you may have had a chance to read, and in there it used to say that we are going to put 750,000 or 1 million dollars for police experiments. We are going to put 1 million dollars for court experiments. We will do the same for corrections and so on and so forth.

I thought that was the wrong way because what it did is it carved it up by special interests. Just as this conference integrates various disciplines, we all have a piece of the problem. Let us take a look at a problem like career criminals -- the National Academy of Sciences has done a fine project on that -- a problem like drugs in our society, a problem like victimization and then let us see how we can plan experiments that might require the use of the police or the courts or probation or parole as to how they contribute to the solution of that problem. Rather than saying, "Here, Mr. Police is $1 million; here Mr. Parole is $1 million, do

a better job," I am saying that what they ought to be doing is saying, "How do we solve this problem in our society?" And that is how we have changed our program plan to try to reflect that.

Now, the opportunities for field experiments I think are boundless, and the range of topics of vital importance to criminal justice professionals that could benefit from experimentation is virtually unlimited. This range includes many innovations, such as the electronic bracelets which are growing in popularity as effective ways to monitor the condition of house arrest. They, also, offer the opportunity for effective supervision of persons on parole and probation. They extend the capacity and the very fine abilities of our parole and probation people, and that is an important thing to talk about. How do you give the people who are human out there a chance to do a better job? NIJ is working in three jurisdictions, Indianapolis, San Diego, and Utah, to test the effectiveness of this new technology. These experiments take advantage of an existing pool of non-violent felons who have been unable to post bond after 2 weeks in detention. A random half -- and everybody knows what that means now,, right? -- a random half will be offered the electronic monitor and released and another random half will be released without the monitor, and we are going to try to see whether the parole and probation people have more effectiveness with getting people to commit fewer crimes. I mean that really is the bottom line. We are all in this business together, and that is the protection of the innocent in our society. That is the most important goal of all of us. One of the ways that we might be able to do that is to keep people from hurting one another or stealing from one another and letting those people who may be the predators try to cut down on the motivation to be a predator and to

move back to that opportunity to where they can live whole and productive lives.

Ultimately we need experimentation because it simply costs too much not to be effective in criminal justice. It costs too much to adopt new technologies and new policies without the rigorous testing possible with randomized field experiments.

You know when people say that we shouldn't spend very much money on criminal justice because it is sort of a gloom and doom area; we ought to spend more money on parks, more money on education, more money on welfare, my answer is if we spent more money on thorough research in criminal justice, we would have more money for our parks and our welfare and our education systems. By not making that investment in the future, we jeopardize all that our education systems, our parks, our welfare and our housing offer to the poor and the middle class in America.

So, I am very high on research, but I think we have got to take it out to the field to communicate that. It, also, costs too much to follow the traditional approaches without analysis and testing to see if they work. I was a philosophy major in the university, and one of the things they said was that the unexamined life is not worth living. I think the unexamined profession is not worth professing.

For instance, in the 911 system which Tony talked about in New York City and which exists in virtually every large jurisdiction in this country, the intention was to get the police and emergency services to help people. What is the consequence? Millions of dollars, right? The consequence is a coup d'etat in police management, the fact that you have an incredible hemorrhage of your police resources going to calls in the field of which the caller has no idea of what their priority is, and where

does it give an opportunity for police management. So, here is a good idea that may have gone dramatically awry, and I think that hopefully we should invest in some research on this.

Another area is that just 5 percent of the identifiable addresses in the city are responsible for 64 percent of the calls for police services. That is an important thing to figure out because if you think about how you solve the problem with disease, you try to isolate; you try to define the vulnerable population, and then you try to solve the problem. If you spend all your time treating the patients without trying to solve the problem, trying to seek a way to inoculate, trying to locate a serum, trying to do better, you try to save some lives, and you spend all your time treating the people, and unfortunately as Tony said, "We spend our time packaging and producing and processing, rather than really curing."

Once we have analyzed the situation, I think we can see that an inordinate amount of police resources are not being diverted to solving problems but are being sapped off to dealing with the same people over and over and over again or the same situations.

We need to be problem oriented in our approach. The bottom line is after our treatment, after the millions of dollars, is there a better life after all the courts, after all of the probation; do we have a better life in our society? I think that is a problem-oriented approach that we have to take a look at and very carefully.

Newport News has tried that. The policing has tried that. We ought to try that throughout the entire system. Now, noting the expanding opportunities for field experiments, we should not forget the difficulties inherent in planning and implementing field experiments.

As Larry Sherman said in his opening remarks, experimental research is very different from evaluation research. In field experiments, extensive collaboration between personnel at all levels of the operational agencies and the research team is necessary just to get started, just to do the design, just to figure out the treatments. I think he talked about Tony Bouza,and about his police officers going out for beer and picnics and sitting around saying, "Geez, what do you think we ought to do? How should we do this? What do you think is in your range of capabilities as a police officer to do? What is in your range of tolerance to do?"

Incidentally, then what is in your range as measurers, researchers and evaluators to test for? That is what we need to have, that kind of collaboration, and what happened as a result is we got an appreciation for the integrity of the process. Like the basic rules of logic, if you do an experiment, and you adhere to the rules, you will end up with something you have confidence in that the result was the consequence of a special treatment, and then you can start talking about the treatment.

If you don't adhere to the rules, and we have a couple of examples of that. Mary Toborg mentioned her Washington, D.C. experiment on drug testing and how many of the judges didn't adhere to the process and how a problem then begins to come up because we cannot be sure the results and the promise in the program is the result of the treatment.

We, also, saw the same thing in the most prominent of all police studies, the Kansas City preventative patrol study where some officers directly sabotaged the design by patrolling those areas that were not supposed to receive any patrols.

That reminds me, and you all know this in court, it is when you have a particularly gruesome crime scene, and all the police want to come up and look through it. Right? They walk through to see the bodies and everything and touch the knife that was used, you know, and sit down in the chair, move the chairs around and everything, and then we get to court, and we have no case because the crime scene is contaminated.

That is the same thing the researchers have to worry about, contamination of the research design. Just like we don't want our cases thrown out of court, we don't want our research thrown out either.

I know we talked about this classic experiment because we got the police officers in Kansas City to sit down and agree that the fact is they adhere to the research procedures. It finally was one of those great and classic experiments. I know my good friend, Al Reiss thinks that the Kansas City experiment was not a true experiment and that the treatments were not randomly assigned to the areas, but on the basic point, I think we all agree that field experiments are difficult to implement.

There is another type of difficulty that the field experiment must face and that is the ethical and legal concerns raised by the random assignment treatments. I want to emphasize that NIJ has adopted special regulations for the protection of human subjects in all its research programs, and I think you have heard a good discussion about that earlier by Chuck Wellford.

Let me just say that we want to be sure that the social benefits outweigh whatever happens to the human subjects but on a sort of a cost/benefit line to all of you who are interested in doing research, let me say that these regulations are helpful for you conducting experiments because they become a protection really for you. The fact is you have not

gone off and done something that has not been reviewed independently, and now, you have some protection. The fact is that you have been concerned about that, and it wasn't a capricious judgment on your part, and I think it helps reduce your individual liability dramatically.

We recognize that ethical and legal considerations might impose limitations on some research designs, but clearly they do not prohibit field experiments in what you do. The number of existing experiments speaks to the ability of large numbers of jurisdictions to address adequately the legal and ethical concerns that have been raised.

Again, I think that we can learn from the medical profession. NIJ has recently decided to study research on the effectiveness of administering social services right at the hospital emergency rooms to victims of assault and rape. So, rather than have to be treated coldly, they have the opportunity to have a caring person to help victims and witnesses right there. I said that we ought to have Dick Berk come in on this project and see if we can do some randomization, to see if we can do something where we actually learn from this experiment.

But, the victim advocate group said, "Oh, no, this is too important a treatment. We have got to give it to these people because who needs it, and we are committed to this. We cannot randomize it because that would mean somebody doesn't get our services that needs it." But they were persuaded at the hospital and by the hospital administrators and doctors who pointed out that patients in this hospital, a very famous one, were regularly involved in similar experimental designs with untested medical treatments. There were people who were trying to survive who were given placeboes and other people who were trying to survive who were given a variety of medical protocols. The same question about ethics and

legality and the desire of the doctors to heal them is the same thing that the victims' advocate looked at, and they said, "Wait a minute. You are right. We ought to take a look and see whether we can communicate what makes a difference."

So, the ethical and legal concerns must be addressed in field experiments, but once they have been addressed, we should move forward to learn which treatments are most effective in which situations and in what types of individuals.

Learning what works is in some ways a moral imperative. It is in my mind a professional imperative. Many of us in this room have dedicated our lives to improving the level of criminal justice services. We have relied by necessity on traditions of our professions and on our own often misinformed judgments about what the most effective response is in a given situation.

It may be our best intentions and our most well thought out intuitive judgments, but it may be wrong. All professions require and will continue to require some degree of professional judgment, but the more advanced professions in our society, the ones capable of aiding their clients, the ones most respected by the American people and not incidentally the ones most highly paid, have moved beyond the folkways and the rules of thumb handed down from generation to generation, the cases that continue to build on the same wrong information.

In our country the medical profession has moved from a discipline dominated by folk medicine to a profession where scientific evidence is expected and in some cases required by law before new treatments are introduced. This change in the medical profession is relatively recent, occurring for the most part in this century and not really coming into

fruition until after World War II, because World War II provided a natural experiment to really advance medical science.

Now, what happened was that thousands of experiments were reported and debated and retested by thousands of practitioners. The basic issue that the criminal justice community has to decide is do we want to change from a craft and become more like a profession? I think the answer is yes, and I think that the field experimentation is a vital part of that change process. This change will not occur overnight or even as a direct result of conferences like this, but this is an investment in the future, and every day that you work to ask yourself, "Could we do this better?" is an investment in the future. It will take time and patience and energy, and it will involve trial and error, and we will make mistakes, lots of them, many mistakes, but here, too, we can learn from our medical colleagues, such as Oswald Avery who was a pioneer in the molecular genetics discovered by DNA.

Many of Avery's predictions turned out after careful testing to be wrong, but he capitalized on his errors. His colleagues reported his saying one time that whenever you fall be sure to pick something up.

We are at the early stages of making ours a modern profession, with stature. We will make mistakes, but we can, and we must learn from them. For those who are willing to adhere to the scientific method in conducting criminal justice experiments the road ahead is going to be difficult, but the potential rewards for our society and for us individually and for our professions is great. Randomized field experiments are the way to developing professional knowledge and professional expertise in controlling crime and creating a truly healthy society.

Thank you very much.

(Applause.)

PROF. REISS: Thank you, Chips, for the $6 wedding. We are very grateful. Chips said that one of the criteria for looking at these things is to ask, "Is there a better life?" and for those of you who have not experienced and been born again in this conference, I thought of giving you an opportunity to come forth at this time.

(Laughter.)

PROF. REISS: And to declare your commitment for randomized experiments, not just experiments and for randomized field experiments, and if you have not done so, this is your opportunity to do so.

I once rode a train, to end this on a note that ties to the beginning, I once rode a train from a place called Ladysmith, Wisconsin to Minneapolis. This is a true story, and it was a Sioux(?) line train, and on this train it mainly consisted of freight cars and so on, but at the end they had this one car which was a combination baggage car and passenger car, and every time we came into a little town, the conductor, baggageman, train agent, etc., would get up and say, "Ladies and gentlemen, this is Rice Lake, Wisconsin. Anyone who wishes to do so, may do so now." So, anyone who wishes to ask any questions may do so now, but I suspect that all of you would rather do so now.

So, it is with my blessing that you are free to leave.

Thank you very much.

(Thereupon, at 3:13 p.m., the workshop was concluded.)

CRIMINAL JUSTICE EXPERIMENTS

compiled by Michael L. Dennis

The following synopses of 22 criminal justice experiments involve a wide range of alternative procedures and programs for offenders. All were carried out in the United States. A large number of experiments with juvenile offenders are not included here. The synopses are classified into four groups: (1) experiments with arrest and pretrial release; (2) experiments with adjudication procedures; (3) experiments using alternative sanctions; and (4) experiments during incarceration or after release. For each experiment, the following information is presented:

Summary: The first paragraph states the experiment's objective, time frame, and geographic location and summarizes what was done. The second paragraph gives the experimental results.

Administering Organization: States the organization with primary responsibility for implementing the experiment.

Principal Investigator: The researcher who had primary responsibility for designing the experiment and analyzing its results.

Sponsoring Agencies: Agencies that funded the experiment. Wherever possible, the amount of these funds are given. Note that these funds are not in 1987 dollars.

Research Team: Lists the types of people and organizations who participated in designing the experiment and analyzing its results. Criminal justice officials who delivered the treatments but were not included in these tasks are not listed.

Michael Dennis is a doctoral student in the Psychology Department of Northwestern University. The Connecticut Restitution Experiment and the Drunk Driving and Jail Time Experiment were summarized by Richard Lovely and Lawrence Sherman, respectively.

Target Group: The criteria for including and excluding subjects or groups from the experiment.

Treatments: The procedures or treatments that differed between the groups within the experiment.

Assignment Method: How subjects were selected and assigned to each of the treatments.

Outcome Measures: What variables were measured to determine whether the treatments had an effect.

Reported Implementation Problems: Problems that arose when the experiment was implemented. These problems are based solely on the cited reports by the author or their critics.

Impact: The impact, if known, of the research results on criminal justice practices and laws.

Sources: The sources from which information for the synopsis was collected.

SYNOPSES BY CATEGORY

Arrest and Pretrial Release Experiments

DOMESTIC VIOLENCE (MINNEAPOLIS) EXPERIMENT: determines the most effective police response to domestic violence calls.

ELECTRONICALLY MONITORED PRETRIAL RELEASE EXPERIMENT: effectiveness of electronically monitored pretrial releases in terms of pretrial outcomes (e.g., rearrest, FTA) and cost.

FELONY ARREST DIVERSION EXPERIMENT: impact of pretrial diversion on case disposition, recidivism, and the utilization of social services.

KANSAS CITY PREVENTIVE PATROL EXPERIMENT: effectiveness of routine preventive patrol on crime rates, community attitudes, and public satisfaction with police services.

MANHATTAN BAIL BOND PROJECT: effect of verified background information about the defendant on the number of pretrial releases and level of court appearances.

PHILADELPHIA BAIL GUIDELINES EXPERIMENT: effectiveness of guidelines in reducing judicial disparity by structuring bail decisions in accordance with crime severity and default risk.

PRETRIAL RELEASE EXPERIMENTS: the impact of pretrial release on recognizance (ROR) in terms of the rates of failure-to-appear (FTA), pretrial recidivism, and the equity of the release process (e.g., what kind of offenders get pretrial releases).

SHOPLIFTING ARRESTS AND RECIDIVISM EXPERIMENT: effects of arrest on shoplifters in terms of their later recidivism.

Adjudication Procedures Experiments

INDIANAPOLIS PROSECUTION EXPERIMENT: effectiveness of a variety of prosecution strategies in reducing subsequent violence among wife batterers.

LITIGATION RESOLUTION EXPERIMENT: effectiveness of court-supervised conferences and court orders reducing litigation.

PRETRIAL SETTLEMENT CONFERENCE EXPERIMENT: the impact of judicially supervised plea bargaining conference on the processing and disposition of cases.

TELEPHONE BAIL RAISING EXPERIMENT: effect of installing telephones in detention centers and defendants' ability to raise bail.

Alternative Sanctions Experiments

COMMUNITY REHABILITATION (ELLSWORTH HOUSE) EXPERIMENT: relative effectiveness of a community-based rehabilitation program in terms of recidivism and employment.

CONNECTICUT RESTITUTION EXPERIMENT: impact of requiring restitution in criminal cases.

DENVER COUNTY DUI COURT EXPERIMENT: relative effectiveness of the common court sanctions for driving under the influence (DUI).

DRUNK DRIVING AND JAIL TIME EXPERIMENT: effectiveness of a jail sentence for convicted drunk drivers on their subsequent traffic violations.

DWI REHABILITATIVE SANCTIONS EXPERIMENT: effectiveness of probation and educational/therapy programs on the rearrest rates for drunk drivers (DWI).

SAN QUENTIN SQUIRES EXPERIMENT: impact the Squires program had on the attitudes and subsequent behaviors (recidivism) of program participants.

SOCIAL SERVICES AND PROBATIONER RECIDIVISM EXPERIMENT: is intensive probation better than conventional probation in terms of probationer recidivism?

PRISON GROUP COUNSELING EXPERIMENT: the impact of group counseling on prisoners' attitudes, cooperation, and parole behavior.

TRANSITIONAL AID RESEARCH PROJECT (TARP): impact of extending unemployment benefits to recently released prisoners on their subsequent employment and recidivism.

WORK RELEASE AND RECIDIVISM EXPERIMENT: effectiveness of prison work release programs in terms of recidivism.

DOMESTIC VIOLENCE (MINNEAPOLIS) EXPERIMENT

This experiment was designed to determine the most effective police response to domestic violence calls. It was conducted from March, 1981 until August, 1982 in Minneapolis. The officer who arrived on the scene of a domestic violence call was responsible for determining whether the case was eligible for the experiment. If it was, the defendant was randomly assigned to one of three response treatments: On the spot advice; separation for at least 8 hours; or immediate arrest. The 33 officers selected for the experiment were given 3 days of training. During this period, they were able to negotiate some of the specific rules for the actual implementation. Subsequent domestic violence calls and the suspects' records were then monitored for 6 months. Victims were interviewed within two weeks of the incident and every other week thereafter for 24 weeks. There were 314 cases involved in this experiment.

Arrest was the most effective police response in terms of deterring future violence. In a 6-month follow-up, arrested assailants had a recidivism rate of only 19%, compared with 33% for those who were separated, and 37% for those who received advice. Arrest was the most effective (9% recidivism) when victims perceived that the officer "listened" to their story before arresting the suspect.

ADMINISTERING ORGANIZATION: Minneapolis Police Department

PRINCIPAL INVESTIGATOR: Lawrence W. Sherman
(formerly, of the Police Foundation)
Crime Control Institute
2125 Bancroft Pl., N.W.
Washington, D.C. 20008
(202) 797-7410

SPONSORING AGENCY: National Institute of Justice

RESEARCH TEAM: The Police Foundation (Washington, DC), Minneapolis Police Chief, and 42 Patrol Officers.

TARGET GROUP: The target population was simple domestic assaults (misdemeanor) in which both parties were present and the officer could establish probable cause to believe that a cohabitant had assaulted the victim within the past four hours. Life threatening or felony cases were excluded.

TREATMENTS: 1) Advice in which the officer tried to mediate the dispute and offer on-the-spot counseling; 2) Separation, in which the assailant was persuaded to leave the premises for a cooling off period of at least 8 hours; and 3) arrest, in which the suspect was immediately arrested.

ASSIGNMENT METHOD: Once an officer determined that a given case was eligible, assignment was made on the basis of a randomly ordered form pad. The forms contained the experimental data questions. They differed in color; one color for each of the three treatments. If the situation changed (i.e., the suspect became violent), the officer took appropriate actions regardless of treatment assignment.

REPORTED IMPLEMENTATION PROBLEMS: There were significant internal hostilities going on between the patrol officers and their superiors. By the third month, it became obvious that only two-thirds of the officers were active. With the case flow slower than what had been expected, 18 additional officers were recruited and the original 12-month time frame was expanded to 16 1/2 months. The number of cases per officer varied considerably; three officers produced 28% of the cases. However, randomization was done by officers which minimized the trends' impact.

IMPACT: Minneapolis has adopted a pro-arrest policy and a recent survey by the Crime Control Institute (CCI) shows that the number of cities with pro-arrest policies has risen from 10% to 44% in two years. NIJ is currently funding six replications (in Colorado Springs, Omaha, Milwaukee, Atlanta, Charlotte, and Dade County, Florida) to examine potential limitations and alternative applications of this study.

SOURCES: Sherman, L. W. and Berk, R. A. (1984) The specific deterrent effects of arrest. American Sociological Review, 49: 261-272.

Sherman, L. W. and Berk, R. A. (1984) The Minneapolis Domestic Violence Experiment. Washington, D.C.: The Police Foundation.

ELECTRONICALLY MONITORED PRETRIAL RELEASE EXPERIMENT

This experiment is designed to test the effectiveness of electronically monitored pretrial releases in terms of pretrial outcomes (e.g., rearrest, FTA) and cost. The 24-month study is currently being conducted by the Municipal Court of Marion County (Indianapolis), Indiana. It will follow 400 Class "D" felons (e.g., theft, DWI, possession of a handgun), who have still been unable to post bond after 2 weeks. Half of them will be offered pretrial release, conditioned on home detention under electronic surveillance by means of a Telso "On Guard" Wristlet. The other half will remain in detention pending bail and constitute a control group. (Class "D" (nonviolent) felony calls for a $2,500-$3,500 surety bond.) Typically 75% of the Class "D" defendants in the jail have been there for more than two weeks. Another 300 Class "D" felons will be selected from those who were able to post bail (100) and those who were released on their own recognizance (ROR, 100), and those released non-financially conditioned on periodic drug testing (100). These will all serve as comparison groups. When the cases reach disposition, a variety of information will be collected. The defendants in the control group will be monitored to see if they secured release before the trial and, if so, the date of release and the release conditions. Similar information will be collected for the two comparison groups. In addition to the official records, defendants in the Electronic Surveillance program will be interviewed during their last week. The interviews will record the defendants' experiences while in the program - in particular, whether they had any difficulties with the equipment or any unanticipated adjustment problems. Finally, key local criminal justice officials will be interviewed about their reactions to the program and how it could be improved or expanded.

The proposed analyses will focus on: 1) the pretrial rearrest rates and their related charges; 2) failure to appear rates and whether the defendant subsequently returned; 3) and the program's cost effectiveness relative to detention. These analyses will be expanded to include the three other comparison groups which were released on bond, ROR or nonfinancially, conditioned on drug testing.

ADMINISTERING ORGANIZATION: The Court Administrators Office of the Municipal Court of Marion County.

PRINCIPAL INVESTIGATOR: John P. Bellassai, J.D.
Toborg Associates
1725 K Street, N.W.
Suite 803
Washington, D.C. 20006
(202) 293-0888

SPONSORING AGENCY: National Institute of Justice

RESEARCH TEAM: Toborg Associates. Practitioners' advice and assistance will be provided by the Presiding Judge of the Municipal Court of Marion County, several of the court's judges, the Prosecuting Attorney, the Court Administrator, the Bail Commissioners, the Public Defender, the Sheriff, and the Federal Court Commissioners who have been appointed to oversee efforts to reduce jail overcrowding in Indianapolis.

TARGET GROUP: Class "D" (nonviolent) felons who have been unable to raise bail after at least two weeks in detention. Three additional groups of Class "D" felons will be selected for comparison: those who were able to post bail, those who were released on their own recognizance, and those who were released conditioned on drug testing.

TREATMENTS: (1) The electronic surveillance program, in which a bond reduction hearing is convened for each defendant. At that hearing, attorneys for the defense and the prosecution jointly request that the defendant be released on the condition of complying with the requirements of the electronic surveillance program for 90 days or until the case goes to disposition, whichever comes first. During the period in the program, a computer periodically calls the defendant at home. It must be answered by the "On-Guard" Wristlet that is strapped to the defendant's arm. After 90 days of successful compliance with the program, the wristlet will be removed and the pretrial release order modified to stipulate third party custody instead of electronic surveillance. (2) The control group, in which defendants remain in detention unless they post bond.

ASSIGNMENT METHOD: Class "D" felons who have been unable to post bond after two weeks are randomly assigned to either the surveillance program or the control group using their "gallery number" -- a unique personal identifier assigned to each person in Marion County at the time of his or her first arrest.

OUTCOME MEASURES: Number of court appearances missed; whether the defendant returned to court after any failure-to-appear; pretrial rearrests (including the charges and outcome) and the cost of the treatment. In addition, qualitative information will be collected from defendants in the program and local criminal justice practitioners.

IMPLEMENTATION PROBLEMS: Because the Indianapolis jail is overcrowded, the investigator feared that the courts would overuse electronic surveillance (i.e., when ROR would have been sufficient) and thereby widen the net. To prevent this, the target population will be limited to only those Class "D" felons who have been unable to raise bond after 2 weeks. The use of the defendants' gallery number, which is rechecked in the follow-up phase, allows a direct monitoring of how well randomization was implemented.

IMPACT: Because of the growing interest in this technology, the investigators are planning a monograph which will analyze the legal, programmatic and policy issues in addition to the outcomes from the various study groups.

SOURCE: Toborg Associates, Inc. (1986) Electronic Surveillance of Pretrial Releases As A Jail Crowding Reduction Strategy: Evaluation of a Controlled Experiment in Indianapolis, Indiana. Grant application submitted to the National Institute of Justice, April 26, 1986. (Grant awarded September 30, 1986).

FELONY ARREST DIVERSION EXPERIMENT

This experiment was designed to assess the impact of pretrial diversion on case disposition, recidivism and the utilization of social services. It was conducted from January, 1977 until October, 1978 in the Manhattan and Brooklyn Criminal Courts. The study evaluated the Court Employment Project (CEP) by assigning 666 felony defendants to either CEP's diversion program (410) or a control group (256) which proceeded through the legal process. Sixty percent of the defendants had no prior record and only 16% had a prior conviction. After assignment, the case dispositions were submitted to judicial review and occasionally (12%) reversed. The CEP program sought to obtain dismissal of the charges pending against diverted defendants. It also assisted defendants in seeking out social services, remedial education, vocational training and employment. Defendants were interviewed three times; at intake, 6 months after intake, and 12 months after intake. During these interviews, subjects were asked about their employment, education, use of social services, drug and alcohol use, and general level of legitimate activity. Criminal history records and subsequent arrest data were collected from the New York City Police Department and the New York City Criminal Justice Agency. Finally, an attempt was made to verify self reports on the defendant's employment, school enrollment and public assistance status.

The diverted defendants were more likely to have their charges dismissed (72% versus 46%) and less likely to be convicted (1.4% versus 6.6%). Failure in the diversion program did not appear to prejudice the eventual case disposition. CEP's intervention did not influence its clients' vocational activities, school involvement, drug use, or level of legitimate activity. Recidivism rates for the diverted defendants (19.8%) and the controls (16.5%) were similar during the initial 6 month and at the 12 month follow-up (30% and 33% respectively). The number of rearrests and convictions are also similar.

ADMINISTERING ORGANIZATIONS: Court Employment Project (CEP) and the Vera Institute of Justice

PRINCIPAL INVESTIGATOR: Sally Hillsman (Baker)
Director of Research
Vera Institute of Justice
New York, New York

SPONSORING AGENCY: National Institute of Law Enforcement and Criminal Justice

RESEARCH TEAM: Vera Institute, CEP, Criminal Court Judges from Manhattan and Brooklyn, and the Bronx, attorneys, academics, New York City prosecutors, and the Legal Aid Society.

TARGET GROUP: Young felony offenders, between the ages of 16-20, being tried in adult court. The emphasis was placed on those with no prior records (60%) or convictions (84%).

TREATMENTS: 1) CEP's diversion program, in which defendants are temporarily removed from the legal process and are given assistance in solving legal, social, and vocational problems. Failure to progress in the program will lead to the defendants' case proceeding to trial; control group, in which defendants eligible for diversion are left in the regular legal system.

ASSIGNMENT METHOD: Recruitment efforts were used to increase demand to twice the level that could be served by the program. The working hours of the experiments' first six months were broken randomly into periods varying between 11 to 21 hours. For each period, the expected number of eligible cases was calculated. The first half of the expected cases in a given period would be placed in the diversion program and the rest in the control group. Codefendants always received the same disposition. Some of the defendants in the diversion program came in directly and were not considered in the analysis.

OUTCOME MEASURES: From official records; case dispositions; criminal history; recidivism in terms of rearrests and subsequent sanctions; level of CEP participation. From self reports; education level; vocational experience; employment; level of legitimate activity; use of community resources; abuse of drugs or alcohol. The record check was done 23 months after the experiment started. Defendants were interviewed at intakes then followed up 6 and 12 months later.

REPORTED IMPLEMENTATION PROBLEMS: The defense bar objected to the random assignment. In response, Vera developed an acceptable approximation. CEP had to accept some outside cases because of interorganizational politics. The judges insisted and utilized their right to review and reverse case dispositions in 12% of the cases.

IMPACT: In response to Vera's findings, CEP abandoned pretrial diversion and has launched a post-conviction supervision program. The new program provided sufficient supervision to serve as a substitution for jail sentences. The social services provided by CEP are now open to anyone in need rather than required as a condition of diversion. It also began development of a training program for exoffenders.

SOURCES: Baker, S. (Hillsman) and Rodriquez, O. (1979) Random Time Quota Selection: An alternative to random selection in experimental evaluation. In Lee Sechrest (Ed.) Evaluation Studies Review Annual, 4. Beverly Hills, CA: Sage Publication.

Baker, S. (Hillsman) and Sadd, S. (1981) Diversion of Felony Arrests: An Experiment in Pretrial Intervention. 1981 Summary Report. Evaluation of the Court Employment Project. Washington, D.C. National Institute of Justice.

Hillsman, S. T. (1982) Pretrial Diversion of Youthful Adults: A decade of reform and research. The Justice System Journal, 7(3): 361-387.

The experiment was designed to determine the effectiveness of routine preventive patrol on crime rates, community attitudes, and public satisfaction with police services. It was conducted from July, 1972 until September, 1973 in 15 of 24 beats in the South Patrol Division of Kansas City. The experiment compared three levels of routine preventive patrol (reactive, control, and proactive) to assess their impact on crime rates and community attitudes. In the reactive condition, officers only responded to calls. In the control condition, they carried out the normal level of routine preventive patrols. In the proactive condition, additional officers were assigned so that the level of preventive patrols could be increased. While they were not on call, officers from reactive beats spent their time conducting preventive patrols in the proactive beats. In addition to police records, researchers used field observers riding with the officers, and before and after surveys of the community and local businesses.

The study found that increasing or decreasing the levels of routine preventive patrol had no major effects on crime, citizens' fear of crime, community attitudes towards police, the delivery of police service, police response time, or traffic accidents. Of the 648 comparisons used to make the study's 13 major findings, only 6% showed any reliable changes. Even these changes were distributed among the three conditions. In fact, 48% of the changes occurred in the control beats. There were also no differences between the number of protective and security measures taken by citizens and businesses in the experimental beats versus those taken in the control beats. Later reviewers have suggested that the experiment also demonstrated the flexibility available to patrol administrators in the distribution of their resources: that within a given division, officers do not need to be evenly distributed geographically.

ADMINISTERING ORGANIZATION: Kansas City Police Department

PRINCIPAL INVESTIGATOR: George L. Kelling

SPONSORING AGENCIES: Police Foundation

RESEARCH TEAM: The Police Foundation and its evaluation advisory board, the Midwest Research Institute, Academics, the Police Chief, K.C. South Patrol Task Force, Police Officers, and private consultants.

TARGET GROUP: The typical patrol beat in which preventive patrol is currently used. Helicopter, K-9, and other specialized units were excluded.

TREATMENTS: 1) Control beats with normal levels of preventive patrols; 2) Reactive beats, where officers did not engage in preventive patrols; and 3) Proactive beats, where the number of officers and the level of preventive patrols were increased.

ASSIGNMENT METHOD: Fifteen of the South Division's 24 beats were selected as representative of the city. A computer was used to form them into groups of 3 beats each. Each group was similar in terms of prior crime rates, demographics, and the prevalence of commercial properties. One beat from each group was assigned to each of the 3 treatments. The reactive beats were never next to each other. This was done to avoid a slow down in the officer's average response times. Since the 15 beats geographically formed a rough rectangle, one reactive beat was placed in the middle of the area and the remaining four in each of its corners.

OUTCOME MEASURES: Outcome was measured with four types of variables: departmental crime reports, departmental arrest data, victimization survey data, and commercial survey data. Crime rates were collected from 10 categories; five thought to be affected by preventive patrol (burglary, auto theft, larceny, and other sex-related crimes). Victims were surveyed about their attitudes towards the current level of public safety and police response time. The survey also sought to estimate their level of fear and willingness to report crime. Local merchants and business establishment were given a similar survey.

REPORTED IMPLEMENTATION PROBLEMS: There was originally considerable concern about the possible risk to the general public caused by a reduction in routine patrols. It was therefore agreed to monitor crime rate data on a weekly basis and to terminate the experiment should a noticeable increase occur. During the first month of the experiment, it became obvious that experimental conditions were not being maintained. There was insufficient manpower to provide the proactive treatment and officers in the reactive beats were violating the guidelines, apparently out of boredom. The experiment was suspended for a month. During this time, Chief Kelley saw to it that the additional manpower was assigned to the South Patrol Division. The researchers held additional training and changed the guidelines to allow officers in the reactive beats to conduct preventive patrol activities in the proactive beats. The guidelines were also changed to acknowledge that officers should take appropriate actions should a criminal incident be observed.

IMPACT: NA

SOURCES:
Fienberg, S. E., Larntz, K., and Reiss, A. J. (1976) Redesigning the Kansas City Preventive Patrol Experiment. Evaluation, 3, pp. 124-131.

Kelling, G., Pate, T., Dieckman, D., & Brown, C. E. (1974) The Kansas City Preventive Patrol Experiment: A Summary Report.

Larson, Richard C. (1975) What happened to patrol operations in Kansas City? Journal of Criminal Justice, 3, 299-330.

Pate, T., Kelling, G. L., & Brown, C. (1975) A response to "what happened to patrol operations in Kansas City?" Journal of Criminal Justice, 3: 299-330.

Risman, B. J. (1980) The Kansas City preventive patrol experiment: A continuing debate. Evaluation Review, 4:802-808.

This experiment was designed to test the effect of verified background information about the defendant on the number of pretrial releases and level of court appearances. It was conducted between 1961 and 1964 in Manhattan, New York. The target group was individuals accused of felonies and misdemeanors; individuals charged with homicide and other serious crimes were excluded. New York University law students, working under Vera Institute staff, reviewed defendants' records of employment, family, residences, references, current charges, previous records, etc. in order to make judgments about whether a pretrial release without bail should be recommended to the court. If a recommendation was made for the defendant to be released without bail, the defendant was randomly assigned to either the experimental or control group. All of the approximately 730 recommendations were delivered to Vera court staff prior to the time of bail setting. Only the recommendations for the experimental group were then delivered to the judge. Vera staff also gave the experimental defendants additional court date notifications and recontacted the ones who missed their court date to encourage them to seek a continuance. Subsequent paroles, releases, defaults, case dispositions, and sentences, were monitored.

Judges granted release without bail to 59% of the experimental group compared to only 16% in the control group; recommendations based on information then served to increase the rate of release without bail. Sixty percent of the recommended group was either acquitted or had their cases dismissed, compared to 23% of the control group. During 1961-64, less than 1% of the experimental group failed to show up in court for trial. This suggests that the relaxation of the bail requirement did not result in unacceptable default rates.

ADMINISTERING ORGANIZATION: Manhattan Criminal Court
Vera Institute of Justice
New York University School of Law

PRINCIPAL INVESTIGATOR: Charles E. Ares
College of Law
University of Arizona
Tucson, AZ 85721
(602) 621-1373

SPONSORING AGENCIES: Institute of Judicial Administration
The experiment cost less than $200,000 (in 1961 dollars) for 3 years.

RESEARCH TEAM: Academic, Vera Institute, Institute of Judicial Administration, New York Supreme and Criminal Court Justices, New York City Department of Corrections, Department of Probation, District Attorney, the Mayor and the Legal Aid Society.

TARGET GROUP: Individuals accused of felonies and misdemeanors; individuals charged with homicide and other serious crimes were excluded.

TREATMENTS: 1) Recommendation for release without bail given to the court along with verified background information; 2) No recommendation or information given to the Court.

ASSIGNMENT METHOD: All cases were reviewed and only cases where a pretrial release was recommended were included. All recommendations were delivered to a staff person at the court. The staff person made assignments based on a list of all case dockets in which each docket was randomly assigned to either the control or experimental group. Only the recommendations for the experimental group were delivered to the judge.

OUTCOME MEASURES: Paroles granted, pretrial releases, case dispositions, sentences, and default rate (failure to appear in court).

REPORTED IMPLEMENTATION PROBLEMS: Professional bail bondsmens' Organizations opposed the selective relaxation of bail requirements. Questions were raised about the confidentiality of case-related information that arose in the pretrial interviews.

IMPACT: Following the experiment, the New York Probation Department extended this program to criminal courts in all five boroughs of the city. Similar projects were launched in Des Moines, Washington, D.C., and St. Louis. Many of the experimental treatment features were incorporated into the Federal 1966 Bail Reform Act.

SOURCES: Ares, C. A., Rankin, A., and Sturz, H. (1963) "The Manhattan Bail Project: An interim report on the use of pretrial parole." New York University Law Review, 38: 67-95.

Botein, B. (1965) "The Manhattan Bail Project: Its impact in criminology and the criminal law process." Texas Law Review 43: 319-331.

Riecken, H. W. and Boruch, R. F. (1974) Social Experimentation: A Method For Planning and Evaluating Social Intervention. New York: Academic Press.

PHILADELPHIA BAIL GUIDELINES EXPERIMENT

This experiment was designed to test the effectiveness of guidelines in reducing judicial disparity by structuring bail decisions in accordance with crime severity and default risk. It was conducted from January, 1981 to March, 1982 in Philadelphia's Municipal Court and Court of Common Pleas. The experiment was initiated after the findings of an earlier study showed that a primary factor in level of bail or release on recognizance (ROR) was individual differences between judges. The guidelines were developed collaboratively by the researchers and the judges to specify a presumptive range of decisions based on ratings of chance severity and risk of failure to appear. Eight judges were randomly assigned to use the guidelines, eight others served as controls and a total of 1920 cases were examined. The information and classifications were presented only to the 8 experimental judges. The distribution of actual bail levels, ROR, detention rates and default rates were then compared for the experimental and control judges. Experimenters conducted random paperwork checks to monitor whether treatments were being delivered.

Compared to the controls, experimental judges gave 8-12% more RORs to misdemeanor defendants and 5-12% less to felony defendants. They gave lower bails on the average ($1,500 versus $1,995), with differences as large as $2,050 for the most severe category of felonies. Despite lower bails, felony defendants before experimental judges were no more likely than other felony defendants to be released. Experimental misdemeanor defendants, however, were 3-5% more likely to be released. The guidelines did not appreciably affect the rates of failure to appear and rearrest, but did tie bail decisions more closely to risk.

ADMINISTERING ORGANIZATIONS: Philadelphia's Municipal Court and Court of Common Pleas.

PRINCIPAL INVESTIGATOR: John S. Goldkamp
Department of Criminal Justice
Temple University
Philadelphia, PA 19122

SPONSORING AGENCIES: National Institute of Justice

RESEARCH TEAM: Judicial Steering and Policy Committee of the Bail Decision making Project; Presidents and Judges from the Philadelphia Municipal Court and Philadelphia Court of Common Pleas; Pretrial Service Division of Philadelphia, Academics.

TARGET GROUP: All defendants in the Municipal and Common Pleas Courts of Philadelphia. Cases for which strict and long standing traditions of bail denial existed (e.g., murder) were excluded.

TREATMENTS: 1) Guideline judges received copies of the bail guidelines, and information classifying the defendant by the current charges of severity and the risk that they will fail to reappear. 2) Control judges continued to set bail and ROR decisions as they had in the past.

ASSIGNMENT METHOD: Sixteen judges were randomly selected from the two courts. Eight were randomly assigned to receive guidelines and recommendation. An equal number of cases were randomly selected for each of the 16 judges. The same distribution of cases, in terms of charge severity, were selected for each judge.

OUTCOME MEASURES: Percentage of releases on recognizance; percentage of release on bail; percentage of detainees; percentage of rearrests during pretrial release; percentage of failure to appear; amount of bail; length of detention.

REPORTED IMPLEMENTATION PROBLEMS: One of the experimental judges had to be excluded because it was discovered too late that he didn't understand how he was supposed to use the guidelines. Since the eight experimental judges were conscripts rather than volunteers, they varied in their desire to utilize the guidelines. With the assistance of the court's president, they did comply. Out of 840 experimental decisions, 75.7% were within the guidelines; 13.6% were higher; and 10.7% were lower. Of the judges who gave justifications, most cited reasons in line with the current research (e.g., prior criminal records, nature of current charges, or prior history of appearances). Of the 35% of the decisions that failed to give an explanation, most were from 2-3 judges. Some cases which fell at a shift change had to be excluded.

IMPACT: The guidelines were adopted by the Philadelphia Municipal Court in 1982. The court has continued to revise and improve the original guidelines. The National Institute of Justice is funding studies to develop similar guidelines in Boston, MA; Dade County, FL; and Maricopa County, AZ.

SOURCES: Goldkamp, J. S. and Gottfredson, M. R. (1985) Policy Guidelines for Bail: An Experiment in Court Reform. Philadelphia: Temple University Press.

Goldkamp, J. S. and Gottfredson, M. R. (1984) Judicial Guidelines for Bail: The Philadelphia Experiment. Washington, D.C.National Institute of Justice.

This series of experiments was designed to determine the impact of expanded pretrial release on recognizance (ROR) in terms of the rates of failure-to-appear (FTA), pretrial recidivism, and the equity of the release process (e.g., what kind of defendants get pretrial releases). They were conducted in 1978 and 1979 in four cities: Baltimore, Md; Lincoln, NE; Beaumont-Port Arthur, Tx; and Tucson, Az. In each case, the experiment was implemented by a local program and monitored by the research team. Eligible defendants were randomly assigned to receive either assistance from the expanded local programs in attaining a ROR, or no assistance (control group and the situation that had existed before program expansion). The number of defendants studied ranged from 130 to 719 per site, with about half in each group. The defendants were then tracked in terms of pretrial release, FTA, pretrial criminality and case disposition. In two of the sites, defendants who were released (from both the control and program groups) were then randomly assigned to receive either program follow-up or no follow-up.

The studies found that more defendants were released in the program than control groups, and that rates of failure-to-appear and pretrial criminality were virtually identical for both groups. The investigators recommended that the program expansions be continued.

ADMINISTERING ORGANIZATIONS: Pretrial-release programs in Baltimore, Md.; Lincoln, NE; Beaumont-Port Arthur, Tx; and Tucson, Az.

PRINCIPAL INVESTIGATOR: Mary A. Toborg
(then with the Lazar Institute,
Washington, DC)
Current Affiliation: Toborg Associates, Inc.
1725 K Street, N.W.
Suite 803
Washington, D.C. 20006
(202) 293-0888

SPONSORING AGENCY: National Institute of Justice

RESEARCH TEAM: The Lazar Institute. Advisory panel consisted of academics, program officials, criminal justice officials.

TARGET GROUP: Defendants in pretrial detention.

TREATMENTS: Pretrial release, in which a release program was expanded and advocated for the defendants' release on their own recognizance. Control group, in which defendants who had previously received no assistance from the release program continued to receive no assistance. In two sites follow-up by the release program was randomly assigned to half of the defendants who were actually released from either of the two treatments.

ASSIGNMENT METHOD: In three sites, the days of the year were randomly assigned to two lists; one for each treatment. Defendants with birthdays that were on the pretrial release program list were assigned to it. In 2 of the 3 sites, half of each list was then assigned for program follow-up. In the fourth site, defendants were randomly assigned by the day of their arrest.

IMPLEMENTATION PROBLEMS: In one site interviewers required several training sessions in order to reliably use the eligibility protocols. The site that randomized the day of arrest had several program days in which no cases occurred. Staff in that site incorrectly re-designated the next day as a program day.

IMPACT: The study has been widely used by the pretrial practitioner community and by federal and state legislators planning changes in pretrial release practices (i.e., in "bail laws"). For example, it was referenced by members of Congress at several points in the legislative history of the Federal Bail Reform Act of 1984 and predecessor bills introduced over several years.

SOURCES: Toborg, M. A. et al. Pretrial Release: A National Evaluation of Practices and Outcomes, Phase II Report. Washington, D.C.: Institute of Justice, 1981.

SHOPLIFTING ARRESTS AND RECIDIVISM EXPERIMENT

This experiment was designed to test the effects of arrest on shoplifters in terms of their later recidivism. It was conducted in 9 branch stores of a department store chain from August, 1983 until April, 1984. The 1595 subjects were apprehended by store security personnel while shoplifting. Security personnel escorted them into a back office, questioned them on the incident, and collected background information. The questionnaire on which this information was recorded also called for the security officer to comment on the apprehendee's appearance, attitude and level of cooperation. The latter included whether they resisted apprehension and whether handcuffs had been used. Each incident was then recorded in a time-governed store apprehension log book and randomly assigned to an arrest (759) or release disposition (836). The arrest condition called for the apprehendee to be turned over to the police. The release condition allowed them to leave. In both cases, they were detained for at least an hour. Juveniles were only released to their parents or guardians. Six months after the experiment was over, researchers collected arrest history information from the State Police, and the nine local police departments serving the branch stores.

There were little differences between the rearrest rates of the studies' 1595 apprehendees. In the six months following each incident, the overall rate of shoplifting rearrest was low (5.8%), as were arrests for other crimes (5.1%). Arrested apprehendees were actually a little more likely to be arrested for other crimes in the subsequent 6 months. The only group for which arrest appeared to produce a deterrent effect (-50%) was apprehendees who had been described as unafraid, uncooperative, and sloppy dressers. For single, employed, high school graduates between the ages of 17-29, arrest appeared to backlash; these arrested apprehendees were 6-10 times more likely to be arrested during the subsequent six months.

ADMINISTERING ORGANIZATION: Security personnel in a department store chain. (Identity confidential)

PRINCIPAL INVESTIGATOR: Brian Forst
Police Foundation
1001 22nd St., NW
Washington, DC 20037

Lawrence W. Sherman
Crime Control Institute
2125 Bancroft Place, N.W.
Washington, D.C. 20008

SPONSORING AGENCY: National Institute of Justice

RESEARCH TEAM: Police Foundation, Academic, Department store management

TARGET GROUP: Shoplifters in 9 branch stores

TREATMENTS: 1) Release, in which apprehendees were eventually released without charges, and 2) Arrest, in which they were turned over to the police. All apprehendees were held for at least one hour. Juveniles were only released to their guardians or parents.

ASSIGNMENT METHOD: Each shoplifting event was assigned to either arrest or release in a randomly predetermined order recorded in the store apprehension log. If more than one individual was involved in a single incident, they were all given the same treatment. There were 4 reasons for a suspect to be excluded: more than 2 prior apprehensions; the apprehension involved a physical injury; a recent apprehension in other stores in the vicinity; or possession of contraband, such as drugs or guns.

OUTCOME MEASURES: Over the six months following each incident, the following were recorded: shoplifting arrests, other arrests, time between apprehension and the next subsequent arrest. For other arrests, the type of crime was also recorded.

REPORTED IMPLEMENTATION PROBLEMS: The security personnel were able to stall the log-in and get uncooperative apprehendees arrested. The incidence of sequence violations was 3%, most of which occurred in the first 3 months. Another 5% were dropped as official exceptions. The author notes that the store's procedures are already quite severe - which might account for the low rearrest rates; 16% of the apprehendees were handcuffed during the incident. Arrest and rapid release by an overworked police station may have had little additional impact. An attempt to survey the apprehendees failed; less than 40% responded and many denied the incident.

IMPACT: NA

SOURCE: Sherman, L. W., and Gartin, P. R. (1986) Differential recidivism: A field experiment on the specific sanction effects of arrest for shoplifting. Draft report to the National Institute of Justice, Summer Research Fellowships. October, 1986.

Glick, B., Hamilton, E., and Forst, B. (1986) Shoplifting; An Experiment in Lesser Crimes and Punishments. Draft final report. Police Foundation, Washington, DC.

NOTE: The Police Foundation and Crime Control Institute reports agree on the overall effect of arrest. Preliminary analyses, however, reveal differences on the effects among some subgroups. Analysis is continuing to determine the source of the difference.

INDIANAPOLIS PROSECUTION EXPERIMENT

This experiment is designed to test the effectiveness of a variety of prosecution strategies in reducing subsequent violence among wife batterers. It is currently being conducted in Indianapolis, Indiana. Approximately 660 wife battering cases will be randomly assigned to one of four prosecution strategies. Two require prosecution; a recommendation for a fine or jail time, and a recommendation for rehabilitative probation. The other two involve non-prosecution alternatives: the first is to inform the woman of her right to drop the charges; the second is to use diversion to a rehabilitative program. About 5% of the cases that are prosecuted are expected to be found not guilty. The cases come from two sources: on-scene arrests by the police (220 expected); and victim-initiated complaints (440 expected). In the latter cases, the assailants will be randomly assigned to receive either a court summons or an arrest warrant. Victims will be surveyed at the time of the complaint or shortly after the assailant's arrest and again within one month of the court date on which the case was disposed. The primary victim follow- up will be face-to-face interviews conducted 6 months after the court date. Assailants will be interviewed immediately after the disposition of their case. Both victims and their assailants will be compensated $10 for each interview. Local criminal justice records will be collected on the assailants.

The data will be analyzed to determine the effects on subsequent violence of prosecution method, of apprehension method (summons versus warrant), and of rehabilitative programs. The design will also determine whether they have interactive or additive effects. It will also give preliminary estimates of differences between the victim-initiated cases and on the scene arrests. While the small "not guilty" group will not be randomly selected, it may provide initial estimates of retaliatory violence following an apparent false accusation.

ADMINISTERING ORGANIZATION: Indianapolis Prosecutor's Office

PRINCIPAL INVESTIGATOR: David A. Ford
Department of Sociology
Indiana University
425 Agnes Street
Indianapolis, Indiana 46202
(317) 264-2449
257-3049

SPONSORING AGENCY: National Institute of Justice

RESEARCH TEAM: Academic; Prosecutor's Office

TARGET GROUP: Men accused of violent behavior in conjugal relationships.

TREATMENTS: Two prosecution groups: rehabilitative probation; and fines or jail time. Two nonprosecution groups: Nolle, in which victims are advised of their right to drop the charges (about 70% are expected to do so); and diversion to a rehabilitative program. Assailants in the victim-initiated cases will receive either a court summons or an arrest warrant. The "not guilty" group will be those so found by the court, irrespective of their assigned prosecution strategy.

ASSIGNMENT METHOD: Cases are expected from two sources: on-scene arrests by the police and victim-initiated complaints. Assailants from the latter source will be randomly assigned to receive either an arrest warrant or a court summons. All of the cases will be randomly assigned to one of the two prosecution treatments (rehabilitative probation, fines or jail time) or the two nonprosecution treatments (e.g., Nolle; diversion to a rehabilitative program). All victims will be interviewed during the complaint, one month after the court date, and six months after the court date. The not-guilty cases will be determined by the court.

OUTCOME MEASURES: Subsequent levels of violence, and arrest; time until the next violent incident.

REPORTED IMPLEMENTATION PROBLEMS: The investigator is concerned about whether there will be enough cases, particularly the 220 on-scene arrests. While allowed under Indiana law, preliminary observations on police arrests suggested an insufficient rate for the experiment's 12-month time frame. The author proposed three strategies to raise the rates: 1) pressure the police officers to make more arrests; 2) use public service announcement to increase public awareness of the necessary procedures, and 3) include non-live-in boyfriend/girlfriend relationships. These could affect the generalizability of the findings.

IMPACT: NA

SOURCE: Ford, D. A. (1985) The Preventive Impacts of Arrest and Prosecution Policies on Wife Battery ("The Indianapolis Prosecution Experiment). Revisions to Proposal-852-CCT-04 to the Crime Control Theory and Policy Program. Washington, D.C.: The National Institute of Justice, October 30, 1985.

LITIGATION RESOLUTION EXPERIMENT

This experiment was designed to determine the effectiveness of court-supervised conferences and court orders reducing litigation. It was conducted from 1974 until 1977 in the United States Court of Appeals for the Second Circuit. This court hears appeals from federal cases in New York, Connecticut, and Vermont. The experiment used only cases which, in the judgment of the courts' attorney, appeared susceptible to settlement. Three hundred and two such cases were randomly assigned to either the conference group (225) or control group (77). In the conference group, attorneys for both sides met with a court appointed attorney. The court attorney would comment on the case's merits in an attempt to be a catalyst for a settlement. The court also issued an order establishing deadlines for critical events, such as filing the briefs. Contrary to common practice, failure to meet the deadline would lead to the case's dismissal. In the control group, both of these procedures were withheld. The number of appeals that were argued before the full panel of 3 appellate judges and the time from notice of appeals to termination were recorded. The judges were asked to rate the quality and clarity of the legal briefs. The lawyers involved were also surveyed about their level of experience and how they modified their case as the appeal progressed.

There were no major differences in the number of cases that were fully argued (54% of the conference group and 57% of the control group). The quality of the legal briefs differed in only one out of seven areas. Judges said the conference group were less likely to have redundant issues in their briefs. In terms of general quality however, the judges believed that the conference attorneys were better prepared and had a higher quality of appeal. The median time from start to conclusion of appellate litigation was shorter for the conference group (161 days) than for the control group (220 days). Most of the difference occurred in cases that were terminated before briefing and argument; the overall effect is only moderate. The only major difference between the two groups of attorneys and the progress of their briefs was that conference attorneys felt that the procedures had helped them to clarify the issues.

ADMINISTERING ORGANIZATION: U.S. Court of Appeals for the Second Circuit

PRINCIPAL INVESTIGATOR: Jerry Goldman
Political Science Department
Northwestern University
Evanston, IL 60202
(312) 491-2637

SPONSORING AGENCY: NA

RESEARCH TEAM: Academic, Federal Judiciary Center, Chief Judge of the Second Circuit

TARGET GROUP: Civil cases that seem susceptible to settlement as determined by the attorney. Important, constitutional questions, cases of first impression, and appeals of significant public moment were excluded.

TREATMENTS: 1) A conference group, in which attorneys for both sides attended a pretrial meeting under the supervision of the court's attorney. This group was also placed under an order establishing deadlines for critical events. Failure to meet the deadline led to dismissal. 2) A control group, which received neither of the above procedures.

ASSIGNMENT METHOD: Eligible cases were selected by the court's attorney. Three quarters were then randomly assigned to the conference group and one quarter to the control group. Assignment was case by case. All judges or attorneys involved in a case were surveyed.

OUTCOME MEASURES: How and when Appeals were terminated; clarity of legal briefs and oral arguments; presence of extraneous or redundant issues in legal briefs or arguments; omission of essential issues in briefs; preparation of counsel; quality of appeal; abandonment, addition, or clarification of issues by the attorneys; extent of adversary contact.

REPORTED IMPLEMENTATION PROBLEMS: Prior to the experiment, the chief judge had found the procedures to be very effective in a trial run on five cases. The principal investigator suggests that a judicial officer and not a court attorney, has the status and influence necessary to bring experienced advocates to a compromise. He also noted the unique experience of New York attorneys. About half of the attorneys, in each group, met to discuss settlement possibilities. This was attributed to the high concentration of attorneys in New York City. The procedures may have been more effective where litigation and attorneys are dispersed.

IMPACT: NA

SOURCE: Goldman, J. (1979) Resolution of appellate litigation: A controlled experiment. Evaluation Quarterly, 3, No. 4: 557-582.

PRETRIAL SETTLEMENT CONFERENCE EXPERIMENT

This experiment examined the impact of a judicially supervised plea bargaining conference on the processing and disposition of cases. It was conducted from January, 1977 until January, 1978 in the Criminal Division of the Circuit Court in the 11th Judicial Circuit of Florida (Dade County). The study involved six of the divisions' 12 judges. Of the 1074 randomly selected cases, 378 were assigned to use a pretrial conference supervised by one of three judges who would also officially hear the cases. The remainder of the cases served as a control group and were heard by any of the six judges. Approximately 40 control cases were selected from each judges' recently completed cases. They served as a baseline comparison for each judge. At arraignment, the judge informed the prosecution and defense as to their selection as an experimental case and scheduled a conference allowing time for pretrial motions and discovery. The defense attorney was required to confirm the conference three court days in advance. Defendants, victims, and police officers were all invited to the conferences that were chaired by the judge. Information on the offense, disposition, and timing of each case were collected from court records. Structured 20-minute interviews were attempted with victims, defendants, and police for all the cases.

Of the 378 experimental cases, conferences were actually held in 287 (76%). Of those, 26% settled during the conference and 46% tentatively settled. The conferences averaged about 10 minutes and had one or more lay persons in attendance (83%). The conference procedures reduced the time from arraignment to disposition, but they didn't change the overall rate of settlement. While the average sentence length decreased, a similar trend occurred in the control cases and thereby disassociated from the conference procedures. The conferences had no effect on the level of knowledge for either the victim or the police. However, both groups were more satisfied with the entire process.

ADMINISTERING ORGANIZATION: Criminal Division of the Circuit Court for the 11th Judicial Circuit of Florida (Dade County).

PRINCIPAL INVESTIGATOR: Wayne A. Kerstetter
Center for Studies in Criminal Justice
University of Chicago Law School
Chicago, Illinois

SPONSORING AGENCY: National Inst. of Law Enforcement and Criminal Justice.

RESEARCH TEAM: Center for Studies in Criminal Justice, Six judges from the 11th Circuit Court, State Attorney, and the Public Defender's Office.

TARGET GROUP: All noncapital felonies that survived arraignment were eligible for selection.

TREATMENTS: 1) The conference group in which a pretrial conference was held. The conference was supervised by the judge who would hear the case and was open to the defendant, victim, and police officers. 2) Three groups of control cases: the first was 40 recently completed cases for each judge that served as a baseline; the second was cases where the three conference judges never called for a conference; and the third was cases heard by the other three judges (who never held conferences).

ASSIGNMENT METHOD: All noncapital felony cases that survived arraignment were eligible. Forty recently completed cases were selected for each judge to serve as a baseline of their sentencing practices. The remaining cases were randomly assigned to one of the six judges. Three of the judges served as controls and proceeded normally after arraignment. The other three judges had their cases randomly assigned to either a conference group or a control group. Attempts were made to interview all defendants, victims, and police officers after the case was closed.

OUTCOME MEASURES: From court records: nature of the offense, type of disposition, method of disposition, timing of disposition. Conference discussions recorded by a staff member. From structured interviews with the victims, defendants, and police: knowledge of case, knowledge of final disposition, perception of process, attitude towards the process.

IMPLEMENTATION PROBLEMS: The researchers had a difficult time finding a site. Many officials were concerned that the conference would absorb inordinate amounts of judicial time; the presence of victims and defendants would lead to emotional or even violent confrontations; or that they would prevent the candid discussions between attorneys that are necessary for a settlement. None of these problems occurred.

IMPACT: NA

SOURCE: Heinz, A. M., and Kerstetter, W. A. (1979) Pretrial settlement conference: Evaluation of a reform in plea bargaining. Law and Society Review, 13(2): 349-366.

TELEPHONE BAIL RAISING EXPERIMENT

This experiment was designed to test the effect of installing telephones phones in detention centers and defendants' ability to raise bail. It was conducted from July, 1967 to August 1967 in the New York City Men's House of Detention ("The Tombs"), in lower Manhattan. The Vera Institute installed and staffed a telephone in the Tombs. Defendants could make two to three calls at a sitting; all calls were free, even long distance calls. To use the telephones, they filled out a short request form indicating whom they were going to call and why. Defendants used the phones for a variety of reasons (e.g., to raise bail, request clean clothes, or seek legal aid). The experiment involved 120 defendants randomly assigned to two groups. The first group was granted access to the telephone the next day. The second group had to wait seven days. Both groups had access to instructions on the require- requirements for release on bail. These were posted in English and Spanish by the phones. Researchers then monitored the delay until the defendants were released on bail if at all.

The group with immediate phone access was more likely to be released within seven days (25.4% versus 13.1%). When the group with delayed access was finally allowed to use the phones, their rate of release rose sharply to 24%. However, they were never able to quite catch up. Thirty-three percent of the immediate access group were eventually released on bail compared with only 25% of the delayed access group. Prior to the experiment, 44% of the defendants had no contact with the outside; 34% had tried to either send a letter or a message via a correctional officer; and 22% had made multiple contacts.

ADMINISTERING ORGANIZATION: Vera Institute of Justice

PRINCIPAL INVESTIGATOR: Kenneth J. Lenihan
Center for Policy Research
Vera Institute of Justice
New York, New York

SPONSORING AGENCY: Office of Economic Opportunity

RESEARCH TEAM: Vera Institute of Justice New York City Department of Corrections

TARGET GROUP: Defendants with bail determinations under $2000 and who had been in detention for 10 days or less. The experiment only considered defendants who requested telephone access.

TREATMENTS: 1) The immediate access group was allowed to use the telephone on the day after their requests. The delayed access group was allowed to use the telephone seven days after their requests. 2) The seven day delay was chosen to cancel any cyclical effect due to the actual day (Monday, Sunday, etc.) on which the request was made.

ASSIGNMENT METHOD: Only defendants with bail determinations under $2,000, who had been in detention for 10 days or less and who requested telephone access were considered. Defendants were classified by how long they had been in detention (1-5 days or 6-10 days) and the amount of their bail ($25-99; $100- 499; $500-999; $1000-$2000). Within each of the resulting eight groups, defendants were randomly assigned to either the immediate or delayed access group.

OUTCOME MEASURES: Whether the defendant was released on bail; when he was released relative to the telephone access request; and whether he made other attempts to contact someone outside or had visitors.

REPORTED IMPLEMENTATION PROBLEMS: Many of the telephone requests were unrelated to raising bail. While all requests were eventually granted, priority was given to defendants making their first call who specifically stated that they were trying to raise bail.

IMPACT: Implementation of the results was originally blocked. The principal investigator attributed this to the courts and the district attorney both of whom preferred to use detention as leverage in plea bargaining. Telephones were finally installed in response to a series of riots that erupted from over crowding. While problems have occurred since the phones were installed (e.g., prisoners dominating the phone, using it to harass witnesses, or to send candygrams) the warden acknowledges their value and stated that "the men should have them."

SOURCES: Lenihan, K. J. (1977) Telephones and raising bail: Some lessons in evaluation research. Evaluation Quarterly, 4: pp. 569-586.

COMMUNITY REHABILITATION (ELLSWORTH HOUSE) EXPERIMENT

This experiment was designed to test the relative effectiveness of a community-based rehabilitation program in terms of recidivism and employment. It was conducted from 1971 through 1973 in a San Mateo County facility called Ellsworth House. One hundred ninety men who had been sentenced for 4 months or more in the county jail were randomly assigned to either a community-based program (92) or the San Mateo county jail (98). The community-based program emphasized a therapeutic environment in which residents are taught to be responsible for their own improvement. Those assigned to the county jail served as a control group, which included an honor camp and a work-furlough facility. Residents of Ellsworth House are in a behavior modification program. They receive progressively more freedom (ability to go out unsupervised) as they show more responsibility. In order to gain 24-hour or 48-hour fur loughs, they must become active participants in the house's "community." All residents must participate in some constructive full-time activity: employ ment, college, or a training program. The house staff includes their probation officer, a vocational rehabilitation counselor and a research psychologist. After residents leave the house (completion of jail term) they return periodically for the duration of their probation. Men from the house and the control group were surveyed about employment and recidivism, 6 and 12 months after their release.

The initial group of Ellsworth residents had higher recidivism rates (27% versus 17%). After a change in their probation procedures, the two recidivism rates became comparable. Ellsworth residents are more likely to have a regular job after both 6 months (61% versus 45%) and a year (69% versus 56%). After a year, both groups had 15% with some employment; 15% (Ellsworth residents) 30% (control group) were unemployed.

ADMINISTERING ORGANIZATION: Ellsworth House

PRINCIPAL INVESTIGATOR: Dr. H. Richard Lamb
Chief of Rehabilitation Services
San Mateo County Department of
Public Health and Welfare
1050 Brittan Avenue
San Carlos, CA 94070

SPONSORING AGENCIES: California Council on Criminal Justice, San Mateo Probation Department, and the Rehabilitation Service of San Mateo County.

RESEARCH TEAM: Rehabilitation Service, Probation Department, Research Psychologist. Twenty residents of Ellsworth House.

TARGET GROUP: Men who have been sentenced for 4 or more months in the San Mateo County Jail. About 33% of the offenders were excluded because they were: considered severe escape risks, had a history of heavy involvement with narcotics, or presented a threat of uncontrollable physical violence.

TREATMENTS: 1) A control group, in which offenders were assigned to the San Mateo County Jail System. This system includes an honor camp and a work-furlough program. 2) Ellsworth House, a community-based rehabilitative program which is designed to teach its residents to be responsible for their own behavior.

ASSIGNMENT METHOD: The courts gave Ellsworth House staff a daily list of all men sentenced to jail. Selection was done on the day of sentencing using a random number table. Approximately one-fourth of the control group proceeded to enter the county jails' work-furlough program.

OUTCOME MEASURES: Recidivism; employment; living situation; criminal or delinquent associates; drug or alcohol abuse; and difficulties encountered by the probation officer in contacting the probationer.

REPORTED IMPLEMENTATION PROBLEMS: High recidivism rates occurred for the initial group. This was attributed to the weak probation program offered after the residents left. The probation officer at Ellsworth House was overburdened with getting the project underway. After the initial findings, the workload was reorganized to allow intensive one-on-one probation sessions. The recidivism rate subsequently dropped.

IMPACT: NA

SOURCE: Lamb, H. R. and Goertzel, V. (1974) Ellsworth House: A community alternative to jail. American Journal of Psychiatry 131:(1): 64-68.

CONNECTICUT RESTITUTION SERVICE PROJECT

This program was an "action research" effort to test the effects of using restitution as a criminal sanction in concert with the implementation of an innovative program in the criminal court. It was one of seven state programs which were part of a "National Evaluation of Adult Restitution Programs" conducted from 1976 through 1978. The program was intended to provide "restitution services" to the court in order to facilitate and increase the use of restitution as an alternative to incarceration. Community service was included as a form of "symbolic restitution." As it was designed to work, upon conviction the court could refer a case to the program in order to have " Restitution Specialists" determine the actual extent of loss suffered by a victim and the capability of the defendant to pay restitution to the victim. The program then submitted a report to the sentencing judge, which supplemented the routing PSI submitted by the Department of Probation if one had been ordered. If a defendant was ordered to pay restitution, the program also provided collection and accounting services.

From its beginning the program and experiment suffered from implementation problems. Slow growth in use of the program hampered both program development and conduct of the experiment. With few cases referred, program personnel resented seeing potential cases for investigation lost when they were randomly selected to be in the control group. Misunderstandings between program and research personnel continually arose despite careful and extensive efforts to avoid them. Even in control cases, for which the program did not provide restitution services, the sentences typically included restitution anyway. No significant findings were ever generated.

ADMINISTERING ORGANIZATION: Connecticut Judicial Department

PRINCIPAL INVESTIGATOR: Criminal Justice Research Center, Albany, NY

SPONSORING AGENCIES: Law Enforcement Assistance Administration
Connecticut Judicial Department; program staff

TARGET GROUP: Individuals convicted of misdemeanors and felonies

TREATMENTS: The intended treatments were ideally supposed to be restitution versus incarceration but in actuality the treatments were: 1) recommendation to the court that restitution was feasible, 2) recommendation to the court that restitution was not feasible, 3) no report submitted to the court. The latter group was to serve as the control or comparison group for the experiment but in those cases the court routinely ordered restitution anyway.

ASSIGNMENT METHOD: Cases referred to the program by the court were selected either as experimental cases for which services would be rendered or control cases for which no services would be rendered. Before a case was referred to the program, court personnel were instructed to call the program to verify that the program would handle the case. Program personnel checked the last three digits of the case's docket number against a list of random numbers which had been pre-assigned as

experimental or control numbers. If the case turned out to be a control the program advised the court that it could not provide services.

OUTCOME MEASURES: Recidivism of convicted defendant, change in attitudes by convicted defendants, success of restitution payments, satisfaction of victim with criminal justice system, change in incarceration rates.

IMPLEMENTATION PROBLEMS: Court planners and program personnel were never really committed to the experiment as other than a perfunctory condition required in order to obtain the grant. From the earliest meetings between court planners and the external evaluator there were conflicts sparked by the experiment. When growth in use of the program was much slower than expected the program director blamed the experiment. Design of both the program and experiment failed to accommodate the plea bargaining dynamics. Although designed and sponsored by the court as a service for judges, in most courts the prosecutors effectively controlled use of the program. Judges, prosecutors and defense attorneys routinely tried to use the program in ways which had not been anticipated and were inconsistent with the experimental design. For example, some prosecutors began to refer cases to the program before conviction as part of their plea negotiations which made the case ineligible for the experiment. In cases which were put through the experiment and selected as control cases, judges still usually ordered restitution in response to a plea agreement. In most cases referred, availability of the program typically led to the use of restitution as an additional sanction tacked onto probation rather than as a true alternative to incarceration. Conflicts emerged between the role of the "restitution specialist" and probation officers which led to strained relations in some courts and between directors.

IMPACT: During the program's operation, several courts and numerous judges increased their use of restitution in sentencing but not as alternative to incarceration. The research program was aborted when LEAA terminated funding for the Restitution program after GAO criticism of the overall national program. After the loss of federal funding, the program was merged into the Department of Adult Probation, where it quickly atrophied.

SOURCE: For information about this project, contact Dr. Richard Lovely, Department of Sociology, John Jay College of Criminal Justice, 445 W. 59th St., New York, NY 10019.

DENVER COUNTY DUI COURT EXPERIMENT

This experiment was designed to test the relative effectiveness of the common court sanctions for driving under the influence (DUI). It was conducted from January, 1969 until March, 1970, in the Denver County Court. All DUI cases for a given month were assigned to the same treatment. The three treatments rotated every month. They were: fines; conventional probation; and a set of counseling, educational, and therapeutic programs, demanded as a condition of probation. In all 487 cases, subsequent traffic reports and license records of the offenders were then monitored.

The judges frequently reviewed "exceptional" cases from the experiment. "Exceptional" cases turned out to be predominantly cases with legal counsel. These cases were more likely to receive fines than their assigned treatment. This occurred in 30% of the cases assigned to conventional probation and 40% of the cases assigned to therapeutic probation. The researchers found that there were no important differences between the three groups. They argued that despite a 50% failure to deliver the assigned treatment, the results are still valid, if less generalizable. To substantiate their view, they reanalyzed the data, grouping the defendants by the delivered treatment instead of the assigned treatment. Again, they found no major differences between the groups in terms of demographics, subsequent number of crashes, violations, or violation points. Nor did they find any differences in the average time from conviction to first subsequent incident.

ADMINISTERING ORGANIZATION: Denver County Court

PRINCIPAL INVESTIGATOR: Murray Blumenthal
Law School
University of Denver
Denver, Colorado

SPONSORING AGENCIES: National Highway Traffic Safety Administration

RESEARCH TEAM: Academic, Presiding County Court Judge

TARGET GROUP: Convicted DUI offenders

TREATMENTS: 1) Fines; 2) Conventional probation; 3) A set of counseling, educational, and therapeutic programs

ASSIGNMENT METHOD: Assignment was systematic, with all cases in a given month assigned to the same treatment. The treatment for the month rotated through the same cycle five times during the experiment's 15 months.

OUTCOME MEASURES: Subsequent number of crashes, violations, violation points; and time from conviction to first subsequent incident.

IMPLEMENTATION PROBLEMS: The two trial judges who actually had to administer the treatments did not "buy in" to the experiment. No rules were set on the use of the "exceptional case" criteria for changing the treatment disposition. In a scheduled election, the presiding judge who requested the experiment was replaced by a member of the opposition party. While the latter was not opposed to the experiment, he did not aggressively support it.

IMPACT: NA

SOURCE: Ross, H. L. and Blumenthal, M. (1974) "Sanctions for the drinking driver: An experimental study." Journal of Legal Studies 3: 53-61.

DRUNK DRIVING AND JAIL TIME INDIRECT EXPERIMENT

This experiment was designed to test the specific deterrent effects of a two day jail sentence for persons convicted of drunk driving. It was an "indirect" experiment because the assignment of treatment took place without any conscious guidance from researchers. Instead, researchers took advantage of natural random assignment of different treatment some years after it had occurred.

In 1982, Hennepin County (Minneapolis) judges agreed to adopt a policy of sentencing all drunk drivers to at least two days in jail. Each judge followed the policy to a greater or lesser extent. With the help of the Minnesota SWI Criminal Justice System Task Force, two judges were identified who had taken extreme positions on the policy. On judge sentenced 75% of his convicts to jail, while the other judge sentenced 24% to jail. The two judges both added fines to about 9090%% of their cases, although the "no jail" judge imposed an average fine of $3453, twice as large as the "jail" judge's average fine of $156. The no jail judge also stayed 45% of the fines, while the jail judge only stayed 12%.

The approximately 250 cases assigned to each judge by court clerks at random had no differences in terms of legal representation (percent private lawyer, public defender, and no lawyer), percent male, prior alcohol-related and non-related driving record, and blood alcohol content level at the instant arrest. The only observed difference was that the no jail judge had a slightly younger group of offenders, which could have occurred by chance.

A two year follow-up of the convicts' statewide driving records showed no significant differences between the two groups. The primary measure was repeat police contact for alcohol-related driving. Twenty percent of the jail judge group, but only 16 percent of the no-jail judge group had repeat police contact over the two years after each case was sentenced; the difference was small enough to have been a chance effect. Other measures of driving also failed to show significant differences.

ADMINISTERING ORGANIZATION: Hennepin County (MN) Courts

PRINCIPAL INVESTIGATORS (In alphabetical order): Brian Forst, Police Foundation, 1001 22nd St. N.W, Wash., D.C. 20007, 202-833-1460; Lawrence W. Sherman, Crime Control Institute, 2125 Bancroft Pl., N.W., Wash., D.C., 20008, 202-797-7410

SPONSORING AGENCY: National Institute of Justice

RESEARCH TEAM: Police Foundation, Minnesota SWI Criminal Justice System Task Force

TARGET GROUP: First and repeat drunk driving offenders

TREATMENTS: Jail and fine, no jail and longer fine (45% stayed)

ASSIGNMENT METHOD: At random by court clerks (normal processing)

OUTCOME MEASURES: Accidents, all driving violations, alcohol-related police contacts

REPORTED IMPLEMENTATION PROBLEMS: NA

IMPACT: NA

SOURCE: Sherman, L.W., P.R. Gartin, S. Miler and D. Doi, "Failed Deterrence: The Specific Sanction Effects of Jail Time for Drunk Driving", Paper presented to the American Society of Criminology, November, 1986.

Martin, S., S. Annan, and B. Forst, "Deterring the Drunk Driver." Police Foundation, 1986.

DWI REHABILITATIVE SANCTIONS EXPERIMENT

This experiment was designed to test the effectiveness of probation and educational/therapy programs on the rearrest rates for drunk drivers (DWI). It was conducted from September, 1976 until December, 1980 in the city of Memphis and the surrounding areas of Shelby County. First time DWI offenders were referred from a court specializing in DWI cases. The court referred offenders as a condition of their diversion (no prior felonies or misdemeanors) or probation. During the DWI Probation department intake, the 4,126 offenders were classified as problem (1621) or social (2505) drinkers. One quarter of the offenders in each drinking group were randomly assigned to one of four groups: control group in which no further contact was required; a supervision group, in which offenders met monthly with a probation officer; an education group, in which offenders attended a 10-hour alcohol safety course; and a supervision and education group which required both. Problem drinkers assigned to one of the education groups were also required to attend 8 therapy sessions that emphasized assertiveness training. Each offender was followed up for a minimum of two years. In the first year, they were tracked daily by the Probation Department's staff. Arrest records for the second year were obtained from the Memphis and Shelby County Criminal Justice Computer System. Additional records were collected from the Memphis Police Department's Breath Testing Unit. The study examined: subsequent arrests for DWI, subsequent other arrests; and time until the next subsequent arrest. It also attempted to compare the four groups on the basis of their relative severity of treatment.

For social drinkers, there were no major differences between the four treatment groups in terms of DWI or non-DWI rearrests. For problem drinkers in the combined treatment group, there was approximately a 22% reduction in non-DWI offenses but no significant changes in DWI arrests. However, for non-DWI offenses, the rearrest rates went down as the treatment severity went up. The supervision and education group had the lowest non-DWI rearrest rate.

ADMINISTERING ORGANIZATION: Tennessee DWI Probation Department

PRINCIPAL INVESTIGATOR: Robert T. Holden
Department of Sociology
Yale University
New Haven, CT

SPONSORING AGENCIES: National Highway Traffic Safety Administration and the National Institute of Mental Health.

RESEARCH TEAM: Academic, Tennessee DWI Probation Follow-up Demonstration Project, and the Tennessee Probation Department.

TARGET GROUP: Persons arrested for driving while intoxicated (DWI) in Memphis and the surrounding Shelby County. Only persons with no prior DWI convictions were eligible; individuals who were not residents of Shelby County, were servicemen, or who had health problems, were also excluded.

TREATMENTS: 1) Education/therapy in which social drinkers were required to attend a 10-hour alcohol safety course. Problem drinkers were required to also attend eight 1 1/2 hour therapy sessions (assertiveness training); 2) Supervision in which offenders had to meet monthly with a probation officer for a year. There were four groups receiving a combination of these treatments: In the first group, offenders were released without either treatment; in the second, they received the education/therapy treatment; In the third, they received supervision; and in the fourth, offenders received both treatments. Participation was mandatory and failure to comply resulted in a one- year jail sentence.

ASSIGNMENT METHOD: Offenders were referred to the DWI, Probation Department by the Memphis City Court, Division VIII; a court which hears only DWI cases from Shelby County; The Probation Department assigned the offenders to one of the four groups using randomly predetermined lists. During intake, offenders were qualified as problem or social drinkers using the Mortimer-Filkins inter view and the individual's blood alcohol concentration at the time of arrest. Problem drinkers who were assigned to one of the two groups receiving the education treatment were also required to attend therapy sessions.

OUTCOME MEASURES: Rearrests for DWI; rearrest for non-DWI offenses; time until subsequent arrest.

REPORTED IMPLEMENTATION PROBLEMS: The principal investigator notes that for 85% of the offenders, this was their first arrest and that the arrest, court procedures, and the threat of a jail sentence were probably severe enough to eliminate any minor differences due to the treatment. He speculated that many offenders were actually relieved merely to be "sentenced" to attend DWI school or report to a probation counselor. Nondelivery of treatment was generally low (2.6%) but varied considerably by group (from .3% for social drinkers in the control group to 9.2 % for problem drinkers in the supervision and education group).

IMPACT: NA

SOURCE: Holden, R. T. (1983) Rehabilitative sanctions for drunk driving: An experimental evaluation. Journal of Research in Crime and Delinquency, 20: 55-72.

SAN QUENTIN SQUIRES EXPERIMENT

This experiment was designed to determine what impact a juvenile awareness program in an adult prison had on the attitudes and subsequent behaviors (recidivism) of program participants. It was conducted between 1978-1979 in San Quentin Prison. A total of 108 delinquent males between the ages of 14 to 18 (mean=16.3) were randomly assigned to the San Quentin Squires program or a control group. All of the youths came from remedial youth camps and had prior records (7.4 arrests on the average). The squires program involved confrontive rap sessions, guided tours of the prison, personal interaction with prisoners and a pictorial review of prison violence. Groups of 20 youths went through the 3-day program on consecutive Saturday mornings. One week before the program commenced both groups were pretested on a variety of attitudinal scales. They were retested within one week of the program's end. The subsequent behaviors, particularly recidivism, were then monitored for 12 months.

Program participants had less delinquent attitudes towards police, crime, and a variety of psychological indices. The Squires program did not reduce delinquent behavior for all participants. However, both positive and negative effects were found for two subgroups. Moderately delinquent youths, particularly Caucasians, had less subsequent offenses than their control counterparts. They were also less likely to commit subsequent drug offenses on the average (0 vs. 4). However, the findings for older participants was less encouraging: while they were arrest-free longer than their control counterparts, they were found to have committed more serious delinquency in the subsequent 12 months.

ADMINISTERING ORGANIZATION: San Quentin Prison

PRINCIPAL INVESTIGATOR: Roy V. Lewis
California Youth Authority
Sacramento, CA.

SPONSORING AGENCY: California Youth Authority

RESEARCH TEAM: San Quentin Squires Program, California Youth Authority, and Youth camps from Los Angeles and Contra Costa countries.

TARGET GROUP: Youths, 16-17 years of age, who have a record of delinquency.

TREATMENTS: 1) San Quentin Squires program, in which youths met for three consecutive Saturday mornings with inmates in San Quentin Prison. The program involved confrontive rap sessions, guided tours of the prison, personal interactions with the prisoners, and a pictorial review of prison violence. 2) Control group in which youth only attended the remedial camp program.

ASSIGNMENT METHOD: Eligible youths were recruited from 6 remedial boy's camps in Los Angeles and Contra Costa Counties. Approximately half were randomly assigned to the San Quentin's Squires program. Only subjects who completed the entire three-day sequence were included in the analyses.

OUTCOME MEASURES: Attitude scales: four delinquency scales (measuring attitudes towards police, school, crime, and prison); one composite index of delinquent attitudes; two semantic differential scales; an attitude towards camp scale; and the Glueck Social Prediction Scale. Five behavioral outcome measures over the following 12 months: 1) number of subsequent arrests; 2) number of subsequent charges; 3) number of subsequent charge types (e.g. crimes against persons or property, drug offenses, minor offenses, and status offenses); 4) average severity of subsequent charges; and 5) length of time to first arrest.

REPORTED IMPLEMENTATION PROBLEMS: Despite the target of 16-17 years of age, a few 15 and 18-year-olds were accidentally placed into the initial pool prior to randomization. One of the experimental youths didn't complete the entire sequence and was consequently excluded from the analyses.

IMPACT: NA

SOURCE: Lewis, R. V. (1983) Scared straight - California style: An evaluation of the San Quentin Squires Program. Criminal Justice and Behavior, 10(2): 209-226.

SOCIAL SERVICES AND PROBATIONER RECIDIVISM EXPERIMENT

The experiment was designed to see if intensive probation was better than conventional probation in terms of probationer recidivism. It was conducted in Wayne County, Michigan from February, 1976 until May, 1977. The probationers came from two sources: The Recorder's Court cases, which were from Detroit, and the Circuit Court cases, which came from the rest of the county. To be included, probationers had to be between the ages of 18-30 years of age and on probation for nonviolent felonious property offenses. The 503 eligible probationers were randomly assigned to either a conventional probation group (227), or an intensive probation group (276) which was operated by a program called START. The control group was then assigned by the standard process to any of the approximately 200 probation officers serving the city and county. The START program began with an elaborate needs analyses and emphasized counseling, extensive referrals and lower caseloads. Probationer records were then tracked and a follow-up survey was conducted a year after the termination of the experiment.

Intensive probation was found to offer little or no improvements over conventional procedures. When probationers received only sparse services or referrals from the probation department, they secured such services on their own. For felonious property offenders, regular probation was less expensive than intensive probation, yet equivalent with regard to recidivism. The caseloads of START probation officers were about 1/2 of the load for the control group. This initially allowed them to give probationers more attention (a mean of 2.44 visits per month versus 1.32 for the controls). This difference however, disappeared within one year. Probationers in the START program also received more referrals per month than the control group (2.99 versus .58) and were more likely to make use of those services (85% versus 63%). Yet in the follow-up interview, the control group reported seeking out these services independently. The only remaining differences in actual services used was for crisis and credit counseling. There were no major differences in terms of recidivism of felonies (START 19.3%, control 14.7%). Similar findings were found using the number of convictions or the time until the first subsequent conviction.

ADMINISTERING ORGANIZATION: Project START

PRINCIPAL INVESTIGATOR: Gary M. Lichtman
Department of Psychology
Wayne State University
Detroit, Michigan

SPONSORING AGENCIES: Law Enforcement Assistance Administration, Department of Justice, Office of Criminal Justice Program grants.

RESEARCH TEAM: Academic, Detroit Recorder's Court, Wayne County (MI) Circuit Court, Project START

TARGET POPULATION: Probationers between the ages of 18 to 30 years of age, who were on probation for nonviolent felonious property offenses (e.g. breaking and entering, auto theft, shop lifting, malicious destruction of property, and passing bad checks). Known users of hard drugs were excluded.

TREATMENTS: 1) Conventional probation, delivered by the existing 200 city and county probation officers; and 2) intensive probation, delivered by project START. The latter emphasized more needs analyses, referrals, counseling, and lower caseloads.

ASSIGNMENT METHOD: The two courts called the researchers daily with information on all newly sentenced offenders. Within the list submitted by each court, half the probationers were assigned to each of the treatments. Actual assignment was made by flipping a coin. By randomizing within each court, the researchers were avoiding any effects due to race; the Recorder's Court cases were 80% black, while the circuit court cases were 80% white. One year after selection had ended, an attempt was made to survey all of the probationers; each was offered $5 to be interviewed.

OUTCOME MEASURES: Recidivism in terms of felonies or misdemeanors; number of subsequent convictions; time to next subsequent conviction; employment status; Probation Officer caseload; number of referrals received; number of referrals utilized; number of social services utilized.

REPORTED IMPLEMENTATION PROBLEMS: While disproportionately more of the probationers went into the intensive group (276 versus 227), the experiment was able to include all of the county's 503 eligible probationers. The researchers reported high compliance from the courts and said they made no attempt to influence assignment. While only 135 probationers could be reached for the follow-up interview, all but 3 were willing to respond.

IMPACT: NA

SOURCES: Lichtman, C. M. and Smock, S. M. (1981) The effects of social services on probationer recidivism: A field experiment. Journal of Research in Crime and Delinquency, 18: 81-100.

PRISON GROUP COUNSELING EXPERIMENT

This experiment was designed to test the impact of group counseling on prisoner's attitudes, cooperation, and parole behavior. It was conducted from January, 1962 until June, 1967 in the then newly opened California Men's Colony-East (CMCE). The experiment utilized 3 of the Colony's 4 quadrangles and studied 965 inmates who were randomly assigned to one of three counseling groups or two control groups. The counseling groups were: Voluntary small group counseling; mandatory small group counseling; and mandatory large group counseling. Each group employed therapeutic techniques designed to help inmates release their aggression and nature. Their techniques, already in use throughout the California Correction system, were introduced by Dr. Norman Fenton in 1944. The two control groups came from two sources: inmates assigned to be mandatory controls and those assigned to voluntary counseling who chose not to participate. Treatment conditions were segregated between the facility's three quadrangles. Only inmates who received at least 6 months of the treatment were included in the study. Their subsequent attitudes and behaviors were monitored for three years, particularly the recidivism of parolees. In addition, participating prison personnel were surveyed about their attitudes towards counseling and its effect on the inmates.

While incarcerated, inmates' attitudes appeared to be continuously shifting in both directions. The counseling group continued to follow antisocial inmate attitude norms. There were no significant differences between the control and counseling participants in terms of: number of problems; severity of problems; returns to prison; jail time; alcohol or drug abuse; or employment. These findings held each time they were measured at 6, 12, 24, and 36 months after release. In addition, no significant differences were found between those inmates who volunteered and those who had mandatory counseling.

ADMINISTERING ORGANIZATION: California Department of Correction

PRINCIPAL INVESTIGATOR: Gene C. Kassebaum

SPONSORING AGENCY: National Institute of Mental Health

RESEARCH TEAM: California Department of Corrections, California Men's Colony-East administrators and staff, California Adult Authority, Academics.

TARGET GROUP: Inmates in a medium security prison.

TREATMENTS: 1) Mandatory large group counseling in which all 50 men in a section met four times a week for an hour. The leaders were administrative personnel with supplemental training. 2) Mandatory small group counseling, in which groups of 10-12 inmates met twice a week with correctional counselors. 3) Voluntary small group counseling, in which groups of 10-12 inmates met weekly and were led by members of the staff who received only the normal training. 4) Voluntary Controls, inmates assigned to voluntary counseling, who declined to participate in experimental counseling programs. 5) Mandatory Controls, who received no group counseling.

ASSIGNMENT METHOD: Only incoming inmates who would be in prison for at least 6 months and available for parole within 3 years were considered. The names of the eligible inmates were then rank ordered by the last two number of their departmental serial number. The first inmate was assigned to either mandatory or voluntary small group counseling. The second inmate was assigned to mandatory large counseling. The third inmate was assigned to voluntary counseling. The fourth inmate was assigned to mandatory controls. This process was repeated through the entire list. Inmates who were classified by departmental headquarters as trouble makers, postpsychotic, arsonists, or aged were excluded. The chief administrators of each quadrangle were allowed to exclude or transfer up to 5% of their men.

OUTCOME MEASURES: Prison record abstracts covering treatment exposure, criminal and personal background, prison activities, rule violations, and release plans. Inmate value questionnaire about inmate code endorsement. Group counseling questionnaire. California psychological inventory. Interviews. Subsequent (to parole) rates of minor legal problems, major problems, returns to prison, jail time, and drug or alcohol abuse.

REPORTED IMPLEMENTATION PROBLEMS: Final funding arrangements were delayed. While the study was unable to commence, CMCE officials implemented the random assignment procedure to living quarters (which determined the nonvoluntary treatments).

IMPACT: The system continued to use some group counseling. The California legislature interpreted the study as showing large scale prisons to be ineffective and that no further funds should be provided for the construction of state prison facilities.

SOURCE: Kassebaum, C., Ward, D., and Wilner, D. (1971) Prison Treatment and Parole Survival: An Empirical Assessment. New York: John Wiley and Sons, Inc.

TRANSITIONAL AID RESEARCH PROJECT (TARP)

The experiment was designed to assess the impact of extending unemployment benefits to recently released prisoners on their subsequent employment and recidivism. It was conducted simultaneously in Texas and Georgia from 1976 until 1977. About 1000 ex-offenders in each system were randomly assigned to one of 6 groups: 4 experimental and 2 control. The experimental groups varied in the number of weekly payments that would be provided (26, 13, and 0) and the rate at which payments would be reduced for every dollar the offender earned from a new job (100% and 25%). The experimental groups also received job placement counseling, and subsidies for work related expenses (e.g., tools, safety equipment). The control groups received no additional assistance. One control group received both interviews and record checks, the second only the record checks. Ex-offenders were eligible for $70 a week in Georgia and $63 a week in Texas. The program benefits were managed by the existing public servants within each state, and monitored by the Transitional Aid Research Project (TARP) staff. Ex-offenders in all but the second control group were interviewed immediately prior to release, after 3 months, 6 months, and 12 months. Criminal justice and unemployment insurance file records on all of the ex-offenders were collected up until 12 months after their release.

No differences in the average number of new arrests were found between the various treatments or between the treatments and the control groups. There was some evidence, however, that the TARP payments reduced the likelihood of recidivism by up to 30%, but that the effects were being undermined by the work disincentives built into the treatments (e.g., the 25% and 100% paybacks). There was also evidence that the payments allowed the experimental ex-offenders to look longer for better paying and more stable jobs.

ADMINISTERING ORGANIZATION: NA

PRINCIPAL INVESTIGATOR: Peter H. Rossi
Social & Demographic Research Institute
University of Massachusetts
Amherst, Massachusetts

SPONSORING AGENCY: Employment and Training Administration of the U.S. Department of Labor.

RESEARCH TEAM: Transitional Aid Research Project (TARP), Texas prison system, Georgia prison system, Academics.

TARGET GROUP: Male and female prisoners from the State of Texas and Georgia who were about to be released.

TREATMENTS: Four experimental groups: Group 1, which received 26 weekly payments lost one dollar of benefits for every dollar they were able to earn from a job (i.e., 100% tax); Group 2, which received only 13 weeks of payment and was also taxed 100% on earned income; Group 3, which received 13 weeks of payments but was only taxed 25% of their benefits for earned income; and Group 4, which only received job placement assistance. Ex-offenders were eligible for $70 a week in Georgia and $63 a week in Texas. Two controls: One which was interviewed and one which wasn't. Record checks were conducted on all 6 groups.

ASSIGNMENT METHOD: Approximately 2000 men and women who were released in the states of Georgia and Texas during a 6-month period of 1976 were randomly assigned to one of four treatment groups and two control groups. All but the second control group were interviewed.

OUTCOME MEASURES: Ex-offenders were interviewed about their post-incarceration experience including: employment, marital status, living arrangements, major illnesses, accidents, and consumer purchases. Records were also collected from: the two prison systems, local criminal justice agencies, and their unemployment insurance files. They included information on their past and subsequent employment, incarceration, arrest for property offenses and arrests for nonproperty offenses.

IMPLEMENTATION PROBLEMS: Despite each state organizing its own data collection activities, and recruiting and training an ad hoc interviewing staff, high completion rates were experienced; the lowest completion rate for any wave of interviewing was 84%. Work disincentive effects appear to have been strong because of the limited employment opportunities available to the TARP participants. Wages of $100 to $150 per week (before taxes) earned at hard and/or unpleasant tasks are unlikely appealing, compared to $63 or $70 per week obtained without working.

IMPACT: NA

SOURCE: Berk, R. A., Lenihan, K. J., and Rossi, P. H. (1980) Crime and poverty: Some experimental evidence from exoffenders. American Sociological Review, 45(Oct): 766-786.

Rossi, P.H., Berk R.A., and Lenihan, K.J. (1980) Money, Work and Crime: Some Experimental Results. New York: Academic Press.

WORK RELEASE AND RECIDIVISM EXPERIMENT

This experiment was designed to test the effectiveness of prison work release programs in terms of recidivism. It was conducted within the Florida Division of Corrections system between July 1, 1969 and December 31, 1969. Two-thirds (188) of the prisoners were assigned to a work release program and one-third (93) to a control group. Releasees spent from 2-6 months in the program. Non releasees stayed in the prison program they had been in. Recidivism was measured using self reports, FBI records, and Florida Division of Correction files. It was operationalized in a variety of ways, including subsequent: arrests, bookings, charges, convictions, and reincarcerations. Subsequent crimes were also scaled in terms of their severity and reincarcerations in terms of their sentences.

No major differences were found between the two groups in terms of their subsequent rates of being arrested, booked, charged, convicted, reinstitutionalized within Florida, or reinstitutionalized anywhere in the United States. No relationship was found between work release and charge severity of subsequent offenses or their sentence length. Releasees remained free prior to reincarceration slightly longer than non-releasees on the average (18.25 months versus 15.80 months). The rate of reincarceration anywhere in the U.S. (from FBI records) was higher than for within Florida alone. However, there were no major differences between releasees and non-releasees for either measure. No relationship was found between recidivism and work release controlling for a variety of demographic and criminal history variables.

ADMINISTERING ORGANIZATION: Florida Division of Corrections

PRINCIPAL INVESTIGATOR: Gordon P. Waldo
School of Criminology
Florida State University

SPONSORING AGENCY: NA

RESEARCH TEAM: Academic, Florida Division of Corrections

TARGET GROUP: Prisoners within the Florida Corrections system who were eligible for its existing work release program.

TREATMENTS: 1) A work release group, in which prisoners participating in Florida's existing work release program for 2-6 months at the end of their sentence; 2) a control group in which eligible prisoners stayed in the same program they had been in.

ASSIGNMENT METHOD: Every third person meeting the minimal requirement was placed in the control group; the other two-thirds were placed in work release.

OUTCOME MEASURES: Self-reports on subsequent rates of: arrest; bookings; convictions; and reincarcerations. FBI records on subsequent rates of: arrests; charges; reincarcerations; and felonies. Division of Correction records on subsequent rates of reincarcerations and sentence lengths. Charge severity data computed from FBI and Division of Correction records.

REPORTED IMPLEMENTATION PROBLEMS: Because work release was an established program in Florida, there were problems of resentment and disappointment on the part of the nonreleasees. Correction administrators had to resist pressure from inmates, superintendents (who nominated the inmates), and even local congressmen.

IMPACT: NA

SOURCE: Waldo, G. P. and Chiricos, T. G. (1974) Work release and recidivism: empirical evaluation of a social policy. Evaluation Quarterly, 1(1): 87-108.

APPENDIX A
WORKSHOP PROGRAM

WORKSHOP ON CRIMINAL JUSTICE EXPERIMENTS

Le Meridien Hotel
New Orleans, Louisiana

March 13-14, 1987

Friday, March 13 — *Morning Session* — *Room - Ile de France III*

9:00 **Welcome and Introduction**

James K. Stewart, Director, National Institute of Justice
Richard Lempert, Chair, Working Group on Field Experimentation in Criminal Justice
Anthony Bouza, Chief of Police, Minneapolis

9:30 **Discovering What Works: Uses and Limits of Experiments**

Presenters: Peter Rossi, University of Massachusetts
Lawrence Sherman, Crime Control Institute and University of Maryland
Moderator: Richard Lempert, University of Michigan

10:30 BREAK

10:45 **Discovering What Works: Designing Experiments**
Richard Berk, University of California, Santa Barbara

11:15 **Legal and Ethical Issues in Criminal Justice Experiments**

Presenter: Charles Wellford, University of Maryland
Discussants: Bruce Beaudin, Judge, Superior Court of the District of Columbia
Delbert Elliott, University of Colorado
E. Michael McCann, District Attorney, Milwaukee
Moderator: Shari Diamond, University of Illinois at Chicago and Sidley & Austin

12:15 **Design Responses to Legal and Ethical Concerns**
Richard Berk, University of California, Santa Barbara

12:30 LUNCH — Frontenac Room

Friday, March 13 *Afternoon Session*

2:00 Discussion Groups: Opportunities and Pay-offs for Policy Experiments

Goal: To identify promising experiments in areas of practice and to discuss design and implementation issues that would arise in carrying out experiments.

1. **Experiments in Policing and Charging** Room 803
   Discussants: Anthony Bouza, Chief of Police, Minneapolis
   George Napper, Commissioner, Department of Public Safety, Atlanta
   Moderator: Albert Reiss, Yale University

2. **Experiments in Pretrial Release, Prosecution and Diversion** Room 804
   Discussants: Ron Clark, Chief Deputy, Criminal Division, King County Prosecutor's Office, Seattle
   Deborah Daniels, Chief Counsel, Office of the Prosecuting Attorney, Indianapolis
   Moderator: Jeffrey Roth, National Research Council

3. **Experiments in Adjudication and Sanctioning** Room 805
   Discussants: Larry Polansky, Court Administrator, Superior Court for the District of Columbia
   Gerald Wetherington, Chief Judge, Eleventh Judicial Circuit, Dade County, Florida
   Moderator: Shari Diamond, University of Illinois at Chicago and Sidley & Austin

4. **Experiments in Supervision of Offenders** Room 806
   Discussants: Malcolm MacDonald, President, American Probation and Parole Association
   Cecil Steppe, Chief Probation Officer, San Diego
   Moderator: James Austin, National Council on Crime & Delinquency

3:15 BREAK

3:30 **Concurrent Roundtables: Experiments and Alternatives**

1. **Randomized Experiments: The Case of Milwaukee** Room 803

   Panel: E. Michael McCann, District Attorney, Milwaukee
   Robert Ziarnik, Chief of Police, Milwaukee
   Dean Collins, Captain, Milwaukee Police Department
   Moderator: Lawrence Sherman, University of Maryland and Crime Control Institute

2. **Alternatives to Randomized Experiments** Room 804

   Panel: David Farrington, Cambridge University
   William Bowers, Center for Applied Social Research, Northeastern University
   David Chambers, University of Michigan
   Moderator: Daniel Rubinfeld, University of California, Berkeley

5:00 RECEPTION AND CASH BAR Orleans Room

9:00 **Randomized Experiments: Anticipating Problems and Finding Solutions**

A. Frustrations and Opportunities in the Field
Mary Toborg, Toborg Associates, Inc., Washington, D.C.
Richard Lovely, John Jay College of Criminal Justice

B. Frustrations and Opportunities: A Practitioner's View
Stephen Goldsmith, District Attorney, Indianapolis

C. Damage Control
Robert Boruch, Northwestern University

Moderator: Robert Boruch

Audience Discussion

10:45 BREAK

11:00 **Continuation of Discussion Groups: Managing Field Experiments**

Goal: Continue discussion of specific experiments and focus on problems that could arise in managing experiments in the community and within agencies.

1. **Experiments in Policing and Charging** Room 803
Discussants: Anthony Bouza, Chief of Police, Minneapolis
George Napper, Commissioner, Department of Public Safety, Atlanta
Moderator: Albert Reiss, Yale University

2. **Experiments in Pretrial Release, Prosecution and Diversion** Room 804
Discussants: Ron Clark, Chief Deputy, Criminal Division, King County Prosecutor's Office, Seattle
Deborah Daniels, Chief Counsel, Office of the Prosecuting Attorney, Indianapolis
Moderator: Jeffrey Roth, National Research Council

3. **Experiments in Adjudication and Sanctioning** Room 805
Discussants: Larry Polansky, Court Administrator, Superior Court of the District of Columbia
Gerald Wetherington, Chief Judge, Eleventh Judicial Circuit, Dade County, Florida
Moderator: Shari Diamond, University of Illinois at Chicago and Sidley & Austin

4. **Experiments in Supervision of Offenders** Room 806
Discussants: Malcolm MacDonald, President, American Probation and Parole Association
Cecil Steppe, Chief Probation Officer, San Diego
Moderator: James Austin, National Council on Crime and Delinquency

12:00 LUNCH Rosalie/St. Claude Room

1:00 p.m. **Interpreting and Utilizing Results from Field Experiments: Two Case Studies**

Sally Hillsman, Vera Institute of Justice
Richard Lempert, University of Michigan

Moderator: Albert Reiss, Yale University

2:00 **Assessing the Role of Field Experiments in Criminal Justice**

Goal: To provide a forum for policymakers and research sponsors to exchange views on the use of field experiments, and to offer some suggestions about launching field experiments, drawing on the earlier workshop discussions.

Malcolm MacDonald, President, American Probation and Parole Association
David Tevelin, Executive Director, State Justice Institute
James K. Stewart, Director, National Institute of Justice

Moderator: Albert Reiss, Yale University

Audience Discussion

3:30 ADJOURN

APPENDIX B
PARTICIPANTS
WORKSHOP ON CRIMINAL JUSTICE EXPERIMENTS
March 13-14, 1987

Tom Arnold, Chief, Metro-Dade Police Department, Miami

James Austin, Director of Research, National Council on Crime and Delinquency, San Francisco

Bruce D. Beaudin, Associate Judge, Superior Court, District of Columbia

Cornelius J. Behan, Chief of Police, Baltimore County, Maryland

Richard Berk, Professor of Sociology, University of California, Santa Barbara

Clarina M. Blackden, Court Programs Analyst, Philadelphia

Robert Boruch, Professor of Psychology, Northwestern University

Anthony Bouza, Chief of Police, Minneapolis

William Bowers, Director, Center for Applied Social Research, Northeastern University

Michael D. Bradbury, District Attorney, Ventura County, California

John Brandl, Professor, University of Minnesota, and Senator, State of Minnesota

Cathy Bruemmer, Special Assistant to the Director, National Institute of Justice

Joe Cannon, District Judge, State of Colorado

Paul Cascarano, Assistant Director, National Institute of Justice

David Chambers, Professor of Law, University of Michigan Law School

Ronald H. Clark, Chief Deputy, Criminal Division, King County Prosecutor's Office, Seattle

Jacqueline Cohen, School of Urban and Public Affairs, Carnegie-Mellon University

Dean J. Collins, Captain, Milwaukee Police Department

Christine Curtis, Assistant Director, Criminal Justice Unit, San Diego Association of Governments

Deborah J. Daniels, Chief Counsel, Prosecutor's Office, Marion County, Indiana

Mike Dennis, Department of Psychology, Northwestern University

Shari Diamond, Professor of Psychology, University of Illinois at Chicago, and Attorney at Law, Sidley & Austin

Robert N. Dick, Chief of Police, Tulsa Police Department

Frank S. Duling, Chief of Police, Richmond

Gaylene Dumouchel, Administrative Secretary, National Research Council

Delbert S. Elliott, Associate Program Director, Institute of Behavioral Science, University of Colorado

A.W. Emerson, Sergeant, Minneapolis Police Department

Lee C. Falke, Prosecuting Attorney, Montgomery County, Ohio

David P. Farrington, Professor, Institute of Criminology, Cambridge University

Joel Garner, National Institute of Justice

Peter S. Gilchrist III, District Attorney, Charlotte, North Carolina

George Gish, Court Administrator, Recorder's Court, Detroit

Stephen Goldsmith, Prosecuting Attorney, Indianapolis

Harvey M. Goldstein, Assistant Director for Probation, New Jersey Administrative Office of Courts

Ann S. Harrington, Assistant State's Attorney, Montgomery County, Maryland

John C. Hendricks, General Manager, Kentucky Pretrial Services, Frankfort

Sally T. Hillsman, Director of Research, VERA Institute of Justice, New York

Andrea B. Huff, Research and Development Bureau Commander, Charlotte Police Department

Richard D. Huffman, Superior Court Judge, San Diego County

Don Hunter, Deputy Chief, Collier County Sheriff's Office, Florida

Terri A. Jackson, Director, Pretrial Services, Phoenix

Timothy M. Kenny, Deputy Chief, Criminal Division, Prosecuting Attorney's Office, Wayne County, Michigan

Richard O. Lempert, Professor of Law and Sociology, University of Michigan Law School

Richard L. Linster, Assistant Director, National Institute of Justice

Martin Lively, National Institute of Justice

Tyler Lockett, Supreme Court Justice, Topeka

Richard Lovely, Professor of Sociology, John Jay College of Criminal Justice, New York

Malcolm MacDonald, President, American Probation & Parole Association, Austin

Michael McCann, District Attorney, Milwaukee

L. Aubrey Moore, Director, Operations Research, Peoria Police Department

Timothy J. Murray, Director, Pretrial Services, Miami

George Napper, Commissioner, Department of Public Safety, Atlanta

Jerry Needle, Manager, Center for Police Administration, International Association of Chiefs of Police, Gaithersburg, Maryland

Stephen D. Neely, Pima County Attorney, Tucson

Jack Novik, Executive Director, New York City Criminal Justice Agency

Antony Pate, Senior Research Associate, Police Foundation, Washington, D.C.

Linda Painter, Director, Pretrial Services, El Paso County, Colorado

Larry P. Polansky, Executive Officer, District of Columbia Courts

Albert J. Reiss, Jr., William Graham Sumner Professor of Sociology, Yale University

Robert A. Riols, Deputy Director, Georgia Division of Youth Services

Michael J. Robak, Court Administrator, Municipal Court of Marion County, Indiana

Peter H. Rossi, Professor, Social & Demographic Research Institute, University of Massachusetts

Jeffrey A. Roth, Senior Staff Officer, National Research Council

Phillip Roth, Judge, Circuit Court of Oregon

Daniel L. Rubinfeld, Professor of Law and Economics, University of California, Berkeley

James Rowland, Director, California Youth Authority, Sacramento

Bill Saulsbury, National Institute of Justice

Annesley Schmidt, National Institute of Justice

Janell Schmidt, Project Manager, Milwaukee Spouse Abuse Project

Michael D. Schrunk, District Attorney, Multnomah County, Oregon

Alan M. Schuman, Director, Social Services, Superior Court, District of Columbia

John E. Shapard, Federal Judicial Center, Washington, D.C.

Lawrence Sherman, Professor of Criminal Justice and Criminology, University of Maryland

Jeffrey M. Silbert, Executive Director, Dade-Miami Criminal Justice Council, Miami

Thomas E. Slade, First Deputy Commissioner, Department of Probation, New York

Darrel W. Stephens, Executive Director, Police Executive Research Forum, Washington, D.C.

Cecil H. Steppe, Chief Probation Officer, San Diego County

James K. Stewart, Director, National Institute of Justice

David Tevelin, Executive Director, State Justice Institute, Washington, D.C.

Mary A. Toborg, President, Toborg Associates, Washington, D.C.

Linda P. Tyon, Executive Director, TASC of Oregon, Inc., Portland

Fred B. Ugast, Chief Judge, Superior Court of the District of Columbia

Barbara D. Underwood, Chief of Appeals & Counsel to the District Attorney, Brooklyn

James G. Vetter, Deputy Chief of Police, Colorado Springs Police Department

Richard Vick, Department of Sociology, Washington State University

Christy A. Visher, Research Associate, National Research Council

Reggie B. Walton, Associate Judge, Superior Court of the District of Columbia

Gerald Wetherington, Chief Judge, Dade County, Florida

Charles F. Wellford, Professor of Criminal Justice and Criminology, University of Maryland

Larry Whalen, Chief, Cincinnati Police Division

Linda Youngquist, Supervisor, Pre-Trial Release Program, Des Moines, Iowa

Robert J. Ziarnik, Chief of Police, Milwaukee Police Department

APPENDIX C
BIOGRAPHICAL SKETCHES OF WORKING GROUP MEMBERS

RICHARD LEMPERT, who chaired the working group, is professor of law and sociology at the University of Michigan. He is vice chair and chair-elect of the Committee on Research on Law Enforcement and the Administration of Justice. He has also served on the Law and Social Science Panel of the National Science Foundation, as editor of the Law & Society Review, and as trustee and member of the executive committee of the Law and Society Association. His research interests span a range of law and social science topics, with special attention to the jury system, capital punishment, deterrence and, dispute settlement. He is the author (with Stephen Saltzburg) of A Modern Approach to Evidence and (with Joseph Sanders) An Invitation to Law and Social Science. He received a JD degree and a PhD in sociology from the University of Michigan.

RICHARD A. BERK is professor of sociology and statistics at the University of California, Santa Barbara. His research is concerned with evaluation research, the sociology of law, and applied statistics. He is currently the chair of the Methodology Section of the American Sociological Association and vice chairman of the Board of Directors of the Social Science Research Council. He is the author of many books and articles. He received a BA degree from Yale University and a PhD degree from Johns Hopkins University.

ROBERT BORUCH is professor of psychology and statistics at Northwestern University. His research concerns the conduct of applied research, especially field tests of social and administrative programs. He is current chair of the Survey Methods Section of the American Statistical Association and a member of the Project Review team for the National Institute of Justice Spouse Assault Replication Project. Boruch is author of Assuring Confidentiality of Social Research Data and Social Experimentation.

ANTHONY V. BOUZA has been chief of police of the Minneapolis Police Department since February 1980. Before that he had a long career with the New York City Police Department, culminating with command of the Bronx forces from 1973 to 1976. From 1977 to 1979, he was second in command of New York's subway police force. He has taught at John Jay College and Hamline University, guest lectured and written extensively. He holds BBA and MPA degrees from the Baruch School of the City College of New York.

SHARI SEIDMAN DIAMOND is associate professor of psychology and criminal justice at the University of Illinois, Chicago, and legal associate at the law firm of Sidley & Austin. Her research involves judicial and jury decision making, regulation of advertising, and methodological problems in the study of law. She is a fellow of the American Psychological Association, president-elect of the American Psychology-Law Society, and a member of the editorial boards of several journals. She received a BA degree from the University of Michigan, a PhD degree in social psychology from Northwestern University, and a JD degree from the University of Chicago.

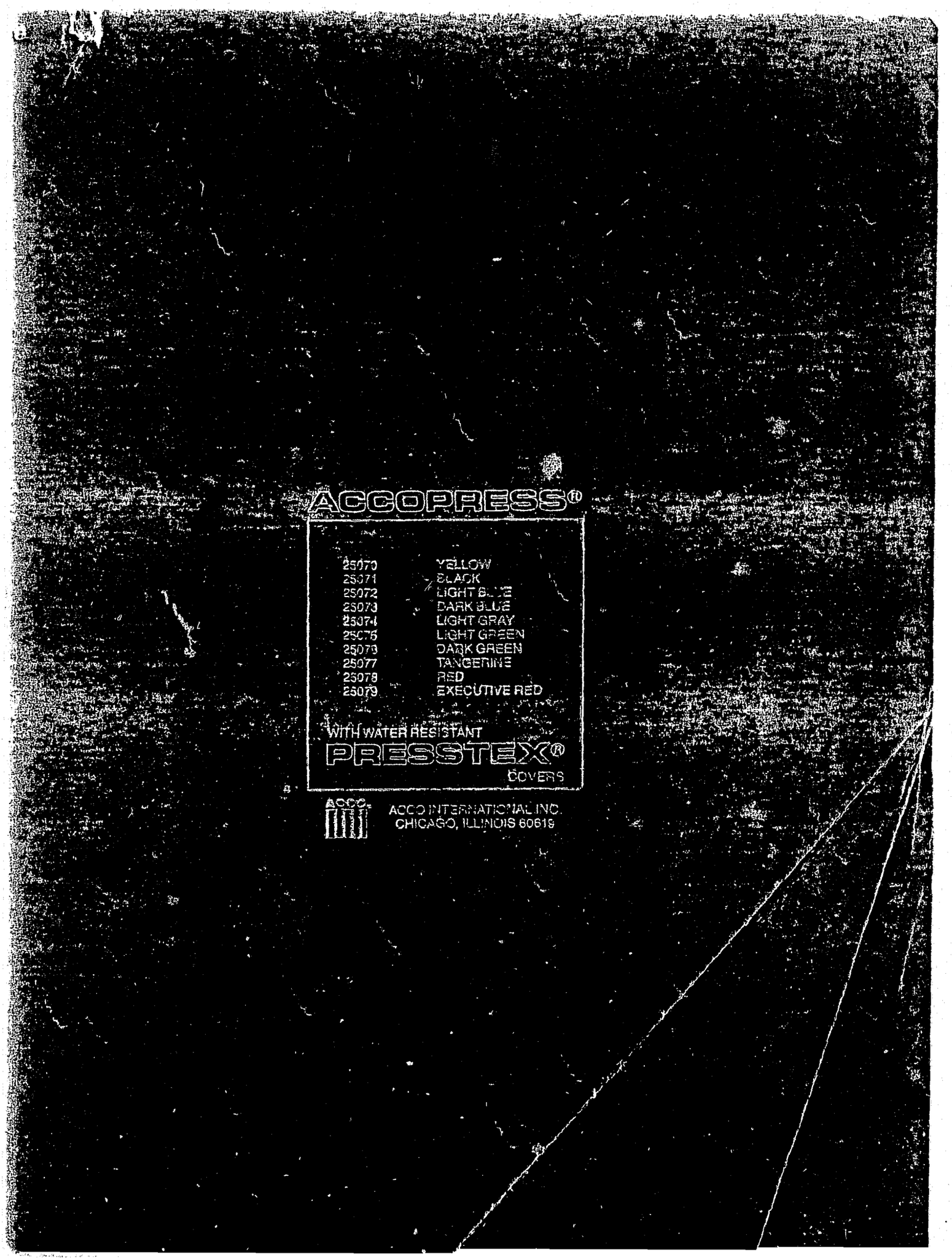
E. MICHAEL McCANN has served as elected district attorney of Milwaukee County since January 1969. He is on the Board of Directors of the National District Attorneys Association and is past president of the Wisconsin District Attorneys Association. He is a member of the Criminal Justice Council of the American Bar Association and has served as chairman of the Victims Committee of that association. He is an occasional lecturer for the National College of District Attorneys and the University of Wisconsin and Marquette University Law Schools. He received a BA degree from the University of Detroit, a JD degree from Georgetown University, and an LLM degree from Harvard University.

ALBERT J. REISS, JR., is the William Graham Sumner professor of sociology at the Institution of Social and Policy Studies and a lecturer in law at Yale University. Among other appointments, he has served as a consultant to the President's Commission on Law Enforcement and the Administration of Justice (1966-1967) and the National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research (1976). He is a past president of the American Society of Criminology. He is a fellow of the American Academy of Arts and Sciences, the American Statistical Association, and the American Society of Criminology. His research concerns police behavior, communities and crime, and methodological issues in the study of criminology and criminal justice. He has authored numerous books and articles. He received a PhD degree in sociology from the University of Chicago, an LLD (honoris causa) degree from the City University of New York, and a Docteur Honoris Causa from the Universite de Montreal.

DANIEL L. RUBINFELD is professor of law and economics at the University of California, Berkeley. His research involves the use of quantitative methods in law, the economics of litigation, and the economics of the state and local public sectors. He is the coauthor of Econometric Models and Economic Forecasts, co-editor of the International Review of Law and Economics, and a member of the American Economics Association, the Econometric Society, the National Tax Association, and the Law and Society Association. He received a BA degree in mathematics from Princeton University and MS and PhD degrees in economics from Massachusetts Institute of Technology.

JEFFREY A. ROTH, who served as the working group's study director, is the senior staff officer of the Committee on Research on Law Enforcement and the Administration of Justice. His interest is in the policy use of social research, especially in the areas of criminal careers, taxpayer compliance, and pretrial release. He is a member of the American Society of Criminology, the Law & Society Association, the American Economic Association, and the American Statistical Association. He received BA, MA, and PhD degrees in economics from Michigan State University.

CHRISTY A. VISHER, who assisted the working group as research associate, is also research associate with the Committee on Research on Law Enforcement and the Administration of Justice. Her research is concerned with criminal careers, juror decision making, police arrest decisions, and public policy issues in criminal justice. She is a member of the American Sociological Association, the American Society of Criminology, and the Law & Society Association. She received a BA degree from Trinity University and MA and PhD degrees in sociology from Indiana University.


ACCOPRESS®
25070 YELLOW
25071 BLACK
25072 LIGHT BLUE
25073 DARK BLUE
25074 LIGHT GRAY
25075 LIGHT GREEN
25076 DARK GREEN
25077 TANGERINE
25078 RED
25079 EXECUTIVE RED
WITH WATER RESISTANT
PRESSTEX®
COVERS
ACCO INTERNATIONAL INC.
CHICAGO, ILLINOIS 60619