

Franklyn W. Dunford,* University of Colorado
 Institute of Behavioral Science, Campus Box 483, University of Colorado, Boulder, CO 80309

The benefits of randomized field experiments are well stated and acknowledged (Campbell and Stanley, 1963; Kerlinger, 1964). Less apparent are practical guidelines for the implementation of randomization under nonlaboratory conditions (Dobson and Cook 1979). The purpose of this paper is to comment on a handful of practical issues common to randomized field experiments.

ETHICAL AND LEGAL ISSUES

When researchers approach practitioners to discuss methodology, the first objections to random assignment typically involve ethical and legal considerations. Practitioners are concerned about the welfare of their clients and the extent to which they as practitioners will be legally culpable for the seemingly arbitrary assignment of clients to what are perceived as experimental (unproven) treatments. In a day of excessive litigiousness this concern is not without merit, nor are the responses to such concerns to be made lightly. A major challenge for researchers interested in conducting randomized field experiments is thus to think about such issues from the practitioner's point of view and be prepared to provide information about, and justification for the use of this technique. A few examples of attempts to follow this counsel are illustrative.

While working with the Juvenile Court and the Juvenile Division of a major midwestern police department to develop research to test the effects of community based youth service diversion programs on reducing delinquent behavior, the police and court expressed concern about the ethics of randomly assigning youth apprehended for misdemeanor violations to one of three dispositions: lecture and release, referral to a community based youth service agency, to juvenile court (Dunford, Osgood and Weichselbaum). At the time of the research, Negative Labeling was a popularly accepted theoretical explanation for the development and continuation of sustained delinquent behavior among youth (Becker, 1963). The diversion of youth apprehended for violations of the law away from the juvenile justice system to "nonstigmatizing" community based youth service agencies was postulated as a way to avoid the negative labeling attributed to the court and to subsequently limit the official

systemic promotion of crime among youth (Schur, 1973).

Cities of all sizes throughout the United States were making sizable investment in diversion programs for juvenile offenders in the absence of any real evidence that such programming was effective in controlling delinquency (Bullington, et al., 1978). This vacuum created a moral dilemma for law enforcement and justice administrators. Were the courts promoting the very behavior they were mandated to control? Were community based youth service agencies less stigmatizing than the court and as a consequence more suitable for reducing delinquent behavior among youth than were the courts? Was no legal action the most suitable official response to all but the most serious delinquent acts committed by youth? Law enforcement and justice administrators agreed, that in the face of the large investments being made in community based youth service agencies, the tremendous pressure they were under to refer officially apprehended youth to diversion programs and the weight of the theoretical arguments for diversion, they had a moral obligation to determine, using the most rigorous methodology available to them, which among a variety of police and justice system responses to delinquent behavior was the most effective in controlling crime. The need to know if the court was less effective in reducing delinquency than lecture and release or referral to community based youth service agencies eventually took precedence over the concern about the ethics of random assignment.

A second example of anticipating and dealing with the ethical concerns of practitioners involves an effort to evaluate a federally supported summer youth employment program in a large Northern California city (Elliott and Knowles, 1976). The goal was to test the effects of summer employment by comparing groups of youth randomly assigned to receiving (experimental group) and not receiving (control group) Summer job placements. Practitioners were initially concerned about the arbitrary nature of granting jobs to some and not to others until the fairness of the procedure, in the face of excessive demand for jobs and limited job opportunities, were clarified. The fairness of the design was recognized and accepted by practitioners who subsequently enjoyed the unanticipated consequences of the procedure by being able to point out to a large and politically active minority population pushing for summer jobs for minority youth the unbiased nature of the selection process.

A concern over the legality of random assignment is exemplified in another experiment involving the police. In seeking cooperation to conduct research in a second major midwestern city in which police

* Much of the data presented in this paper came from grants supported by the National Institute of Justice, Department of Justice. Points of view or opinions expressed in this document are those of the author and do not necessarily represent those of the National Institute of Justice. Appreciation is extended to Delbert S. Elliott for his review of an earlier draft of this paper.

response to misdemeanor domestic disturbance calls would be randomly assigned to three different police dispositions (advise, separate or arrest), police and city officials expressed concern about two issues: charges/suits for false arrest and charges/suits for failure to arrest in the presence of probable cause. These two questions had been anticipated and were addressed at the outset of negotiations with the city. Historically arrests for misdemeanor domestic assaults were discouraged in the city of interest by virtue of department tradition and training (but not by official written department policy) unless the assault occurred in the presence of a police officer. Such cases were handled as civil matters, a common police procedure during the 70's and early 80's. The consequence of this practice was that the police seldom made arrests for domestic misdemeanor assaults even though legal authority to do so had always existed. For the purpose of the proposed research, eligibility for inclusion in the pool of cases to be randomly assigned to one of the three dispositions was dependent upon the presence of probable cause for arrest for a misdemeanor assault. The research design thus mandated that all cases randomly assigned to the arrest outcome would meet the requirement of probable cause. The police, it was noted, had always had the legal authority to make arrests under such circumstances and were simply choosing anew to exercise that authority. The concern for claims of false arrest were thus easily resolved on the basis of longstanding legal statute since probable cause for arrest was a requirement for all cases entered for random assignment.

The resolution of the concern over a failure to arrest in the presence of probable cause (which could be made for cases randomly assigned to advise or separate) followed from the research design as well. It was reasoned that with the implementation of the research design the city and police department would reduce the risk of such charges by one-third. That is, one-third of the cases in which arrests were made under the proposed random assignment design would not have been made absent the research. Thus the city and police department would literally reduce liability for charges of failure to arrest with the initiation of the research because some arrests would be made in a category of cases previously handled informally.

Researchers interested in the implementation of random assignment procedures in field experiments are wise to explore the ethical and legal ramifications of such procedures for the practitioners and populations involved, and to insure and be able to communicate that the design is reasonable and defensible on both legal and ethical grounds before approaching those whose permission or support is required to conduct the research.

RISK

A related legal issue is the liability risk potentially associated with some classes of randomized experiments. Experience suggests that practitioners are potentially at legal/fiscal risk in some field efforts that use random assignment techniques. Law enforcement and judicial agencies are particularly sensitive to this issue. The aforementioned field experiment with a midwestern city and police department is one example. As a condition to agreeing to conduct the research in the police department, the city initially required the researchers to indemnify the city against legal action or to provide an insurance policy to cover liabilities incurred as a result of the experiment. The language of that condition was as follows:

The (Research Group) shall provide to the (City) insurance coverage to hold the (City) free from any liability which may be resulting directly and uniquely from the experiment. Or if such liability insurance is unavailable, they hereby agree to indemnify the (City) for any losses incurred as a direct result of liability incurred uniquely and directly by the experiment.

In response to the demands for protection of this kind, the researchers investigated the possibility of purchasing liability insurance only to discover that insurance companies would not underwrite a policy for a subset of officers within a department (the proposed research was limited to a few officers in the department), nor for a limited set of liabilities (i.e., property damage, bodily harm, false arrest) the most likely areas of litigation for the proposed research). The researchers communicated to City officials their findings, indicating that the purchase of an insurance policy for an entire department at full coverage was an unworkable option and that the quoted cost of \$600 per officer was well beyond the research organization's ability or the sponsoring agency's willingness to fund. Further, because of the lack of resources the researchers expressed their inability to indemnify the city against legal action. The researchers consulted an attorney regarding the liability issue and were able to provide the city with a legal opinion stating that given the research design the conduct of the study would not alter the potential for suits or claims against the city beyond that which existed prior to the initiation of the research. The researchers provided the city with a copy of the attorney's opinion and reiterated the characteristics of the research design pertaining to the attorney's brief.

After an extended discussion the city dropped demands for insurance or indemnification and agreed to the conduct of the research. A postmortem review of the interaction with the city over this issue

suggested that (1) the city wanted to participate in the research from the outset, but was genuinely concerned about the potential for legal action against the city, that (2) it cost the city nothing to push the researchers as far as possible in an effort to get them to assume potential liabilities created by the proposed research, and (3) that when the researchers responded with a good faith effort to meet the demands of the city, and were able to demonstrate an inability to comply with the city's request and could present data minimizing the city's fears regarding litigation, the city backed away from their original demands.

Two practical lessons emerge from this discussion. First, practitioners face real and legitimate judicial risks with real fiscal consequences when participating in some kinds of research involving random assignment. Insofar as litigation has been a historical issue for practitioners and they, or the researchers, are seen as not lacking resources, risk issues will be raised and researchers must be prepared to deal with them thoroughly and convincingly.

Second, the forces that motivate practitioners and officials to participate in experimental research, are not always obvious. Perseverance and good faith responses to the demands and concerns of practitioners and administrators often result in success where little was foreseen and enhances the relationship between researchers and practitioners. The first seeds of trust and cooperation are often planted at the entry phase of research implementation.

ELIGIBILITY DECISIONS

Field experiments often involve selecting and randomly assigning cases as they naturally occur or come to the attention of the agency or organization involved in the research. Typically, judgements must be made to identify those cases out of the variety coming into the agency or organization that are relevant to the research. If, for example, we were seeking cases for a study of alternative police responses to drunk drivers, we would screen out all traffic stops save those involving drunk drivers. Decisions regarding eligibility for inclusion in the research would be, in this manner, critical for the success of the research experiment. When eligibility decisions are made is a significant issue. At the heart of the matter is the researchers ability to exercise some control over the potential manipulation of the random assignment procedure. When knowledge of random assignment outcome precedes eligibility determination, in time order, the random assignment process can be manipulated quite simply by defining cases as eligible or ineligible depending upon what the next known random assignment option is. If, for example, the individuals who define cases as eligible know that the next disposition is arrest and a nonarrest disposition is preferred, they can effect the preferred disposition by defining the presenting case

as ineligible for inclusion into the study and releasing the potential subject. The seriousness of this time order issue is potentially mitigated by the interests of those defining eligibility. In those instances in which the definers have a vested interest in outcomes the problem is most serious. Where those determining eligibility have no such vested interest the potential for manipulation is lessened. The realities of experimental field research are such, however, that whenever practitioners are defining eligibility, disinterest cannot be assumed. The sure way to proceed is to require that random assignment always follows eligibility decisions.

An example of this particular problem is illustrated in the Minneapolis Experiment as reported by Sherman and Berk. The research design called for the randomization of suspects involved in misdemeanor domestic disturbances to one of three treatments. Police officers carried a pad of report forms color coded for three different treatments. Each time the officers encountered a situation that they determined to be eligible for the experiment, they were supposed to take whatever action was indicated by the report form on the top of the pad. The forms were numbered and arranged for each officer in an order randomly determined.

By allowing the officers who determined eligibility access to the randomized outcomes prior to their eligibility determinations, the researchers potentially tempted the police officers involved in the experiment to manipulate the random assignment procedure. Sherman and Berk recognized the problem and made the following observations about the possibility of police officer manipulation of the random assignment process.

" . . . we are left with at least two disturbing possibilities. First, police officers anticipating . . . a particular kind of incident, and finding the upcoming experimental treatment inappropriate, may have chosen to exclude certain cases in violation of the experimental design. This amounts to differential attrition, which is clearly a threat to internal validity. . . .

Second, since the recording officer's pad was supposed to govern the actions of each pair of officers, some officers may also have switched the assignment of driver and recording officer after deciding a case fit the study in order to obtain a treatment they wanted to apply. If the treatments were switched between driver and recorder, then the internal validity was again threatened." (Sherman and Berk, 1984)

It is clear that opportunities to manipulate the randomization process are reduced when eligibility decisions are made before and independent of knowledge of random assignment outcome.

UNKNOWN VIOLATIONS OF RANDOM ASSIGNMENT

Manipulation of random assignment procedures of any kind is a serious threat to the internal validity of a research design. Instances of unknown manipulation represent,

however, a more serious threat to the integrity of a randomized experimental design than are known instances of manipulation primarily because the latter can be tracked and assessed for impact more readily than the former. Where violations of the randomized design are known, data analyses can be completed on samples as assigned, on samples as delivered or on samples in which the misapplied cases have been removed. In this manner the effects of manipulation can be observed from several angles and can be made more explicit.

This problem is quite common in field experimentation and is well illustrated by research conducted by the Police Foundation on the effects of arresting shoplifters. The research was conducted in cooperation with a chain of Midwestern area department stores. The research design called for security personnel to follow predetermined and alternating arrest or release dispositions as kept in logs in the security offices of each of ten stores. When security officers apprehended shoplifters, they were not supposed to know whether the shoplifter would be arrested or released until the log was consulted for a disposition. Technically, eligibility for inclusion in the study was to occur before random assignment to arrest or release were made. In practice the design was easily compromised. Manipulation of the random assignment procedure was confirmed in a chance conversation with a former store security director who had left the department store chain during the experiment. "If they (shoplifters) gave us a hard time or we didn't like them, and we wanted to prosecute but we knew it wouldn't be approved as an exemption, we checked what disposition was up next: if it was release we just waited until another case came in and then logged our bad guy in as a prosecution case" (Sherman and Gartin, 1986). The researchers had no way of determining the frequency with which this type of manipulation occurred and were unable to specify the effects of the manipulation upon their findings.

A similar circumstance was experienced in research conducted with a large Southern California law enforcement agency. The research design called for law enforcement officers in the field to assign cases found eligible for the research to one of four randomly determined outcomes. Officers were to call one of nine police stations for a disposition upon encountering an eligible case. A log was maintained at each station in which the randomly predetermined dispositions were kept covered individually with nontransparent tape. It was discovered, early on in the study, that the officers in three of the nine stations were manipulating the random assignment process. They were peeling back the tape on a number of cases simultaneously and choosing among the revealed dispositions depending upon the seriousness of the case, officer interest, etc. (Klein, 1986).

Given the typical costs and care associated with most randomized field experiments, researchers must not settle for randomization procedures which will not insure the identification of misapplications or exceptions to every randomized assignment.

INTRUSIVENESS

The intrusiveness of the actual mechanics of the random assignment process has both access and operational consequences. If the procedures for obtaining a random assignment option are complicated and time consuming they may be rejected at the outset as inappropriate and unworkable by administrators.

A more serious concern is the operationalization of an administratively approved random assignment design that proves to be so intrusive to the standards, goals and procedures of practitioners that they fail to submit eligible cases for assignment after the research is in progress. At best, this "dropping out" may jeopardize the viability of the research by extending the sampling period and the associated costs of the research. At worst, the withholding of cases may introduce a systematic bias into the sampling procedures when excluded cases represent a systematic class of people/events whose absence skews the data set. A concern about the intrusiveness of the actual mechanics of random assignment is not an idle issue. To the extent that the procedures for random assignment impede and/or severely prolong the work-tasks of practitioners, their reluctance to participate by submitting cases for random assignment will increase.

The most unobtrusive and beguiling method for randomly assigning cases to alternative experimental groups is one in which practitioners know the random assignment outcome of the next case before the next case comes along. In this manner no time is wasted on the mechanics of securing a random assignment disposition once eligibility has been determined. The apparent practicality and unobtrusiveness of such a process should not be allowed to obscure the potential liabilities inherent in its use (as detailed above).

It should be assumed that the logistics of most field random assignment procedures will be intrusive to the work of practitioners. The mechanisms for random assignment typically intervene the normal flow of practitioner business whether it involves consulting a chart, making a telephone call or interacting with a machine. The process almost always interrupts the normal flow of events. The researcher's concern is to limit the intrusiveness of these interruptions, to minimize the extent to which they place practitioners in jeopardy (as in police work) and hinder practitioner effectiveness.

CASE FLOW

A Parkinson's like law states that the number of eligible cases required for the conduct of a randomized field experiment is reduced by one-half the day that the experiment is funded. As in all humor, the

element of truth contained in this proposition is what makes it funny. Given the costs of field experiments accurate estimates of the number of cases available for random assignment to experimental and control groups is of foremost importance. In the five randomized field experiments which the author has directed, case flow has always been less than expected. It is a common problem and one that can be anticipated and considered prior to the implementation of a field experiment. The reasons for less than expected flow rates are attributable to a number of factors.

The absence of good baseline data is probably the major explanation for less than expected flow rates in field experiments, and is problematic for two important reasons. First, without fairly accurate estimates of the number of eligible cases likely to be encountered in a given time span, investigators are ignorant of the amount of time that will be required to reach appropriate sample sizes. This has obvious implications for a variety of issues: the ability to estimate and complete the research within time-budget constraints, the ability of the organization to cope with the intrusiveness of the research design and the systems ability to deal with abnormalities generated by the research. The latter was a serious issue for justice agencies involved in diversion experiments which diverted cases normally headed for court to community based treatment agencies. The reduction of the number of cases penetrating the Juvenile Court in one city decreased the work load of the Court and Probation Department to the point that the County Supervisors considered reducing the Court's budget. About the time that the Court/Probation Department adjusted to the lower case flow resulting from the experiment it ended and the juvenile justice system was overwhelmed with the increase in case flow. If practitioners are advised that research will take 6 months to be completed and they are subsequently required to be involved an additional 6 months, resentment and disinvolvement may follow. If a research staff is supported by grant funds, any extension of time will require expenditures for salaries and overhead costs that were not anticipated and will result in adjustments of some sort - additional funds, truncated research objectives, staff reductions, crude analyses, rushed reporting. If the host organization is devoting resources to the experiment, an extension may represent an organizational inconvenience or other organizational priorities may be compromised. The accuracy of time estimates for data collection are greatly enhanced with good baseline data.

A second and perhaps more troublesome consequence of poor baseline data relates to representiveness and accountability. When researchers do not know how many cases of a preferred (eligible) type exist in a given population, they will have no way of assessing how well that population is

represented in the experimental sample nor will they have a good basis upon which to judge practitioner cooperativeness in referring cases to the experiment. Both are potentially serious problems. The former has to do with the generalizability of the findings associated with the acquired sample. If a case cannot be made that the experimental sample is representative of the population from which it was selected, then the relevance of the findings must be limited to the sample itself which may seriously jeopardize the utility of the findings. The concern is that outcomes based upon seriously biased samples may have little utility for more general or typical populations. The other issue, accountability, has very practical implications for the conduct of the research as well. A low flow of cases into an experiment may be due to a reluctance of practitioners to refer cases to the experiment or to a paucity of cases eligible for referral. The absence of good base line data prohibits researchers from making the distinction between rare events and biased samples.

When researchers have an established base rate for a particular event they want to observe, their ability to effectively observe it increases. Sustained or significant departures from the base rate alerts the researcher that referral procedures may not be operating satisfactorily. Where researchers are dependent upon practitioners for referrals, base rate data helps to assess the degree to which practitioners are adhering to referral procedures. As important, such data represent a means for holding those responsible for the referral of eligible cases to the experiment, accountable. When researchers are ignorant of the rate at which events of interest occur, their ability to identify and deal with uncooperative practitioners responsible for referring eligible cases is limited.

To state that good baseline data may be essential to the conduct of high quality field experiments is not to discount the challenge often associated with its acquisition.

This difficulty is exemplified in law enforcement research focusing upon specific types of crimes coming to the attention of the police. For example, if researchers are interested in determining the incidence of misdemeanor assault involving conjugal partners coming to the attention of the police over a given period of time, what are their options? 1. They can go to police reports of action taken involving domestic disturbance calls, assuming that the reports are accessible and retrievable. The major problem with this approach is the absence of reports for misdemeanor domestic assault. In many, if not most police departments in this country, reports for misdemeanor domestic assaults are frequently not made unless an arrest is made, effectively preventing the use of police reports to assess incidence rates. 2. Researchers might consider using

dispatch data to determine baseline incidence rates. However, even in departments using Computer Assisted Dispatch and Enhanced 911 dispatch systems the discrepancy between calls as dispatched and calls as found is sufficiently different to effectively destroy the utility of dispatch records as a source of accurately defining baseline data. Perhaps the only sure way to establish accurate baseline data for domestic disturbance calls of the type defined above is for independent observers to respond to calls with the police as disturbances occur to define and record the nature of each call. Collecting observational data of this sort is usually not cost effective, which accentuates a commonly existing research fact--good baseline data is hard to find for many types of field experiments.

UNEXPECTED EVENTS

When the conduct of field experimentation is contemplated, consideration should be given to the length of the sampling period. While it is important to insure sufficient time to capture a representative sample of cases during the sampling period, the shorter that period is the less likely changes in the experimental environment will complicate the research design. Shorter experimental periods provide less time for forces and circumstances beyond the control of the experiment to threaten the integrity of the research design and/or the relevance of findings.

An experience or two from one of the mid-western cities mentioned above will illustrate this point. About six months after random assignment began, the Chief of Police of the department whose men were referring cases to the project for random assignment to alternative treatments was fired by the mayor. An acting chief was appointed while the former chief set in motion a long and formal appeal process that was filled with obstacles and delays. Within the first month of his appointment the acting chief eliminated one of the two shifts participating in the study (8:00 p.m. to 4:00 a.m.). This action not only decreased case flow into the project, but also eliminated cases occurring between 12:00 midnight and 4:00 a.m. The Acting Chief then changed the policy regarding shift changes from an annual to a semi-annual procedure which was immediately implemented. The effect on the experiment was that experienced and trained officers could bid off of the experimental shift at six months rather than 12, creating higher turnover rates, the need to train new people and integrate them into the study. Further, had the Acting Chief failed to lend active and well publicized support for the experiment, support from the officers referring cases from the field would have dwindled dramatically. All of this could have been avoided if sampling had been completed in 6 months. The longer sampling takes the greater the likelihood that changes in the

environment in which it takes place will occur that have the potential to negatively impact the implementing of the research design.

SUMMARY

The issues raised in this paper are obviously not exhaustive of all of the types of problems that those attempting to implement randomized field experiments potentially face. The adage that anything that can go wrong will go wrong should be understood to apply to field experiments, such that those responsible for insuring the integrity of such experimental designs must pay attention to business. A first principle for all researchers, and most emphatically for those involved in randomized field experiments, is that nothing about an experiment can be left to chance or taken for granted. Attending to business every day is the name of the game.

REFERENCES

- Becker, H. S. 1963. *Outsiders*. New York: Free Press.
- Bullington, B., J. Sprawls, D. Katkin and M. Phillips 1978. "A Critique of Diversionary Juvenile Justice." *Crime and Delinquency* 24(1): 59-71.
- Campbell, D. T. and J. Stanley 1973. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally
- Dobson, D. and T. J. Cook 1979. "Implementing Random Assignment: A Computer-Based Approach in a Field Experimental Setting." *Evaluation Quarterly* 3: 472-489.
- Dunford, F. W. D. W. Osgood, and H. F. Weichselbaum 1982. *National Evaluation of Diversion Programs: Executive Summary*. Office of Juvenile Justice and Delinquency Prevention (August).
- Elliott, D. S. and B. Knowles 1976. *An Evaluation of the Oakland Youth Work Experience Program*. Final report to the National Office for Social Responsibility.
- Kerlinger, F. 1964. *Foundations of Behavioral Research*. New York: Holt, Rinehart and Winston.
- Klein, M. W. 1986. "Labeling Theory and Delinquency Policy: An Experimental Test." *Criminal Justice and Behavior* 13(March): 47-79.
- Riecken, H. W. and R. F. Boruch 1974. *Social Experimentation: A Method for Planning and Evaluating Social Intervention*. Academic Press: New York, San Francisco, London.
- Schur, E. 1973. *Radical Non-Intervention: Rethinking the Delinquency Problem*. New Jersey: Prentice Hall.
- Sherman, L. W. and P. R. Gartin 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): 8.
- Sherman, L. W. and R. A. Berk 1984. "The Specific Deterrent Effects of Arrest for Domestic Assault." *American Sociological Review* 49(2): 264.