

DISCUSSION

Michael L. Dennis, Northwestern University
Psychology Department, Evanston, Illinois 60208

Both Dr. Berk and Dr. Dunford have raised several key issues that I would like to reiterate using other kinds of evidence. Their similar themes stress that field research is often a muddy process and that successful implementation can be no more readily assumed than the internal validity of comparing non-equivalent groups. Yes, randomized experiments can provide the highest quality data in social research; but only when they are successfully implemented.

The first point is the importance of gaining cooperation from the service providers. They must be sold on the experiment, consulted on the implementation rules and listened to for potential problems. The ideas seem simple, even simplistic. But the history of field research is strewn with efforts that failed because service providers were not considered seriously enough.

In the Denver County Court experiment, for instance, the principal investigators were asked by the chief justice to conduct an experiment comparing the effectiveness of several sanctions for drunk driving (Ross and Blumenthal, 1974). They designed an experiment that required the bench judges to assign all cases within a given month to one of three treatments: fines; conventional probation, or counseling. But at no point did the investigators attempt to assure that the judges were persuaded of the need for the experiment or the merits of doing research. Even when it became obvious that the bench judges were overriding the random assignments with an exceptional case clause, the investigators continued to rely only on directives from the chief justice. The results included an assignment override rate of over 50% in the counseling treatment.

Based on police research, Dr. Berk suggested several methods for avoiding both intentional and unintentional subversion:

- *selecting motivated officers;
- *making a concerted effort to sell them on the research;
- *conduct additional training to establish the correct procedures
- *hold regular meetings with the officers
- *ride along in the squad cars

Together these techniques help to create a sense of camaraderie and commitment to the experiment. Sherman and Berk also presented their research as "Your Experiment" to the patrol officers. They cited research showing that more police officers die responding to domestic violence calls (22%) than any other single call (Straus, Gelles, and Steinmetz, 1980). And while officers are often seen as the world's leading cynics, they appealed to their desires to do a better job. They then regularly held beer parties in which the officers were encouraged to talk about any problems or ask further questions.

In the Omaha replication, Dr. Dunford went one step further by issuing a newsletter that kept officers informed about how the research and their cases were proceeding. This proved to be a major boon when it became evident that the assistant city's attorney was actually prosecuting and winning the arrested cases. Several officers, reported feeling that for the first time, the system was working, that they felt good about their jobs (personal communication with Sgt. Truax, O.P.D., July 19, 1986).

A second major point, was the importance of monitoring actual assignment. Simply reporting the subversion of random assignment is not enough. Data on the treatments actually delivered and on the subversion should be sufficient to allow additional analyses (e.g., Berk and Sherman, 1987). It is better to know about and control for these systematic biases than to assume that there are none. To realize that there will be legitimate legal and ethical reasons for overriding random assignment (e.g., when the assailant violently attacks a police officer or victim shortly after random assignment to an advise condition). The key is to prevent abuse of this loophole (e.g., the Denver County Court Experiment).

Dr. Dunford also mentioned the importance of randomization occurring after eligibility is determined and concurrently monitoring the caseload. In a variety of experiments in judicial, criminal, and civil research, it has become increasingly apparent that given the opportunity to peek at the next assignment, some people will (Partridge and Lind, 1983; Ross and Blumenthal, 1974; Sherman and Gartin, 1986). Furthermore, they allow this information to affect their decision to include the case.

It is not simply a question of external validity, in which they are systematically excluding cases.

Instead, they are often delaying inclusion in order to manipulate assignment. The very nature of this deception makes it unlikely that there will be any documentation in order to reanalyze the data. In cases such as the Detroit Shoplifting experiment, security guards who wanted to arrest a particular subject simply delayed logging them in until the arrest disposition came up (Sherman and Gartin, 1986).

The concurrent monitoring required by this advice has substantial positive side effects. By monitoring the caseload and dispositions, the principal investigator can catch problems and make decisions early enough to improve the research or terminate it. In the Kansas City Preventive Patrol Experiment (Kelling, Pate, Diekmann, and Brown, 1974) for instance, the investigators suspended the experiment after the first month when police officers in the reduced patrol beats began subverting the experiment. These officers had originally been given nothing to do in place of routine preventive patrols. Many incorrectly thought that the experiment

prohibited them from stopping a crime they inadvertently witnessed. The experiment was redesigned to have these officers conduct their patrols in beats with experimentally increased preventive patrols. They were also reassured that they were to respond to any crime that they witnessed.

One major lesson then, is to beware of designs that only propose to monitor randomization in a post hoc fashion. Toborg and Associates tried using some low cost methods of monitoring randomization. They indirectly randomly assigned subjects by using either the day of their arrest or their birthday. It was assumed that in spite of prior knowledge about the next assignment, it would be difficult for the service providers to manipulate these two variables. If they did, it would be relatively easy to measure how well randomization was implemented. Well, it was easy to measure, but without concurrent monitoring, randomization didn't work well. Out of five sites in Toborg's Pretrial diversion experiment, randomization failed completely in two and partially in a third (Toborg, 1981). The lesson here is that even the best designs on paper still require constant attention and skilfull management in the field.

The third point concerns the legal and ethical issues involved in doing a randomized field experiment. Dr. Dunford briefly mentioned the fear both police officers and ctij officials had over increased liabilities. These fears are not unique to either Omaha or even to Police Research. During the 1930's, Simon and Devine (1940, 1985) reported similar apprehensions from social workers in an experiment to determine the ideal caseload. The high caseload social workers were worried, both legally and ethically, about denying some people their full attention and services. The low caseload workers were worried about the extended period of low productivity being held against them in latter job moves. Like Omaha, the solution was simply getting the chief administrator to sign off on the experiment in writing.

Other questions, such as additional liability insurance, can be negotiated. Dr. Dunford showed the Omaha city officials that such insurance would be too expensive, if attainable at all. Fortunately, their position wasn't set in concrete, and Dr. Dunford's serious and good faith effort were apparently enough to allay their concerns. These are perhaps simple solutions, yet each of these issues must be addressed if the experiments are to get the high levels of cooperation they require.

A more complicated issue arises during the replication of a randomized field experiment. If an experiment finds a particularly strong effect, sould it be replicated? As social scientists, most of us would say yes. But how much evidence is enough? In the two years following the Minneapolis Domestic Violence study, the number of police departments with pro-arrest policies has risen from 10% to 44% (Cohn and Sherman, 1986). In fact, when Dr. Sherman proposed replicating the study in Milwaukee last year, several local groups

opposed him. They were citing Sherman and Berk (1984) as proof that arrest worked, implying that a replication involving anything less than arrest would be unethical (personal communication with Dr. Sherman, September 2, 1986).

Like Sherman, field researchers will not face these questions in academic halls, but in city councils and public forums. For this reason we must be aware of the utility and rationale of replications; always with ready illustrations at hand.

A case in point would be two experiments done on the Civil Appeals Management Program (CAMP) in the Second Circuit Court of Appeals. The program was designed to speed up the litigation process and to encourage pretrial settlements. In the original experiment, the court's counsel selected cases that seemed susceptible to settlement and then randomly assigned them to either CAMP or the traditional procedures. No significant effects were found (Goldman, 1979). In the second experiment, all cases in certain, objectively defined, categories were randomly assigned, again to either CAMP or the traditional procedures. This time CAMP was significantly better and more cost effective (Partridge and Lind, 1983).

Despite first appearances, there isn't necessarily a conflict between thse two findings. A program that is effective for a general population need not be effective for one of its subpopulations. Recall that first experiment only used cases that already appeared susceptible to settlement. Without replications then, we are in danger of generalizing findings that may be heavily dependent on how the original study was implemented. We would be like one of the proverbial blind men, describing an elephant, based only on touching it's trunk.

In summary then, we must make concerted efforts to gain the service provider's coeaprtion, to involve them in the planning process, and to maintain on-going lines of communication. Experiments should be designed to anticipate the type and frequency of assignment overrides. Actual assignment and delivered treatments should be monitored concurrently with the experiment and recorded in detail to assist in latter analysis. And finally, issue of legality and ethics mut be addressed in good faith.

Bibliography

- Berk, R. A. and Sherman, L. W. (1987) "Police responses to family violence incidents: An analysis of an experimental design with incomplete randomization. Journal of the American Statistical Association, forthcoming 1987.
- Cohn, E. G. and Sherman, L. W. (1987). Police policy on domestic violence, 1986: A national survey, (Report Number 5). Crime Control Institute, 2125 Bancroft Place, N.W., Washington, D.C.
- Goldman, J. (1979). Resolution of appellate litigation: A controlled experiment. Evaluation Quarterly, 3(4), 557-582.
- Kelling, G., Pate, T., Dieckman, D., and Brown, C. E. (1974). The Kansas City preventive patrol experiment: A summary report.

- Washington, D.C.: Police Foundation.
- Partridge, A. and Lind, A. (1983). A reevaluation of the Civil Appeals Management Plan. Washington, D.C.: Federal Judicial Center
- Ross, H. L. and Blumenthal, M. (1974). Sanctions for the drinking driver: An experimental study. Journal of Legal Studies, 3, 53-61.
- Sherman, L. W. and Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. American Sociological Review, 49, 261-272.
- Sherman, L. W. and Gartin, P. (1986). Differential recidivism: A field experiment on the specific sanction effects of arrest for shoplifting. Washington, D.C.: Crime Control Institute.
- Sherman, L. W. and Hamilton, E. E. (1984). Impact of the Minneapolis domestic violence experiment: Wave I findings. Washington, D.C.: Police Foundation.
- Simon, H. A. and Devine, W. R. (1985). Controlling human factors in an administrative experiment, (reprint of 1940 article). In Eleanor Chelmsky (Ed.). Program evaluation: Patterns and directions. Washington, D.C.: American Society for Public Administration.
- Straus, M. A., Gelles, R. J., and Steinmetz, S. K. (1980). Behind closed doors: Violence in the American family. Garden City, N.Y.: Anchor Books.
- Toborg, M. (1981). Pretrial release: A national evaluation of practices and outcomes. Phase II report. Washington, D.C.: National Institute of Justice.

Author's Notes

This work was partially supported by grants from the National Institute of Mental Health (No. 5 T32 MH15113-09) and the Rockefeller Foundation.