CHAPTER 23

Place Randomized Trials*

ROBERT BORUCH, DAVID WEISBURD, AND RICHARD BERK[1]

BACKGROUND AND RECENT HISTORY

An important and emerging vehicle for generating dependable evidence about what works or what does not work falls under the rubric of place randomized trials. In criminology, for instance, such a trial might involve identifying a sample of high crime hot spots and then *randomly* allocating the hot spots, the *places*, to different police or community interventions. The random assignment assures a fair comparison among the interventions, and when the analysis is correct, a legitimate statistical statement of one's confidence in the resulting estimates of their effectiveness. See, for example, Weisburd et al. (2008) for illustrations of such trials and other references in what follows.

This chapter provides basic definitions and practical counsel, with illustrations from criminology, education sciences, and health and welfare research. The problems of generating dependable evidence about what works in crime prevention to detection transcend conventional academic disciplines. Their potential solutions do so. This is a reason for depending on work from different areas. The chapter also identifies issues, ideas, and challenges that might be addressed in further research, such as recent efforts to enhance the quality of reporting and in designing place randomized trials. We also discuss the special analytic difficulties that may occur in the development of place randomized trials as opposed to more traditional trials, in which individuals are the units of random allocation and analysis.

*Research and development work on this topic has been supported by the Rockefeller Foundation and done pro bono under the auspices of the Campbell Collaboration. We are grateful to Dorothy de Moya for assistance and intellectual support.

[1] Robert Boruch is University Trustee Chair Professor at the University of Pennsylvania's Graduate School of Education and Statistics Department of the Wharton School, co-directs Penn's Center for Research and Evaluation in Social Policy, and is a member of Penn's Graduate Group in Criminology. David Weisburd is Chair Professor at the George Mason University and Hebrew University of Jerusalem, and directs the Center for Evidence Based Crime policy at GMU. Richard Berk is Professor of Statistics at the Wharton School and Professor of Criminology, School of Arts and Sciences, at the University of Pennsylvania.

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_23, © Springer Science + Business Media, LLC 2010

In the vernacular here, we use the phrase "place randomized trial." In related statistical and social science literature, phrases such as "cluster randomized trial," and "group randomized trial" are used and they mean the same thing *at times*. References are given below. "Place randomized trial" is used because the term is becoming common in the criminological literature and is used in some welfare and education reports on impact evaluations. More importantly, perhaps, the phrase is likely to be better understood by nonresearchers than are other technical and less transparent phrases.

PLACES THAT ARE RANDOMIZED: THEORY AND THE UNITS OF RANDOMIZATION

Place randomized trials depend on a clear understanding of the role of place. On the one hand, a place can be an entity unto itself. For example, business establishments have a legal status separate from their owners and employees and relevant outcomes such as profit and loss can be analyzed as such even though the thousands of individuals may work in the establishments or be their customers. In many cases, schools, police departments, and neighborhoods have the same properties. Measures of such places can have little to do with the smaller units within them, e.g., the market share of a retail establishment, arrest rates in bars, or reputation of a neighborhood.

On the other hand, a place can be primarily an organizational convenience for smaller units within it. Schools can be construed as a good example. One may care about schools because they educate the young and because they provide employment for teachers and administrators. Further, when one asks, for instance, whether a school is "good," attention is often directed to the average academic performance of students at the school level as well as to performance of students (and teachers) within the schools.

Whether the place in place-based randomized trials is an entity in itself or little more than a receptacle for lower level observational units, such as people within the units, can affect dramatically how a randomized trial is designed and analyzed. For instance, the policy maker's goal may be to change how a place functions or to change the way the people within a place function. Sometimes the two are linked. However, the key definitional point in this chapter's context is that in place randomized trials, the random assignment is by place. One important implication is that the methodological benefits of random assignment adhere primarily to places, not the units within them. It follows that statistical analyses at the level of places conform to well understood and widely accepted statistical practice. It also follows that statistical analyses at the level of the units within randomized places, when units are nonrandomized, can be difficult and subject to controversy.

In statistical parlance, the "units of randomization," in a place randomized trial are the places. These units may vary considerably. Weisburd and Green's (1994) study, for instance, focused on drug crime hot spots, street segments, as opposed to specific institutions such as housing developments, schools, or business units. The broad theory underlying the trial posited that focusing police and other resources intensively on hot spots will reduce crime. This is in counterpoint to a theory that says such a focus will not have an effect, and a possible further theory that focusing police resources on hot spots will lead to migration of criminal activity to neighboring areas. Earlier, hot spot trials had been undertaken by Sherman and Weisburd (1995) in Minneapolis, where the unit of analysis was the single street segment from intersection to intersection. The earliest related precedent appears to have been the Kansas

City police patrol experiment (Kelling et al. (1974) where whole police beats were randomly allocated. See Braga (2001) for a review of nine related studies covering five such trials on policing crime hot spots.

Other places have been targeted for different kinds of interventions, such as saloons in the context of preventing violence (Graham et al. 2004) and preventing glassware-related injuries (Warburton and Sheppard 2000). Private properties, including apartment houses and businesses, have been targeted in a study on the effects of civil remedies and drug control (Mazzarole et al. 2000). Housing projects were randomized in a study on preventing elder abuse (Davis and Taylor 1997). Convenience stores, crack houses, and other entities have also been randomly allocated to different interventions.

In health research and evaluation, place randomized trials are usually called "cluster randomized trials." They are becoming frequent as researchers recognize the need to change entire healthcare institutions, or subunits of them, in the interest of enhancing peoples' health. For instance, Grimshaw et al. (2005) randomly allocated medical family practices to different interventions. Each cluster was a medical practice, with physicians, staff, and patients nested within this practice. Leviton and Horbar (2005) randomized tertiary care facilities and neonatal intensive care units within hospitals to different interventions. They did so in order to learn whether better practices in perinatal and neonatal medicine could be deployed well in multiple facilities and would have a statistically dependable effect on patient outcomes. Donner and Klar (2000) give illustrations of such trials in health research reported between the 1950 and the late 1990s.

Sikkema (2005) randomly allocated entire public housing developments to different regimens to learn whether a particular opinion leader-based intervention, tested earlier in other trials and sustained by coherent theory, produced a detectable and substantial effect on women's health behavior. Although their particular focus was AIDs prevention, health-related work has been done in other arenas, in which brothels, bars, and other entities have been the units of random allocation and analysis. For instance, hospital/clinic catchment areas were the units of allocation in rolling out Mexico's place randomized trials on the effects of universal health insurance programs for the rural poor (Imai et al. 2009).

In the Jobs Plus trials, Bloom (2005) and Bloom and Riccio (2005), and their colleagues also randomized housing developments to different interventions just as Sekkema et al. did, but they did so for an entirely different purpose. It was to understand the effects of a particular form of development-wide approach to increasing social capital, including increases in wage and employment rates, and other outcomes. Sekkema et al. depended on theoretical and empirical research on the effects of diffusion strategies, including opinion leader approaches in the health sector. Bloom and Riccio depended on what could be construed as a theory of developing human capital. A theory of diffusion of innovation and change is implicit, rather than explicit, in their work.

Brian Flay is among the first social scientists to have succeeded in randomizing entire schools to different health risk reduction interventions (Flay and Collins 2005). In the US and Canada, Flay, and Botvin et al. (2000), and others have advanced theories of change so as to develop better interventions, and tested theories which posit how changing individual behavior depends on changing school conditions, including group processes within the school. The outcome variables in prevention trials usually include substance abuse and, at times, other behaviors. This, of course, impinges on work by criminologists dealing with juveniles and adults.

Entire villages containing low income families were the targets in Progresa/Oportunidades in Mexico, reported by Parker and Teruel (2005), Gulematova-Swan (2009), Schultz (2004), and others. In this randomized trial, the theory was that changing village behavior, notably through conditional financial incentives (conditional cash payments to mothers), would increase the rate at which children stayed in schools rather than working in the fields. Further, the theory posited that such an intervention would build social capital at the regional and country level. A main outcome variable was children dropping out of school, which has implications not only for human capital development but also for juvenile and adult crime. A partial replication of this conditional income transfer program is being undertaken in New York City.

The United States Department of Education has, since 2001, sponsored testing of curriculum packages and other interventions in multiple places as well as trials involving random allocation of individuals to different interventions. The place randomized trials involve entire school districts, or schools, or classrooms as the entities that are randomly assigned to different interventions. Nearly 100 randomized education trials were mounted from 2002 to 2009. For instance, in one place randomized trial, Merlino (2009) engaged nearly 200 schools to understand the effects of different approaches to enhancing children's understanding of science. The different approaches exploited theory from the cognitive sciences to augment commercial curriculum packages being used in the schools. See Garet et al. (2008) and Porter et al. (2005) for examples of place randomized trials involving large numbers of schools in the context of testing the effects of school based professional teacher development programs.

Each of these examples educates us and invites basic questions. How and why can such trials be deployed well? Can we develop better theory about what should happen as a consequence of an intervention at high levels of the units: province or county, city or village, institution or housing development, hospital catchment area or crime hot spot? Can theories be developed to guide thinking about change or rate of change at the primary aggregate level – the places-and below it. What new statistical problems emerge from randomization of places? Each also invites a question about how to learn about other trials of this sort, involving yet other units of allocation and analysis.

RELATIONSHIPS AND AGREEMENTS

How have people gotten place randomized trials off the ground? Part of the answer to the question lies in agreements between the trialist's team and the prospective partners in the place based trial. "Partners" here mean individuals or groups whose cooperation and experience are essential in deploying both the intervention and the trial.

In dealing with tests of prevention programs in schools, for instance, Flay et al. (2005) emphasized written agreements signed by people who are authorized to sign such agreements. In Flay's case, this includes not only a school principal but also the school superintendent. Given the mobility (turnover) of school administrators and teachers in parts of the US, one or the other signatory might disappear before the end of the trial. Having both kinds of people as signatories helps reduce the obstacles to running a fair trial to completion. Elsewhere, Slavin (2006) and colleagues took a similar tack when they required that most teachers in a school sign on to testing the Success for All program at one point in time; new people would presumably sign on as they move into the system. The education sector in the US requires formal agreements that are localized at the school or school district level rather than at the state or federal levels.

In some areas of the health sector, agreements have been able to depend on an overarching theme and organization. Leviton and Horbar (2005) work, for instance, depended partly on the Vermont Oxford Network. Healthcare institutions in such a network commit, in advance, to be willing to engage in research, including perhaps, randomized trials in the interest of contributing to the cumulation of knowledge about what works. Part of the Grimshaw et al. (2005) work also depended on formal networks of people and institutions whose interests lay in collaborating on studies that help us understand what works better.

The trials on Jobs Plus, reported by Bloom (2005) and Bloom and Riccio (2005), depended on identifying entities and people who wanted to improve employment prospects for people in public housing projects and were willing to participate in an experiment to generate evidence on the effects of the intervention. The process, of identifying prospective partners, inviting proposals, and reaching agreement on willingness and capacity to participate, which the authors describe is instructive. The relationships were formalized through written agreements at the level of agencies within the city, such as state and municipal departments responsible for public housing, as well as with the federal entities responsible for regulations. Reaching agreements requires time, talent, and industry.

In the crime research context, for instance, Weisburd (2005) emphasizes the need to develop personal relationships that lead to trust and willingness to experiment on innovations that might work better than conventional practice. In his Jersey City experiment, for instance, the strong involvement of a senior police commander as principal investigator in the study, played a crucial role in preventing a break-down of the experiment after 9 months. In the Jersey City experiment, on the other hand, the Deputy Chief who administered the interventions was strongly convinced of the failure of traditional approaches and the need to test new ones.

The commander took personal authority over the narcotics unit and used his command powers to carefully monitor the daily activities of detectives in the trial. This style of work suggests the importance of integrating “clinical” work and research work in criminal justice, much as they are integrated in medical experiments (see Shepherd 2003). It also reinforces the importance of practitioner “belief” in the importance and necessity of implementing a randomized study. The Kingswood experiment described by Clarke and Cornish (1972) illustrates how doubts regarding the application of the experimental treatment led practitioners to undermine the implementation of the study

Developing relationships in place randomized trials as in many other kinds of field research, depends of course on reputation. The topic invites attention to questions for the future. How do we develop better contracts and agreements with networks of organizations, public and private, ones that permit us to generate better evidence about the effects of an innovation? And how do we develop and publish “model” contracts and memorandums of understanding (MOUs) and make them available to other trialists and their potential collaborators so that they, and we, can learn?

JUSTIFICATIONS FOR A PLACE RANDOMIZED TRIAL

For many social scientists, an important condition for mounting a randomized trial on any innovative intervention that is purported to work is that (a) the effectiveness of conventional practice, policy, or program is debatable, and (b) the debates can be informed by better scientific evidence. In the crime sector, cops, of course, are local theorists, and they disagree about what could work better. Crime experts have also disagreed about what might work in high crime areas. More generally, of course, people disagree with one another about what might work in the policy sector and there is, at times, some agreement that better evidence would be helpful. Berk et al. (1985) took this position in the context of a MacArther Foundation initiative and it has been independently taken by others.

Weisburd's (2003) position, for instance, accords with the justifications for trials in the medical arena: disagreement among experts about the effectiveness of an intervention. This is an important factor in justifying a randomized trial using places or individuals as the units of random allocation. For instance, neither the Salk trials on polio vaccine nor trials on streptomycin for treating tuberculosis would have been mounted had not the purported effects of conventional treatments been suspect. See, for instance, Evans et al. (2006), and the references therein for contemporary history of medical trials and their scientific and ethical justifications and limits. For Leviton and Horbar (2005), the "quality chasm" between what constitutes good healthcare, based on dependable evidence and contemporary medical practice in hospital units, constituted is a major justification for running a trial. Ditto for Grimshaw et al. (2005).

In considering the prospects of a conditional income transfer program in Mexico that Parker and Teruel (2005), Schultz (2004), and others describe, economists disagreed that the effects of an incentive program, such as Progresa/Oportunidades would keep Mexican children away from working in the agricultural fields and increase the likelihood that the children would stay in schools. The economists' disagreements were important in making plain the uncertainty in outcomes. The place randomized trial was mounted partly for this reason.

The United States has a history of attempting to reduce poverty and poverty's consequences. Many of these efforts depend on community-wide or organization-wide efforts in many places. Judging by the Jobs Plus effort described by Bloom and Riccio (2005), for instance, the incentive for people in local departments of housing and other government agencies to participate in the trial include their interest in reducing the problem in a way that introduces good evidence in a sometimes politically volatile context. In a classic paper, Gueron (1985) made a similar point. She reiterated the interest of some public servants and elected officials in discovering how to do better, in the context of employment and training programs and trials on them during the 1970s and early 1980s.

In place based trials of a teacher development program that Porter et al. (2005) described, the broad justifications for the research were U.S. interests in improving mathematics and science education in the middle school grades, especially in large urban school districts. It lay also in the fact that, prior to the Porter et al. work, no sizeable controlled trials on the effects of any teacher professional development had ever been run. More recent trials, undertaken by Garet et al. (2008) at the American Institutes for Research, have involved over 90 schools with similar justification in recruiting schools into the study and in securing funding for the trial.

The scientific justification for place randomized trials is, of course, the assurance that, if the trial is carried out properly, there are no systematic differences between groups of places randomized, which in turn carries a guarantee of statistically unbiased estimates of the intervention's effect. It also assures that chance and chance imbalances can be taken into account and that a legitimate statistical statement of one's confidence in the results can be made. Further, Weisburd (2005) in criminology, Gueron (1985) in labor and welfare economics, Boruch (2007) in education and social sciences, and others have pointed out that the simplicity and transparency of the idea of fair comparison through a randomized trial has strong appeal for policy people and decision makers who cannot understand and do not trust complex model-based analyses of data from nonrandomized studies. For the abiding statistician, the crucial aspect of simplicity is that the statistical inferences as to the effect's size relative to chance need not depend on econometric or statistical or mathematical models. The randomization feature permits and invites less dependence on such speculation, and modern computing methods permit the use of randomization tests.

The empirical evidence on the vulnerability of evidence from nonrandomized trials, in comparison to the evidence from randomized trials, has been building since at least the late 1940s. Assuring that one does not depend on weak and easily assailable evidence when stronger evidence can be produced is an incentive at times in parts of the policy community. Assuring that one does not needlessly depend on heroic assumptions to produce good estimates of effect, assumptions often required in the nonrandomized trials, is an incentive for the scientific and statistical community. See references to recent papers by Deeks on healthcare, Glazerman and others on trials on employment and training, and Duflo and Kraemer on economics that are summarized by Boruch (2007). See also Kunz and Oxman (1998).

Shadish et al. (2008) provided a persuasive illustration that is especially compelling because comparisons between a randomized experiment and an observational study were anticipated as part of their study's design and before the data were collected. Their results suggest comparability, as opposed to major differences, if the quasi-experiment is designed well in a particular domain. Theirs is an interesting and potentially important specific case. More generally, the biases in estimating an intervention's effect, based on the quasi-experiments can be very large, small, or nonexistent (see Lipsey and Wilson (1993) and Weisburd et al. (2001) for instance). The variance in estimates of effects appears to be typically larger in the quasi-experiments than in the randomized tests. So far, and with some narrow exceptions, there is no way to predict the directionality or magnitude of such biases, or in the variances of the estimates, of the intervention's effect, based on a nonrandomized trial.

Studies of whether and by how much estimates of the effect of a nonrandomized trial differ from those of a randomized trial, when *individuals* are the units of random assignment and analysis (as in most medical trials), are important. It remains to be seen whether similar methodological studies on aggregate level analyses using randomized versus nonrandomized places, clusters of individuals, or groups yield similar results, i.e., uncover serious biases in estimating effects or the variance of the estimates, or both. It is however, reasonable to expect biases here also. Bertrand et al. (2002), for instance, focused on biases in estimates of the standard error of effects, assuming no effect at all using conventional different in differences methods, and found Type I error rates that were nine times the error rate presumed (0.05) in using conventional statistical tests. This was partly on account of serial correlation. More methodological research, however, needs to be done on the quasi-experimental approaches to aggregate level units so as to understand when the biases in estimates of effect appear, when the biases in estimates of their standard errors appear, and how large the mean square error is, relative to place randomized trials.

It is important to identify dependable scenarios in which bias and variance of estimates generated in non-randomized trials are tolerable. Doing so can reduce the need for randomized trials (Boruch 2007; Heinsman and Shadish 1998). It is also not easy, as yet, to identify particular scenarios in which bias in estimates of effect or variance will be small, as was the case in some work by Shadish et al. (2008) and Cook et al. (2008). See Berk (2005) for a more general handling of the strengths and weaknesses of randomized controlled trials at the individual level. This kind of work is pertinent to the analysis of data generated in place randomized trials as well. Concerns about the generalizability of results from trials, about attrition and missing data especially across arms of randomized or non randomized trial, and others are important. Further, if the treatment to which a given unit is assigned affects the response of another unit, there is an interference that can bias average treatment effect estimates, and, at a deeper level, implies that an average treatment effect is not even defined. This can be a serious problem in place-based studies when effects on the units nested within the paces are of interest.

The scientific justifications that are identified here are important in the near term. In the long term, it would be good to understand what the other incentives are and to make these explicit at different levels, e.g., policy, institution (agencies), and individual service-provider levels. Incentives for better evidence may differ depending on whether the stakeholders are members of the police force at different levels, the mayor's office, and the community organizations that have a voice, and so on. Many police executives, for instance, want to improve policing and produce evidence on whether things do improve, and also usually want to keep their jobs. The two incentives may not always be compatible if a city council or mayoral preferences are antagonistic toward defensible evidence generated in well-run field tests. Sturdy indifference to dependable evidence of any kind is, of course, a problem in some policy sectors.

DEPLOYING THE INTERVENTION

Implementation, Dimensionalization, and Measurement

Justifications and incentives are essential for assuring that places, and the influential people in them, are willing to participate in a randomized trial. Understanding how to deploy a new program or practice in each place requires more. It requires expertise at ground level and at others.

The Drug Market Analysis Program, which fostered a series of randomized experiments on crime hot spots, suggests that "ordinary" criminal justice agencies can be brought on board to participate in experimental study if there is strong governmental encouragement and financial support that rewards participation (see Weisburd et al. 2006). A similar experience in the Spouse Assault Replication Program (SARP) reinforces these observations. Joel Garner, who served as program manager for SARP, noted that he knew that the program was a success the "day that we got 17 proposals with something like 21 police agencies willing to randomly assign offenders to be arrested" (Weisburd 2005, p. 232)

Understanding how to implement an intervention at the place level in multiple places is no easy matter, one that is apart from the challenge of executing the randomized trial in which the places are embedded. Jobs Plus, for instance, required research teams to engage and guide coordination of the local housing authority, welfare department, workforce development agency, and public housing residents within each site of many sites (Bloom 2005; Bloom and Riccio 2005). The challenge of getting these agencies to work together more closely than any had in the past was piled on top of other challenges to implementation, including getting housing development residents involved as partners despite their inexperience or distrust among themselves and of government agencies, integrating services across providers (so as to meet housing needs and encourage employment in complex welfare environments), and generating job search and acquisition process that fit the place.

For school based trials that aim to test the effect of a program designed to affect the achievement of children, the implementation challenges vary: school bureaucracies differ from bureaucracies in housing, labor/employment, and so on. Nonetheless, Porter et al. (2005) advance the state of the art by developing indices of people's participation in meetings and teams that were major ingredients for change. For these trials, one simple index set includes counting training sessions in which at least one team member participated, computing the average number of participants per session, the proportion of sessions to which local leaders (school principals) contributed, and consistency of people's participation over time. This also included qualitative reconnaissance on factors that impede participation, such as the limited time that schools allow for teachers to meet during work hours.

The challenges to Sikkema et al. (2005), in deploying an AIDS risk reduction program to women in the U.S. housing developments, is similar to others in some respects but differs in others. She and her team developed three different intervention approaches prior to the trial. They used focus groups to reconnoiter the virtues and vulnerabilities of each intervention, relative to local standards of acceptability. The strategy involved identifying opinion leader cadres within housing developments, selecting them, and providing workshops to support participants. Sikkema as others did relied on basic count data, to index level of participation in workshops, and community events. As in the Jobs Plus effort, learning how to negotiate agreements with different influential entities in housing developments, and documenting this, was an essential part of implementation. Sikkema et al relied on explicit and tentative theory of change, as did Flay.

The reports on deploying programs that appear in papers published in peer reviewed journals can be excellent, but they are typically brief. This sparceness of information invites broad questions about how the authors' experience *in detail* can be shared with others, e.g., web based journals, or reports without page limits, workshops, and so on. It invites more scientific attention to the question of how one can dimensionalize implementation and the engineering questions of how to measure implementation level inexpensively and how to establish a high threshold condition for implementation and how to achieve it.

RESOURCES FOR THE TRIAL'S DESIGN, STATISTICAL PROBLEMS, AND SOLUTIONS

Some of the important technical references for trial design and model based data analysis include Donner and Klar (2000) on cluster randomized trials, a fine summary that considers the basic technical ideas and approaches and examples from different countries, mainly in the health arena. The resources include Murray's (1998) book on group randomized trials which is more detailed, focuses more on individuals within groups, considers diverse applications in the U.S. Bloom's (2005) monograph, and is excellent on account of its clarity and use of economic examples in the US. Bryk and Raudenbush's (2002) text, a tour de force, directs attention to model-based statistical analyses of multi-level data that may or may not have been generated by randomized trials. The models are complex and entail assumptions that the analyst may not find acceptable.

In the context of place randomized, cluster randomized, or group randomized trials, we have discovered no books that cover the simplest and *least* model dependent approaches to analyzing data from such a trial. Such approaches fall under the rubric of randomization tests or permutation tests, for which Edgington and Onghena's (2007) book is pertinent. A simple randomization test, for instance, involves computing all possible outcomes of the trial, ignoring the actual allocation to intervention or control conditions, and then making a probabilistic judgment about the dependability of the effect detected, based on the distribution of possible outcomes so generated. There is no dependence on linear or other explicit models.

Given the relatively small sample sizes involved in place randomized trials (less than 200, typically) and contemporary computing capacity, generating the distribution of possible outcomes in these approaches is relatively straightforward. The basic ideas were promulgated by Sir Ronald Fisher in the 1950s for trials in which individuals or plants, for instance, were the independent units of random allocation and analysis. But the idea is directly relevant to trials in which places are randomly allocated and inferences are made about the effects of intervention at the place level. Kempthorne's (1952) classic volume describes the matter better than his mentor (Fisher) did, and Neter et al. (2001) do even better. But their handling of the topic is very brief and does not take into account complex trial designs for which model based approaches are nowadays in common use.

The Campbell Collaboration organized conferences on place randomized trials in 2001 (Bellagio) and 2002 (New York). A special issue of the *Annals of the American Academy of Political and Social Sciences*, covering different social sciences came of this initiative (Volume 599, 2005). After 2002, the William T. Grant Foundation, under Robert Granger's leadership, initiated efforts to enhance the technical capacity of researchers to design such trials and to analyze results through workshops and in other ways described below. During 2002–2009, the Institute for Education Sciences, under the leadership of the IES Director Grover (Russ) Whitehurst and Commissioner Lynn Okagaki, committed funding to institutes and pre- and post-doctoral training programs on designing and executing such trials in education. We are aware of no similar efforts sponsored by the National Institute of Justice or private foundations in the criminological arena.

A major scientific issue in designing place randomized trial, one that is important in policy also, is assuring that the size of the sample of places that are randomized is large enough to permit one to detect relative effects of interventions with some statistical confidence. Put in other words, the concern is that the number of places assigned to different interventions will be sufficient to discern an important effect (or minimally detectable effect size) if, in fact, it occurs. A statistical power analysis, informing us of how many places are needed, is an essential planning device. Such an analysis, however, usually depends on informed speculation about the expected effect size (or the minimum detectable effect size), the randomization, and on particular statistical tests of hypotheses and related procedures. The power analysis may entail assumptions about an underlying model that would be employed in analyzing data from the trial. The models, of course, can be suspect in a variety ways, important and otherwise. User friendly software for estimating the model-based statistical power of a trial under various assumptions about sample size and other factors are accessible on the William T. Grant Foundation's web site (http://www.wtgrantfdn.org). The mathematical underpinnings of the software are based on Raudenbush (1997) and more other papers referenced in Spybrook (2008).

The aforementioned resources are basic but will change as the technology changes. The statistical issues that place randomized trials have uncovered over the last decade, and their resolutions, are also important. Some advances in the area are described below, along with pertinent source references.

In a simple two-level scenario, the number of places that are randomized is generally crucial for assuring statistical power, rather than the number of individuals or other units within the places. This is regardless of whether the approach is based on statistical models or based on inference directly from randomization (Small et al. 2008; Raudenbush 1997). In the past, and at worst, researchers have often wrongly employed the number of people in places for statistical power analysis and statistical hypothesis testing rather than the number of places that were randomly allocated to different interventions. In the Kansas City

Preventive Patrol Experiment, for example, the researchers analyzed survey data obtained at the level of individuals rather than police beats, beats being the units of randomization. Although there are model-based methods for analyzing "nested" units within places (Bryk and Raudenbush 2002), the most dependable (least ambiguous) units of analysis in place based trials for estimating effects are the places that were randomized.

The data generated in place randomized trials are often analyzed using regression models of various kinds rather than randomization tests that are not model dependent. In particular, because of the random assignment, it is commonly assumed that the usual regression models are necessarily correct. Nonetheless, there is usually no explicit justification for the regression model. For a place randomized trial in which the units analyzed are not the ones randomized, of course, the model based assumptions are not necessarily true; the models must be justified.

Even when the analysis is based on outcomes for the places randomized, simply assuming that the regression model is appropriate for analysis can be wrong. In particular, within the Neyman–Rubin framework of randomized trials, each observational unit has a fixed *potential* response under each treatment and control condition. The *actual* response observed depends on the condition to which that unit has been randomly assigned. Because that assignment is by chance mechanism, the observed response, not the potential response is a random variable. The uncertainty then stems solely from the random assignment process and this uncertainty also has clearly defined statistical properties.

When data from randomized trials are analyzed with linear regression, a very different framework is imposed. The uncertainty is represented by the unobservable "disturbances," i.e., the error terms in the regression model. For the regression model to perform as advertised, these disturbances must have certain properties. For instance, each disturbance is characterized by a single common variance, but when the uncertainty is, in fact, a function of the random assignment mechanism, these "certain properties" do not hold. The regression model is automatically wrong in some cases. See Freedman (2006, 2008a, b). The question then is how misleading the results of a regression analysis are likely to be relative to straightforward randomization tests. Empirical answers are very difficult to generate. With large samples, the bias in estimated average treatment effects will be small, but in place randomized trials, the number of randomized units is often modest. Even with very large samples, the standard errors can be systematically too small or too large (Freedman 2006, 2008a). Analyses based on regression models with categorical, count, or time-to-failure outcomes create additional problems (Freedman 2008b), and there are alternatives such as a simple t test for mean differences (which is tantamount to a regression with only one or more indicator variables for the intervention) or Fisher's exact test for proportions. The problems with conventional regression materialize when covariates are included in the model, and, in the case of hierarchical modeling, the problems may materialize in other ways.

Confusion often arises regarding the number of cases that are employed for statistical analysis. For instance, place base trials often involve the study of people within places, and thus, one may have a relatively small number of units at the randomized place level and a very large number of units at the nested individual level. For example, the Redlands Valley Risk Focused Policing Experiment randomized 26 census block groups for intervention, with 13 block groups assigned to each arm of the trial. The research included 800 individuals who were studied within the block groups (Weisburd et al. 2008). Simply including all individuals in an analysis that ignored the place level random assignment would have led to an underestimation of the standard error of the intervention's effects, and (consequently), a positively biased test of statistical significance.

A key issue in this scenario is that the individuals or other units within a place were not assigned to treatment or control conditions. Indeed, all individuals within each place are automatically assigned the same condition, either treatment or control, for instance, in a place randomized trial. This scenario does not mean that one cannot analyze data on individuals or entities within places that are randomized in a place randomized trial but as suggested above, the analysis is riskier (Bloom and Riccio 2005; Bryk and Raudenbush 2002; Spybrook et al. 2004). In the Redlands study, the main research question focuses on examining the impact of the Redlands intervention on individual level outcomes. In analyzing the data, the researchers acknowledged and took into account the fact that although treatment was assigned and administered at the block group level, the study was intended to examine individual level outcomes. They could be fairly certain that pre-existing differences between the treatment and control block groups in the study were not systematic due to the matching and subsequent randomization process. However, they could not make the same claim regarding individuals who reside within the treatment and control block groups because the individuals were not randomly assigned to condition. In turn, when the researchers examined the effects of the treatment on individuals, unmeasured characteristics of the groups could, of course, influence estimates of individual level effects.

Prior techniques for clustered data involved either aggregating all information to the group level (e.g., block group level) as in a straight up randomization test or t test of mean differences, or disaggregating all block group level traits to the individual, which involves assigning all block group level traits to the individual (e.g., including a dummy variable for each individual for membership in a specific census block group). For Bryk and Raudenbush (2002), the problem with the first method is that all within group information is wasted and omitted from the analysis and this can be as much as 80–90% of the total variation. The problem with the second method is that the observations are no longer independent, as we know that all individuals within a certain block group will have the same value on a certain variable. Specifically, individuals from different block groups are independent, but individuals within the same block groups will share values on many more variables, some of which will be observed and some, not observed. The effects of the omitted or unobserved variables are absorbed by the error term in a linear regression model, resulting in the correlation between individual level error terms. To account for the dependence that arises when using samples of individuals nested within block groups, Weisburd et al. (2008) used several statistical methods that attempt to correct for clustering, such as hierarchical linear modeling techniques (i.e., HLM) and generalized least squares estimation techniques with robust standard errors. These techniques are discussed in more detail in other parts of the volume.[2]

Clearly, one can properly analyze place-based randomized experiments in a manner that is largely model free. Small et al. (2008), for instance, emphasize randomization/permutation tests with different kinds of adjustments for covariates, none relying on any form of HLM. Imai et al. (2009) advance the state of the art by showing how pair-wise matching in place randomized trials can often enhance precision in estimates of the effects of interventions, increase statistical power of analysis in detecting effects, and better assure unbiasedness of the estimates of variability of the effect with a relatively small number of places (clusters, in their vernacular).

[2] In the Redlands study the researchers also conducted aggregate level (block level) post test analyses to assess aggregate level delinquency and substance abuse differences between experimental and control block groups. Results from this set of analyses closely mirror results generated using the HLM approach.

In criminological and other trials that are place based, matched pairs designs can be used that build on a large number of recent developments (Greevy et al. 2004). But the analysts concern has usually been that the matching may unnecessarily reduce degrees of freedom, and thus, the statistical power of the studies. There is a loss of one degree of freedom for each restriction (pair or block) in a study design.

The Redlands Valley Risk Focused Policing Experiment, for instance, used the matched pairs approach, but it is often not possible to define clearly in crime and justice work, the relationship between "blocking" or matching variables and the outcomes observed. In the Minneapolis Hot Spots Experiment (Sherman and Weisburd 1995) and the Jersey Drug Hot Spots Experiment (Weisburd and Green's 1994), for instance, subjects were classified into broad groups ("blocks"), and then randomization was carried out within the groups or blocks (see chapter on Block Randomized Trials). This approach allowed the researchers to decrease heterogeneity within the groups but led to a loss of only a few degrees of freedom in the study design. A fully matched pair design will provide benefit over a partially blocked design where there is strong knowledge that allows careful distinctions of place units in terms of the impacts of the blocking factors.

As a practical matter, the Imai et al. (2009) work shows that the number of units within a place (cluster) are an important matching variable per se, and that using it can reduce costs and increase statistical power of the trial. That is, when the numbers of units within randomized places vary, the numbers can be exploited in a matching process to produce a better trial design.

Imai et al. (2009) work updates and generalize earlier methodological work done mainly in health and in education where schools or classrooms are the units of randomization. Data on students at the schools level may be used to enhance precision in estimates, i.e., covariates at the place level. Within-school level data can be used to model speculated processes at the student level within school. Schochet (2009) reviews such work upto about 2008, provides numerous references, and attend to different experiment designs involving data at school, district, classroom, and student levels. Konstantopoulos (2009) incorporates cost consideration into power analysis of three level designs. Spybrook (2008) considers power reporting issues in funded federal projects in education.

Bloom and Riccio (2005) advanced the state of the art in another important respect. They couple conventional comparison of randomized places with time series analyses of data from the places. This approach permits one to also take into account the changing composition of individuals and conditions across the places (housing developments). The idea of coupling randomized trials and nonrandomized approaches to estimating effects of interventions is not new in the sense that others have done conceptual work on it. But actualizing the ideas and being thorough in the design and analysis is unusual.

Since about 2005, the idea of "step wedge" and "dynamic wait list" designs has emerged as an interesting addition to the trialists' armamentarium in estimating short term effects of interventions. The most familiar variation on this idea is a wait list design in which a random half sample of the eligible places are assigned to the intervention for some period of time. The remaining half sample receives the intervention after this period is ended. The critical policy presumption is that all units will eventually receive the intervention. The important assumption for some statistical analyses is that the time delay is inconsequential for simple comparisons. Put in other words, the policy presumption is usually based on political or ethical concerns that all places should eventually receive a purportedly desirable intervention, or on logistical concerns about the need to stage an intervention's deployment to a large number of units. The import of the statistical assumption regarding inconsequential effect of delay in an intervention is often unclear.

The general design approach in such trials is to plan how to randomly deploy the intervention to subsets of the target sample of places periodically and over some specified period of time, so as to achieve a given level of precision in estimating an effect of a specified size. A set of 32 places, for instance, might be divided randomly into eight sets of four places. Each set is then randomly subjected sequentially to the intervention at eight time points. The earliest set so assigned then receives the intervention for the longest period of time. Relevant advances in this kind of experimental design here have been health related. For instance, Brown et al. (2006) develop and illustrate dynamic wait list designs and apply the analytic work in the context of a suicide prevention program in Atlanta schools. Their paper shows how statistical power is enhanced over a simple two-group delayed intervention design and that power increases with the number of times a new subset is introduced to the intervention, using count data (incidence) as the outcome variable. In a related stream of work, Hussey and Hughes (2006) develop the idea of step wedge designs, the "steps" being the point at which a subset of places (clusters) is randomly assigned to the intervention, and apply the ideas to a trial in the state of Washington. For these authors, the design is an expansion of cross over designs. They too show how increasing the number of steps can enhance statistical power of a formal test of hypothesis, conditional on a statistical model of how nature works, and how effect size attributable to increasing exposure to the intervention can be taken into account in modeling the underlying processes. De Allegri et al. (2008) provide an illustration in the context of deploying a community based health insurance intervention in West Africa.

More generally, the Cochrane Collaboration's methods group served as the auspices for a systematic review of recent applications of the step wedge design (Brown and Lilford 2006). The step wedge and dynamic wait list designs are innovative and have very attractive features. But the power analyses and data analyses, to date, have relied on hierarchical linear models with unusual assumptions that are not always testable. There may be alternatives that are far less model dependent.

REGISTERS OF RANDOMIZED TRIALS AND STANDARDS OF REPORTING

Learning about place randomized trials, whether completed or underway, can be difficult. The authors of studies cited here, for instance, have contributed to place randomized trials in China, Canada, England, the United States, Kenya, Peru, Mexico, Colombia, Thailand, and elsewhere. Their reports have been published in a variety of scientific journals that differ, sometimes remarkably, from one another, e.g., in criminology, education, prevention science, medicine, and so on. This variety is a challenge for searchers of published and unpublished literature. Further, relying on web based searches such as ERIC and PsychInfo have not fared well relative to hand searches of the research literature, at least in regard to locating randomized trials in the social sector (Turner et al. 2003). More recently, Taljaard et al. (2009)

Found that fewer than 50% of published cluster randomized trials in health are appropriately classified as cluster trials in titles or abstracts, contrary to classification recommendations by the 2004 CONSORT statement. They had to rely on over 50 search terms for "places" to locate relevant reports. Even in the health sector then, electronic searches of electronic databases are likely to be a challenge despite advances in machine based approaches.

The difficulty in identifying randomized trials was reduced with the creation of the international Cochrane Collaboration in healthcare in the early 1993 and the international Campbell Collaboration in 2000. Both organizations, which rely heavily on voluntary participation, developed registers on randomized trials that could be accessed through their web sites. Both organizations relied heavily on hand searches (full text readings) of peer reviewed journals rather than on conventional machine based (key word, abstract, title) searches.

For instance, between 2002 and 2005, the Campbell Collaboration Social, Psychological, Educational, and Criminological Trials Register (C-SPECTR) contained about 13,000 entries on randomized and possibly randomized trials and included references to about 200 reports on place randomized trials. In C2 SPECTR, one could find entries on place randomized trials involving schools and classrooms in Kenya, El Salvador, and India, brothels in Thailand, factories in Russia, barrios (impoverished neighborhoods) in Colombia, and *lins* (neighborhoods) in Taiwan, among others. Unfortunately, the SPECTR registry has not received continued support and its operation was suspended. The Cochrane Collaboration library, on the other hand, has been supported and is routinely updated. It contains over 500 references to such trials to judge from our search based on the phrase "cluster randomized" in 2009.

In the United States, one can learn about trials in the health sector, including place randomized trials, from http://www.clinicaltrials.gov. It is a resource that criminologists and others may find helpful at times, when the interventions being tested or the outcome variables being examined are pertinent, or when information about the place trial's design can be exploited. In 2007, David Greenberg and Mark Shroder developed a new register of randomized trials, oriented toward tests of economic interventions under the auspices of the Social Science Research Network (SSRN). Their Randomized Social Experiments Abstracts are given at ERN@publish.ssrn.com.

Standards for reporting on randomized trials that involve individuals as the unit of random allocation and analysis are a product of the late 1990s. See an illustration and references to early work in Boruch (1997). More recently, the CONSORT statement provided guidance on the ingredients that a good report on a randomized trial in the health arena should contain (Moher et al. 2001). The CONSORT statement has been modified to include standards for reporting on place/group/cluster randomized trials. This extension, given in Campbell et al. (2004), covers important topics embedded in a place randomized trial. These include, for instance, consideration of the rationale for this research design and the fact that there are often at least two levels of sampling and inference: the cluster (place) level and the individual (person within place) level, different rates of attrition and measurement at each level, and so on. For instance, an experiment that involves bar rooms (places) as the primary unit of random allocation in a study of an intervention to reduce violence, may also obtain information on the individuals within bars who are a party to violence. The randomized bar room level of analysis can produce unbiased estimates of the intervention's effects. The within bar data may help to speculate about the estimates of effect and inform secondary analyses that aid in the interpretation of effects. Standardized reporting at each level, and perhaps others, is important for scientific understanding.

As of this writing, there are no uniform standards for reporting on randomized trials in crime and justice or related areas. This problem has been raised by Perry in this volume. Empirical work on the problem is reported by Perry and Johnson (2008), by Perry et al. (2009), and by Gill (2009) who assessed reporting methods in crime and justice journals. Farrington (2003) and, much earlier, Loesel and Koferl (1989), noted the lack of reporting standards. Farrington coined the term "reporting validity" to emphasize that such standards are critical for developing an experimental science in criminology.

The lack of clear and uniform standards for reporting on results of trials in criminology and in related disciplines means that valid reanalysis of experimental studies is difficult, and often impossible. Further, and as important, there are no professional standards for assuring access to the original data used in a published report on results of a trial, place randomized or not. Reanalysis at the report level or micro-record level, for instance, is often essential for verifying an original analysis, testing new hypotheses, or exploring new theories or assumptions. Dozens of reanalyses of data from Progresa/Oportunidades, for instance, have been reported in published papers, dissertations, and research monographs. These reports are on diverse topics, such as verifying the apparent equivalence of the randomized groups and conventional intent to treat impact analyses, and examining many, often elaborate structural modeling efforts that get at different aspects of the program's impact on outcomes. These have made the trial famous in the world of economics of education.

Some funding agencies in the US, such as the National Institute of Justice and the National Science Foundation, have tried to make policy to assure that independent analysts have access to micro record data generated in the research that they sponsor, including data from controlled trials. The Institute of Education Sciences policy on this is unclear as of this writing. Regardless of clarity of policy, the actualization of the intent – to make micro-records accessible – seems uneven. Given the intellectual and financial investment in the production of such data, it seems sensible to try to improve independent researchers' access to it for secondary analysis. Regardless of the question of access to micro-records, criteria like those produced by the CONSORT statement for uniform reporting on randomized trials are critical for advancing experimental criminology.

ETHICS AND LAW

As of this writing, no professional society or government agency has promulgated explicit statements about the ethical propriety of place randomized trials. Contemporary rules and ethics statements, for instance, attend to the rights and well being of individuals, rather than places, groups, or entities that are engaged in a trial. The earliest fundamental statements on research ethics, such as the Helsinki Statement and the Belmont Report direct attention to individuals' rights. Contemporary government regulations regarding human subjects research in the US, Canada, the UK, and other countries also do not consider place/cluster/group trials explicitly.

Applying contemporary standards to place randomized trials can be awkward and imperfect. See, for instance, the report of the Planning Committee on Protecting Student Records and Facilitating Education Research (2009) and Boruch's (2007) attempt to the Federal Judicial Center's guidelines to understand the ethical justification for the Progresa/Oportunidades place randomized trial.

At times, contemporary professional standards of ethics and governmental regulations are not clearly relevant to place randomized trials. At other times, standards and regulations may be only partially relevant. This invites exploration. Taljaard et al. (2008, 2009) have taken the lead in an effort based at University of Ottawa, to understand and explicate the ethics issues engendered by place/cluster/group randomized trials. They attend mainly to health and risk reduction in health and school based risk prevention settings. Sabin et al. (2008) took another approach in exploring the ethics of cluster trials in the context of drug testing in health plans. The issues that Taljaard et al. (2008, 2009) and Sabin et al. (2008) confront are likely to be relevant to criminological place randomized trials.

To illustrate one issue, the commonly used phrase "human subjects" is arguably inappropriate in settings in which a city's crime hot spots are the targets for random allocation and the interventions are approved by a city council or police department under their administrative authority. The target of research is jurisdictions in which people happen to commit crime, reside, or do business. A recent Institute of Medicine Report has taken issue with the use of the phrase, "human subjects" for different reasons, maintaining that the word "participant" is arguably more accurate and less rebarbative.

The concerns of an institutional review board (IRB) may be moot when the law determines actual random allocation of places to different interventions and the individual's rights are entrained in the determination or, the IRB's concern may be pertinent in that the trialist may seek detailed information which goes beyond that obtained in conventional administrative/police record systems, such as attitudes, beliefs, wages, etc. Most importantly, perhaps, the matter of whose consent ought to be sought in place randomized trial is not easily resolved. It may be the "agent" responsible for the place, such as a police chief or director of a health plan, or it may be individuals within the place such as teachers within schools that are testing curriculum packages, or it may be both, and the responsibility may lie with entirely different agents.

Boruch et al. (2004) suggested that it may be possible to avoid some ethical and moral dilemmas commonly associated with experimentation by randomly allocating at the organizational or place level, rather than randomly allocating individuals. At first glance, one might question why the change in unit of analysis should affect ethical concerns. Why should it matter, for example, whether students in a specific school are allocated to treatment and control conditions versus all students in specific schools? The end result is the same. Some individuals will gain treatment and others will not. However, where individuals do not experience the inequality of treatment directly (e.g., by seeing other students in their school being treated differently), ethical issues may just not be raised.

The general proposition that place based studies are likely to be faced with relatively few ethical objections is illustrated by both the Minneapolis Hot Spots Experiment (Sherman and Weisburd 1995) and the Jersey City Drug Hot Spots Experiment (Weisburd and Green's 1994). City officials in these studies did not raise significant ethical concerns during negotiations over randomly allocating either crime hot spots or drug markets to intervention and control conditions. Sherman, for example, notes that "no one ever raised ethical objections" to the Minneapolis Hot Spots Experiment (Weisburd 2005, p. 233). Moreover, as neither study collected information directly from individuals within hot spots (human subjects), but relied rather on official police data and observations of the sites, they were not subject to significant human subject review. This contrasts strongly with controversies often surrounding the random allocation of individuals to different interventions in some criminal justice settings.

While common ethical objections to random allocation did not surface in the police hot spots studies, different types of objections were raised by citizens and the police. The objections are suggestive of more serious problems that might develop. These objections can be construed as a part of a political or institutional ethic concerning group rights, as opposed to an ethic that concerns individual rights. For example, in Minneapolis, the City Council was asked to approve the reallocation of police resources in the hot spots experiment. One city councilman in a low crime area would not give his approval unless "an early warning crime trend analysis plan" would monitor burglary trends and send more patrols back into his neighborhood if "burglary rashes developed" (Weisburd 2005). Monitoring did not reveal such increases in burglary, and thus, the experiment was not affected.

In the Jersey City experiment, when a citizens group in one area of Jersey City found out that their neighbors were getting extra police attention, they demanded to be made part of the hot spots study. The police convinced the citizens group that they continued to get good police service, but that their problem (to their benefit) was not sufficiently serious to make them eligible to join the experiment. In Jersey City, all of the drug areas that showed consistent and serious activity were included in the study. This coverage, combined with the equality of "service" (this was a treatment and "comparison" group design in which some type of treatment was delivered to every site) in the experimental and control areas made it possible to avoid objections that some serious drug markets were receiving more police attention than others. Nonetheless, the rule that "(e)xperiments with lower public visibility will generally be easier to implement" (Weisburd 2005, p. 186) appears particularly relevant to place randomized trials.

A more complex problem was raised by police officers participating in the Minneapolis study. Many patrol officers objected to the hot spots approach of "sitting" in or riding through specific areas. While researchers tried to draw support from rank and file police officers for the experiment through briefings, pizza parties, and the distribution of t-shirts bearing the project logo ("Minneapolis Hot Spot Cop"), many officers argued that the hot spots approach was unethical and violated their obligations to protect the public. In particular, the officers argued that the intensive approach to handling crime hot spots allowed crime to shift to around the corners from the hot spots. The officers' theory is important. In fact, and in opposition to the theory, there were no wide scale attempts by officers to undermine the experiment and no displacement of crime activity from one are a to another in this trial (Weisburd et al. 2006). But the early objections by officers, in this instance, are similar to the practitioners' concerns that undermined the Kingswood study that forms the basis for Clarke and Cornish's (1972) well known critique of experimentation in the Home Office.

Regardless of the ethical dilemmas engendered by place based trials over the last decade, there appear to have been no serious challenges in the US courts to the conduct of place randomized trials. For instance, the former director of the Institute for Education Sciences (IES), Russ Whitehurst, reported in a personal communication (2008), that he had encountered no court challenges as a consequence of the IES's sponsoring many such trials in education during 2002–2008. Similarly, we are aware of no judicial challenges in the context of place randomized trials in the crime sector, such as the trials on crime hot spots, bar room violence, convenience store vulnerability to hold ups, and so on.

CONCLUSIONS AND IMPLICATIONS FOR THE FUTURE

Place randomized trials have become important in criminology, health, education, prevention research, and other sectors. This is because they employ substantive theory about the effects of intervention at the place level, which is different from theory that is focused on individuals. It is because the statistical technology for designing such trials has advanced remarkably since the late 1990s. Their importance has increased because the resources and capacity for mounting such trials have increased.

The interest in place randomized trials transcends academic disciplines. Such trials are being mounted in the health sector to understand how hospital units can be encouraged to change practice in the interest of enhancing people's health or reducing costs of healthcare. They have been mounted in the education sector to judge the effectiveness of curriculum packages in elementary, middle, and high schools, and to evaluate risk prevention programs that focus on substance abuse and non-criminal disorder. They have, of course, been employed in criminological research to understand how to reduce crime or prevent it. One implication for the future is that cross discipline work will enhance understanding: the problems of engaging places in a trial, assuring ethical propriety and good statistical design and analysis transcend discipline. For instance, criminological place randomized trials, with few exceptions, have not attended to health (injury) outcomes as a variable, nor have health trials paid much attention to crime interventions. Prevention research that is school based is a remarkable exception in cutting across disciplines.

The statistical armamentarium for design and analysis of place randomized trials is fundamental. Its origins lie in the simple idea of randomization and permutation tests that do not depend on complex statistical models. The statistical models and methods that get beyond a basic test of the statistical dependability of the intervention's effects in a place randomized trial has also developed apace. Statistical theory and the development of relevant software for power analysis and data analysis have improved and are accessible on account of some private foundation and US federal agency efforts. Specialized institutes for advanced study have become more common over the last decade. For the criminological community, it is a fine prospect to look to other advances in training and in statistical design and analysis of such trials.

The experience of people who have been involved in the design, execution, and analysis of place randomized trials is an important source of intellectual and social capital. Part of the future lies in exploiting and building on this capital, in graduate and post graduate education, for instance. The future lies also in producing published documentation on negotiated agreements, MOUs, "model agreements," etc., among criminologists, police, and other stakeholders, such as community groups that have agreed to participate in such trials.

Ethics in a place randomized trial are important. The ethical issues in such a trial arguably differ from those encountered in a trial in which individuals, rather than entities, are randomly allocated to different interventions. The advanced thinking lies in exploring the implications of randomization of entire counties, businesses, schools, and crime hot spots to different interventions when the expected effects of the interventions' effects are in "equipoise." The issues remain unclear until colleagues who are doing advanced work on the ethics of cluster/place/group randomized trials in health make their empirical findings and intellectual understandings known.

Standardized reporting on the design, execution, and analysis of the data from place randomized trials has become improved in health research and in education research. Reporting has not been standardized in criminological research and related social research sectors, including social services. Criminology's future lies partly in developing better and uniform reporting standards. In addition, learning how to assure better access to well documented microdata stemming from randomized trials will be important for reanalysis and secondary analysis, especially in controversial studies.

Walter Lippmann, an able social scientist and newspaper writer, had a strong interest in cops, and crimes by adults and adolescents, and was familiar with political ambivalence about or opposition to sound evidence. He was a street level criminologist, remarkable writer, and good thinker. In the 1940s, Lippmann (1963) said: "The problem is one for which public remedies are most likely to be found by choosing the most obvious issues and tackling them experimentally... the commissions of study are more likely to be productive if they can study the effects of practical experimentation." Nowadays, trialists in criminology would have little difficulty in subscribing to Lippmann's counsel.

REFERENCES

Berk R (2005) Randomized experiments as the bronze standard. J Exp Criminol 1(4):417–433

Berk R, Boruch R, Chambers D, Rossi P, Witte A (1985) Social policy experimentation: a position paper. Eval Rev 9(4):387–430

Bertrand M, Duflo E, Mullainathan S (2002) How much should we trust differences in differences estimates? Working Paper 8841. National Bureau of Economic Research, Cambridge, MA.

Bloom HS (2005) Learning more from social experiments: evolving analytic approaches. Russell Sage Foundation, New York

Bloom HS, Riccio JA (2005) Using place random assignment and comparative interrupted time-series analysis to evaluate the jobs-plus employment program for public housing residents. Ann Am Acad Pol Soc Sci 599:19–51

Boruch RF (1997) Randomized experiments for planning and evaluation. Thousand Oaks, California, Sage Publications

Boruch RF (2007) Encouraging the flight of error: ethical standards, evidence standards, and randomized trials. New Dir Eval 133:55–73

Boruch RF, May H, Turner H, Lavenberg J, Petrosino A, deMoya D, Grimshaw J, Foley E (2004) Estimating the effects of interventions that are deployed in many places. Am Behav Sci 47(5):575–608

Botvin G, Griffin K, Diaz T, Scheier L, Williams C, Epstein J (2000) Preventing elicit drug use in adolescents. Addict Behav 25(5):769–774

Braga A (2001) The effects of hot spots policing on crime. In: Farrington DF, Welsh BC (eds) What works in preventing crime? Special issue. Ann Am Acad Pol Soc Sci, vol 578, pp 104–125

Brown CH, Wyaman PA, Guo J, Pena J (2006) Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide. Clinical Trials 3:259–271

Brown C, Lilford R (2006) The step wedge trial design: a systematic review. BMC Med Res Methodol 6:54. Available at: http://www.biomedcentral.com/info/1471–2288/6/54

Bryk AS, Raudenbush SW (2002) Hierarchical linear models. Sage, Thousand Oaks, CA.

Campbell MK, Elbourne DR, Altman DG (2004) CONSORT statement: extension to cluster randomized trials. BMJ 328:702–708

Clarke RV, Cornish D (1972) The controlled trial in institutional research: paradigm or pitfall for penal evaluators. London, Her Majesty's Stationary Office

Cook T, Shadish W, Wong V (2008) Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within study comparisons. J Policy Anal Manage 27(4): 724–750

Davis R, Taylor B (1997) A proactive response to family violence: the results of a randomized experiment. Criminology 35(2):307–333

De Allegri M et al (2008) Step cluster randomized community-based trials: an application to the study of the impact of community health insurance. Health Res Policy Syst 6:10. Available at: http://www.health-policy-systems.com/content/6/1/10

Donner A, Klar N (2000) Design and analysis of cluster randomized trials in health research. Arnold, London

Evans I, Thornton H, Chalmers I (2006) Testing treatments: better research for healthcare. British Library, London

Edgington E, Onghena P (2007) Randomization tests (Fourth Edition), New York, Chapman and Hall/CRC

Farrington DP (2003) Methodological standards for evaluation research. Ann Am Acad Pol Soc Sci 587(1):49–68

Flay BR, Collins LM (2005) Historical review of school based randomized trials for evaluating problem behavior prevention programs. Ann Am Acad Pol Soc Sci 599:115–146

Freedman DA (2006) Statistical models for causation: what inferential leverage do they provide. Eval Rev 30(5): 691–713

Freedman DA (2008a) On regression adjustments to experimental data. Adv Appl Math 40:80–193

Freedman DA (2008b) Randomization does not justify logistic regression. Stat Sci 23:237–249

Garet M et al (2008) The impact of two professional development interventions on early reading instruction and achievement. Institute for Education Sciences, US Department of Education and American Institutes for Research, Washington DC

Gill C (2009) Reporting in criminological journals. Report for seminar on advanced topics in experimental design. Graduate School of Education and Criminology Department, University of Pennsylvania, Philadelphia, PA

Graham K, Osgood D, Zibrowshi E, Purcell J, Glicksman K, Leonared K, Perneanen K, Alitz R, Toomey T (2004) The effect of the safer bars programme on physical aggression in bars: results of a randomized trial. Drug and Alcohol Review 23:31–41

Greevy R, Lu D, Silber JH, Rosenbaum P (2004) Optimal multivariate matching before randomization. Biostatistics 5(2):263–275

Grimshaw J, Eccles M, Cambpell M, Elbroune D (2005) Cluster randomized trials of professional and organizational behavior change interventions in health settings. Annals of the American Academy of Political and Social Science 599:71–93

Gulematova-Swan M (2009) Evaluating the impact of conditional cash transfer programs on adolescent decisions about marriage and fertility: the case of oportunidades. PhD Dissertation, Department of Economics, University of Pennsylvania

Gueron JM (1985) The demonstration of state/welfare initiatives. In: Boruch RF, Wothke W (eds) Randomization and field experimentation. Special issue of new directions for progam evaluation. Number 28. Jossey Bass, San Francisco, pp 5–14

Heinsman D, Shadish W (1998) Assignment methods in experimentation: when do nonrandomized experiments approximate answers from randomized experiments? Psychol Methods 1(2):154–169

Hussey M, Huges J (2006) Design and analysis of stepped wedge cluster designs. Contemporary Clinical Trials, doi: 1.1016/j.cct2006.05.007

Imai K, King G, Nall C (2009) The essential role of pair matching in cluster randomized experiments, with application to the Mexican universal health insurance evaluation. Stat Sci 24(1):29–53

Kelling G, Pate A, Dieckmann D, Brown C (1974) The Kanasas city preventive police patrol experiment. Washington DC, The Police Foundation

Kempthorne (1952) The design and analysis of experiments. New York: Wiley, and Malabar Florida: Robert E. Krieger Publishers (Reprint 1983)

Konstantopoulos S (2009) Incorporating cost in power analysis for three-level cluster-randomized designs. Eval Rev 33(4):335–357

Kunz R, Oxman A (1998) The unpredictability paradox: review of the empirical comparisons of randomized and non-randomized clinical trials. BMJ 317:1185–1190

Leviton LC, Horbar JD (2005) Cluster randomized trials for evaluation of strategies to promote evidence based practice in perinatla and neonatal medicine. Annals of the American Academy of Political and Social Science 599:94–114

Lippmann W (1963) The young criminals. In: Rossiter C, Lare J (eds) The essential Lippmann. Random House, New York, Originally published 1933

Lipsey M, Wilson D (1993) The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. Am Psychol 48:1181–1209

Loesel F, Koferl P (1989) Evaluation research on correctional treatment in West Germany: a meta-analysis. In: Wegerner H, Loesel F, Haisch J (eds) Criminal behavior and the justice system: psychological perspectives. Springer, New York, pp 334–355

Mazzarole L, Price J, Roehl J (2000) Civil remedies and drug control: a randomized trial in Oakland California. Eval Rev 24(2):212–241

Merlino J (2009) The 21st century research and development center on cognition and science instruction. Award from the Institute of Education Sciences, U.S. Department of Education. Award Number R305C080009. Author, Conshahocken, PA

Moher D, Shulz KF, Moher D, Egger M, Davidoff F, Elbourne D et al (2001) The CONSORT statement: revised recommendations for improving the quality of reports on parallel group randomized trials. Lancet 387: 1191–2004

Murray D (1998) Design and analysis of group randomized trials. Oxford University Press, New York

Neter J, Kutner M, Nachtsheim C, Wasserman W (2001) Applied linear statistical models. McGraw Hill, New York

Parker SW, Teruel GM (2005) Randomization and social program evaluation: the case of progresa. Annals of the American Academy of Political and Social Science 599(May):199–219

Perry AE, Johnson M (2008) Applying the Consolidated Standards of Reporting Trials (CONSORT) to studies of mental health for juvenile offenders: a research note. J Exp Criminol 4: 165–185

Perry AE, Weisburd D, Hewitt C (2009) Are criminologists reporting experiments in ways that allow us to access them? Unpublished Manuscript/Report. Available from the Authors: Center for Evidence Based Crime Policy (CEBCP), George Mason University, Virginia US

Planning Committee on Protecting Student Records and Facilitating Education Research (2009) Protecting student records and facilitating education research. National Academy of Sciences, Washington DC

Porter AC, Blank RK, Smithson JL, Osthoff E (2005) Place randomized trials to test the effects of instructional practices of a Mathematics/Science professional development program for teachers. Annals of the American Academy of Political and Social Science 599:147–175

Raudenbush S (1997) Statistical analysis and optimal design for cluster randomized trials. Psychol Methods 2(2): 173–185

Sabin JE, Mazor K, Metereko V, Goff SL, Platt R (2008) Comparing drug effectiveness at health plan: the ethics of cluster randomized trials. Hastings Cent Rep 35(5):39–48

Schochet P (2009) Statistical power for random assignment evaluations of education programs. J Educ Behav Stat 34(2):238–266

Schultz TP (2004) School subsidies for the poor: evaluating the Mexican Progresa poverty program. J Dev Econ 74:199–250

Shadish WR, Clark MH, Steiner PM (2008) Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. J Am Stat Assoc 103(484):1334–1356

Shepard J (2003) Explaining feast or famine in randomized field experiments: medical science and criminology compared. Evaluation review 27:290–315

Sherman LW, Wesiburd D (1995) General deterrent effects of police patrol in crime "Hot spots." A Randomized Controlled Trial. Justice Quarterly 12:626–648

Sikkema KJ (2005) HIV prevention among women in low income housing developments: issues and intervention outcomes in a place randomized trial. Ann Am Acad Pol Soc Sci 599:52–70

Slavin R (Oct 6 2006) Research and Effectiveness. Education Week

Small DS, Ten Have TT, Rosenbaum PR (2008) Randomization inference in a group-randomized trial of treatments for depression: covariate adjusted, noncompliance, and quantile effects. J Am Stat Assoc 103(481):271–279

Spybrook J (2008) Are power analyses reported with adequate detail? J Res Educ Eff 1:215–235

Taljaard M, Grimshaw J, Weijer C (2008) Ethical and policy issues in cluster randomized trials: proposal for research to the CIHR. Authors: University of Ottawa, Ottawa, Canada

Taljaard M, Weijer C, Grimshaw J, Bell Brown J, Binik A, Boruch R, Brejhaut J, Chaudry S, Eccles M, McRae A, Saginur R, Zwarenstein M, Donner A (2009) Ethical and policy issues in cluster randomized trials: rational and design of a mixed methods research study. Trials 10:61

Turner H, Boruch R, Petrosino A, Lavenberg J, de Moya D, Rothstein H (2003) Populating an international web based randomized trials register in the social, behavioral, criminological, and education sciences. Ann Am Acad Pol Soc Sci 589:203–225

Warburton AL, Sheppard JP (2000) Effectiveness of toughened glass in terms of reducing injury in bars: a randomized controlled trial. Inj Prev 6:36–40

Weisburd D (2003) Ethical practice and evaluation of interventions in crime and justice: the moral imperative for randomized trials. Eval Rev 27(3):336–354

Weisburd D (2005) Hot spots policing experiments and lessons from the field. Ann Am Acad Pol Soc Sci 599: 220–245

Weisburd D, Green L (1994) Defining the drug market: the case of the Jersey city DMA system. In McKenzie DL, Uchida CD (eds) Drugs and crime: evaluating public policy initiatives. Thousand Oaks California, Sage Publications

Weisburd D, Lum C, Petrosino A (2001) Does research design affect study outcomes? Ann Am Acad Pol Soc Sci 578:50–70

Weisburd D, Wycoff L, Ready J, Eck JE, Hinkle JC, Gajewski F (2006) Des crime move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44(3):549–591

Weisburd D, Morris N, Ready J (2008) Risk focused policing at places: an experimental evaluation. Justice Q 25(1):200