

# Estimating the Effects of Interventions in Multiple Sites and Settings: Place-based Randomized Trials<sup>1</sup>

robert boruch, ellen foley & jeremy grimshaw

## 1. Introduction

A place-based trial here means a study in which a number of places or organizations are randomly assigned to one of two or more interventions so as to learn which intervention works best. The »places« may be villages or neighborhoods, schools or juvenile facilities, housing projects, or other organizations. The places that are assigned to interventions will not differ at the outset. They are statistically equivalent on account

---

Robert Boruch is University Chair Trustee Professor at the Graduate School of Education, the Statistics Department of the Wharton School of Business, and the Fels Center for Government at the University of Pennsylvania.

Ellen Foley is the Senior Associate in District Redesign at the Annenberg Institute for School Reform, Brown University and serves as Research Director for School Communities that Work: A National Task Force on the Future of Urban Districts.

Jeremy Grimshaw is the Director of the Clinical Epidemiology Unit, Ottawa Health Research Institute and Director of the Centre for Best Practices, Institution of Population Health, University of Ottawa.

of the random assignment. This equivalence permits a fair comparison, i.e. an unbiased estimation of the relative effects of the intervention and a statistical statement of one's confidence in the results.

Trials in which individuals are randomly assigned to different interventions are familiar in medical and other research. Random allocation of units such as places and entities are less frequent. As Donald T. Campbell suggested in »Reforms as Experiments«:

Where policies are administered through individual client contacts, randomization at the person level may often be inconspicuously achieved....

But for most social reforms, larger administrative units will be involved, such as classrooms, schools, cities, counties or states. We

---

1 This paper is based on work supported by the U.S. Department of Education's Planning and Evaluation Service, The Rockefeller Foundation, and the Swedish Center for Evaluation of Social Services (CUS) in Stockholm

need to develop the political postures and ideologies that make randomization at this level possible. (Campbell, 1969; Campbell, 1988)

Campbell, did not consider deeply the use of places or entities in randomized trials because such trials, at the time, were rare. In what follows, we depend on Campbell's insight and build on others' more recent work. The topic is germane to evaluation of complex social programs that are designed to enhance health and well-being, welfare and, education, and to reduce crime and delinquency.

### 1.1 Definitions

The unit of allocation refers to who or what is randomly assigned to different interventions in a trial. Conventional textbooks in psychology and design of medical trials, for instance, typically handle experiments in which individuals are the units of allocation. Here, we focus on sites, administrative units or groups, rather than on individuals. We refer to »place-based randomized trials« in this paper. Such trials are also called »group randomized trials« (Murray, 1998) and »cluster randomized trials« (Donner and Klar, 2000).

The units of analysis are those for which data are available and used. Juvenile facilities may be the units of random allocation in a trial that compares two facility-wide approaches to reducing recidivism. The units of analysis may be the facilities or both facilities and individuals within facilities.

### 1.2 The Contents of this Paper

In what follows, we discuss assumptions about the use of randomized trials and their rationale. Further, we identify difficulties in their use. The examples in this paper are diverse, partly to demonstrate that useful trials can be carried out in a variety of settings.

### 1.3 Assumptions

The first assumption is that the government agencies and private foundations are interested in estimating the relative effect of new programs that they sponsor. Put another way, we assume that the public is interested in answering the question: »What works better, for whom, and for how long?«

A second assumption is that a defensible estimate of an innovation's effect depends on determining how sites or other entities would behave in the absence of an innovation. As a practical matter, one might, for example, develop such an estimate from time series forecasts. Kuusi's (1957) study on the effect of alcohol sales in Finland is a remarkable precedent in using administrative records in short time series. Here, we assume that time series data and ad hoc comparisons are insufficient to produce unbiased estimates of a program's effect. Some of these alternatives to randomized trials, including time series and their vulnerabilities, are covered by Campbell and Stanley (1963) and by Shadish, Cook, and Campbell (2002).

Most important, a simple, and scientifically defensible method of composing

a comparison group, one that permits fair estimates of the relative differences among programs, is the method of random assignment. For instance, a sample of juvenile facilities might be randomly selected from the pool of eligible facilities and engaged in a new intervention program. The outcomes at these facilities would then be compared to the eligible facilities that were randomly assigned to continue operating under the existing programs. The random assignment assures that the two groups of facilities do not differ systematically, apart from the influence of the intervention program under study.

A third assumption is the future of impact evaluation in the many countries lies with controlled trials that are mounted on a small scale so as to understand which programs work before such programs are mounted at the national or regional level. In fact, such experiments have been undertaken and their frequency has increased. Boruch and Foley (2000), for example, list over 50 different studies involving communities or geographic sectors, schools or classrooms, housing projects, and other kinds of organizations as the units of allocation in a randomized field trial. See Boruch (1997), Donner and Klar (2000), and Murray (1998) generally, and the Campbell Collaboration's Social, Psychological, Educational, and Criminological Trials Register (<http://www.campbellcollaboration.org>).

## 2. Rationale: Why Use Sites as the Units?

Why should we consider places or other

entities as the units of assignment to programs in evaluating the effect of a program? The reasons include: program theory; law and ethics; policy; the counsel of advisory groups; and statistical theory and rules of evidence.

### 2.1 Program Theory

By »theory« here, we mean how an intervention is supposed to have the effects that we believe they will have. In other words, the theoretician proposes a »logic model« to explain tentatively what happens when a program is implemented. Or, the theoretician may outline a formal path model or a causal chain.

Numerous theories of societal change posit that a program will work if it is delivered by organizational elements acting in concert. Research on preventing sexually transmitted diseases for example depends on theories about what institutional and group factors influence risky behavior. See Wasserheit, Aral, Holmes, and Hitchcock (1991) generally and Hornik (1991) in particular. Randomized field trials undertaken in California and Texas have employed 20 schools as the unit of allocation and analysis so as to test programs based on several such theories (Coyle et al., 1996; Basen-Engquist et al., 1997).

A variety of place-based randomized trials have also used schools as units to assess theory-driven programs that were designed to prevent or reduce substance abuse. The Midwestern Prevention Project (Pentz, 1994), for example, was based on a theory that adolescents' drug use depends on their characteristics, such as prior

drug use, and on the adolescents' ability to handle peer pressure toward using drugs. The theory also recognized that environmental and situational factors beyond the individual are important because community norms, for instance, can influence adolescent behavior.

Theory has also driven multi-stage research on how to engage and encourage mental hospital practices shown in earlier research to be more effective for treating certain forms of mental illness. Such theory involved ideas about the level at which the hospital staff might first be engaged (top down or bottom up) and the best mode of engagement. The latter included involvement in workshops or demonstration projects as opposed to merely sending brochures. The expectation was that people would react differently to these various engagement strategies (Fairweather et al., 1974; Fairweather and Tornatzky, 1977).

## 2.2 Law, Ethics, and Culture

One reason why sites might be used as the units of random assignment in a trial is that the random assignment of individuals to alternative programs within a site may not be legal or ethical. Or, this kind of randomization may not be acceptable on cultural or political grounds. Random allocation of entire sites to alternative programs might be regarded as both legally and ethically responsible.

For instance, in a randomized trial testing the Drug Abuse Resistance Education model (D.A.R.E.), researchers randomly assigned entire schools to treatment and

control groups partly because it would have been difficult to get cooperation from schools if some of their students received the program and some did not (Curtin, personal communication, April 3, 1996). A kind of institutional ethic or culture prevailed. Using schools as the units of random assignment helped insure the cooperation of control schools in the trials. Schools in the control group were promised access to the D.A.R.E. program the year after the completion of the study.

Similarly, each of the 80 or so juvenile facilities in Sweden, for example, may object to random allocation of their clients to different programs so as to discern which program is most effective in reducing recidivism. Other ethical values in the local facility may take precedence, e.g., giving the »same« service to everyone in the facility. A randomized trial in which eligible and willing facilities try out one of two different approaches may then be regarded as more just. This point was made by Karin Tengvald at Stockholm's meetings on evaluating social service programs (Soydan, 1998).

Again, the emphasis here is on comparing alternative interventions in different communities, not on giving one set of these groups a »treatment« and leaving the others high and dry. The focus, then, is not simply on whether a treatment works but on which treatment works better.

## 2.3 Policy and Politics

As a matter of policy and politics, the government agency or foundation that sponsors programs make rules that affect

organizations directly rather than individuals directly. Such rules require sites or organizations to take particular actions, create transactions, and so on. The implication is that a study of the effects of such a program has to recognize sites as the immediate targets in an evaluation design. The individuals within sites are the ultimate targets.

For example, federal policy on demonstration projects in the United States has emphasized, at times, that communities are essential in ameliorating certain social problems. Preventing substance abuse is a case in point. The Center for Substance Abuse Prevention (CSAP) was created to reduce the incidence of alcohol, tobacco, and drug use. It has tried to do so through efforts such as the Community Partnership Demonstration Program which has focused on learning how diverse community-based organizations can be engaged in effective intervention. Different ways to do so were described by Kaftarian and Hansen (1994). The emphasis was on communities as the units of allocation and analysis in randomized field trials (Pentz, 1994; Wagenaar et al., 1994; Ellickson, 1994; Murray and Wolfinger, 1994; Lorion, 1994).

Other examples of programs in which the most direct connection are between entities and government or foundation assistance rather than between individuals and such assistance are easy to identify. They appear in compensatory education and other programs sponsored by the U.S. Department of Education; and the U.S. Department of Health and Human Services. Loans made by the World Bank to governments are operationalized by organizations such as banks,

agricultural stations, or schools. The World Bank rarely supports randomized trials, but there are a few examples of programs sponsored by bank loans that have been tested in place-based trials.

## 2.4 Statistical Theory and Analysis

Contemporary statistical analysis methods rely on the assumption that an observation on any given individual or entity is independent of observations on all the others. When the assumption does not hold, and the analyst fails to recognize this, the analysis will be compromised. For instance, difference in program effectiveness may be declared statistically significant because the analysis is wrong in failing to recognize non-independence. See for instance Donner and Klar (2000) and Murray (1998).

Assuming that the units of observation are independent is not plausible in many settings. For example, a particular gang member's response to a juvenile crime reduction program may not be independent of other gang members' responses even though the program involves only some members. A child's grade on a test of ability to work in teams presumably will not be independent of grades given to other children on the same team.

For the statistician, all this implies that it is not individuals who ought to be randomly assigned to programs. And it is not individual level data that must ordinarily be used to estimate the program's effect. Rather, allocation and analysis should focus first on entire groups or organizations and second on individuals within each group or entity.

## 2.5 The Counsel of Advisory Groups on Research and Evaluation Policy

At times, preventing dangerous diseases, including sexually transmitted ones, requires that the programs be deployed through organizations or geopolitical jurisdictions. As a consequence, the National Academy of Sciences Panel on Evaluating AIDS Prevention Programs suggested that diagnostic testing and counseling sites be considered as the units in controlled experiments to improve the services (Coyle, Boruch, and Turner, 1991). Multidisciplinary conferences on sexually transmitted diseases (STDs), sponsored by the National Institute on Allergy and Infectious Diseases (NIAID), have led to the observation that clinical practices, factories, churches, and other organizations, as well as communities, might properly serve as the units in randomized trials (Green and Washington, 1991).

In considering approaches to preventing abuse of controlled substances, the participants in the »Communities that Care« Evaluation Design Conference said:

rigorous evaluation of a comprehensive community intervention requires an experimental design whereby communities are randomly assigned to experimental and control conditions.

See Peterson, Hawkins, and Catalano (1992). England's Joseph Rowntree Foundation has been influenced by similar concerns (Farrington, 1997).

The National Research Council's Panel on the Understanding and Control of Vio-

lent Behavior offered the following:

Recommendation 4: The panel calls for a new multi-community program of developmental studies of aggressive, violent, and antisocial behaviors, intended to improve both causal understanding and preventive interventions... (p. 25).

Edited by Reiss and Roth (1993), this Panel's report argued that »Randomized controlled field experiments usually have important advantages as an evaluation strategy« (p. 320).

Finally, consider that »Design and Analysis Issues in Community Trials« was the primary topic on the agenda of a 1992 National Institutes of Health conference. The participants agreed that the use of the communities as the units of allocation and analysis presented challenges, but that there were a variety of techniques for overcoming these challenges (Murray et al., 1994).

## 3. Examples

People often do not realize that it is possible to execute randomized trials that use organizations or other entities as the units of random allocation in trials that permit fair comparisons. In what follows, we give evidence on the feasibility of such trials

### 3.1 Schools, School Districts, and Classrooms as the Units of Random Assignment

Schools and classrooms, for instance,

have been randomly assigned to different approaches in educating children about avoiding substance abuse (Schaps et al., 1982; Moskowitz, 1984; Botvin et al., 1995; Murray, Moskowitz, and Dent, 1996). In tests of the Drug Abuse Resistance Education (D.A.R.E.) program in Illinois, for example members of 12 pairs of schools were randomly assigned to different programs in the interest of fair comparison (Rosenbaum et al., 1991). Other entity-based experiments on this program were reviewed by Ennett et al. (1994). The Flay et al (1985) work in Canada is a remarkable precedent in this arena.

In efforts to evaluate a theory-driven program to reduce alcohol use by underage youth, Wagenaar et al. (1994) mounted a randomized field trial involving 15 school districts.

Seven of the willing districts in Minnesota and Wisconsin were randomly assigned to employ a special community-based prevention program. Eight of the willing districts were randomly assigned to the control group.

Schools have also been the units in at least two smoking prevention experiments. The Television, School and Family Smoking Prevention Project, used multi-attribute balancing to randomly assign 35 Los Angeles area schools to different media-based smoking prevention campaigns. Flay et al. (1985) randomly assigned 22 matched schools to experimental and control conditions in the Waterloo Study, a Canadian smoking prevention effort. Tests of school-wide cardiovascular risk reduction programs for children have been undertaken. For example, schools have been randomly

assigned to such programs in four states (Killen et al., 1988; Hansen and Graham, 1991; and Perry et al., 1992).

In a mobile societies, it is important to understand how to reduce the psychological and educational risk of children who are moved from one education context to another. Jason et al. (1992, 1993a, 1993b) focused on children who transferred into new schools and who were, as a consequence, vulnerable. One project involved randomly assigning members of ten matched pairs of schools to an innovative treatment program or to a control condition in order to determine whether their special transition program worked.

Until the late 1990s, high quality evaluations of violence reduction programs in schools were rare. Among the notable exceptions is the Grossman et al. (1997) study of the effectiveness of violence prevention curricula for second and third graders. Six matched pairs of schools were randomly assigned to employ the curriculum or to serve in a control group. Differences in children's behavior were discernible and persisted for at least six months.

Until the 1970s, no controlled-field experiments of any scale appear to have been run to understand the effects of standardized testing on students in any country. In 1975, the Irish Republic decided to consider for the first time standardized testing for children in the Republic's elementary schools. Kellaghan, Madaus, and Airasian (1982) and their colleagues at St. Patrick's College (Dublin) mounted a study in which 175 eligible schools, matched and stratified, were allocated randomly to different conditions. The control condition involved

no standardized testing. The intervention was standardized testing, with and without feedback to teachers, on student performance.

Randomized trials have been mounted to understand what kinds of programs might be deployed in education settings as to enhance children's understanding of high risk sexual behavior and how to avoid it. In the U.S. for example, Gay's (1996) dissertation research involved matching eight middle school classrooms and allocating half to a new Red Cross program and half to a control condition in which no such education effort existed. In the Philippines, Alpasca et al. (1995) also targeted classrooms within schools. In a large-scale trial in California, Kirby et al. (1997a) randomly assigned 102 classrooms in six middle schools to a theory-driven risk prevention program that relied heavily on young »peer education« to implement the program. Another California based program, Postponing Sexual Involvement (PSI) was evaluated using a complex research design in which classrooms were randomized in one component (Kirby et al., 1997b). Over 50 schools were involved.

A different stream of health related work has concerned nutrition education. Woodruff (1997), for instance, described a San Diego experiment that involved eight intervention classes and nine control classes being randomly assigned to a new nutrition program from three community colleges.

Earlier examples to test different approaches in different countries to enhancing children's achievement deserve recognition. Consider examples from Nicaragua, El Salvador, and the U.S. Classrooms

in Nicaragua have been randomly assigned to radio-based mathematics education and to conventional education so as to learn whether the former would enhance mathematics achievement and reduce education costs relative to the latter (Dean et al., 1981; Jamison, Searle, and Suppes, 1980). A similarly designed randomized trial in El Salvador disintegrated; Hornik et al. (1972) gave an admirably candid description. During the 1970s, the U.S. Department of Education sponsored a large scale study to understand whether funding could be effectively employed by schools to reduce racial isolation and enhance the achievement of students. Eligible schools that were willing to participate in the experiment were randomly allocated to a special funding opportunity and to a control group that received no special treatment. See Coulson (1978), Reichardt and Rindskopf (1978), and Weissberg (1978).

### 3.2 Communities and Geopolitical Entities as the Units of Random Assignment

In a study of how to encourage voter registration in Chicago, Gosnell (1927) appears to have randomly assigned distinct neighborhoods in political precincts to treatment and control conditions. The »treatment« involved publicity, mail, and in-person contacts, provided at times in different languages to diverse ethnic neighborhoods. The intent was to provide information about voter registration and to encourage registration in different ways, and to test the treatment.

Communities have been the units of



allocation in evaluations of health related programs. LaPrelle, Bauman, and Koch (1992), for instance, reported on a study of the relative effectiveness of three media campaigns to prevent cigarette smoking among adolescents. They screened, matched and then randomly assigned communities from a sample of ten communities to one of three treatments and to a control group. The Community Intervention Trial for Smoking Cessation (COMMIT), assigned eleven matched pairs of communities to its treatment and comparison groups (Freedman, Green, and Byar, 1990 cited in Peterson et al., 1992).

In randomized trials on fertility interventions in the Far East, communities and villages have been randomly assigned to different approaches to understand how to decrease birth rates (Freedman & Takashita, 1969; Riecken et al., 1974). Small numbers of communities have also been used as units in randomized studies of HIV risk prevention tactics (Kelly et al., 1991). In media-based smoking prevention campaigns, standard metropolitan statistical areas (SMSAs) have been allocated randomly to the campaigns or to control conditions (Bauman et al., 1991). Federal statistical agencies specify these SMSA geographic areas in a uniform way so as to make clear what is meant by »metropolitan area« in contrast to a rural area, for example, and use these areas to design the census and national surveys. Education studies in Cali, Colombia involved randomly assigning very small geographic areas in the low-income barrios to a cultural enrichment and health enhancement

program for preschoolers to determine its effect relative to randomly assigned control areas (McKay et al., 1978).

Some randomized trials have been mounted because the integrating multiple services at the community level are thought to be important to people who are mentally ill and live in the community. Access to Community Care and Effective Service Supports (ACCESS) involved eight cities, each of which contained two independent jurisdictions that were randomly assigned to the ACCESS or to the control condition (Randolph et al., 1997). About 50 agencies within each jurisdiction cooperated on the study.

Finally, consider early research on crime prevention. In the Kansas City patrol experiment, fifteen police beats were matched and randomly divided into three groups of five beats each. This precedent compared the relative effects of reactive, proactive, and control (normal) patrols on victimization (Kelling, Pate, Dieckman, and Brown, 1974). Twenty years later, Sherman and Weisburd (1995) executed a better-randomized trial in Minneapolis. The researchers identified over 100 »hot spots«, local areas of predictably high crime and randomly allocated half of these areas to more intensive police patrol or to a normal patrol activity.

### 3.3 Other Private and Public Organizations as Units of Random Assignment

In some countries, a sensible way to enhance the well-being of individuals is through private organizations. Programs

designed to reduce the risk of sexually transmitted diseases, for example, might be more effective if the program is directed toward all the workers in corporate factories rather than to individuals who may or may not work in the factories. It is partly for this reason that the National Institute of Allergies and Infectious Diseases in the U.S. has invested in tests of factory-based peer education (NIAID, 1997). No one knows whether peer education among factory workers will reduce infection. The project involved some 40 factories in Zimbabwe, half being randomly assigned to programs designed to reduce incident HIV infection and the remaining to a control condition. Other randomized trials have used work sites as units in assessing nutrition programs and weight control and smoking cessation programs (Simpson et al., 1995).

Non-profit service organizations have, at times, committed resources to randomized trials. For instance, Good Will Industries in the U.S. agreed to participate in controlled experiments on how to improve the management of the organization's stores (Glaser et al., 1967). In this instance, independent stores were the units of allocation.

In the medical arena, nearly forty Minnesota community hospitals agreed to participate in a trial to discover whether local medical opinion leaders and a formal feedback system could influence the rate at which the hospitals adopted new beneficial therapies for acute myocardial infarction patients (Soumerai et al., 1998). The theory underlying the program is that the entire hospital staffs' understanding, not just the physician's education, together

with the monitoring of therapy, are necessary to produce change. Hence, allocating hospital physicians randomly to a program was not sensible. The trial's design involved the random allocation of 20 hospitals to this approach to clinical education and random allocation of 17 hospitals to a control condition.

Our final illustration involves a program designed to enhance employment of individuals at high risk of unemployment who live in low-income public housing developments in communities that need economic revitalization. In each of seven cities, the trials involved the random allocation of one public housing facilities to the program and one or two public housing facilities to a control condition. The presumptions underlying the program's design were that local collaboration and collective decisions are essential in transforming local communities in ways that affect, among other things, education, training and employment, and wage rates (Riccio, 1998; Bloom, Bos, and Lee, 1998).

## 4. Difficulties and Possible Resolutions

Challenges to using places or other entities as the units of allocation in a randomized trial are numerous. Strategies that have been invented to surmount obstacles are valuable and discussed in what follows.

### 4.1 Statistical Power

Consider a randomized field trial in which two literacy programs are compared to one

another to establish which is more effective and less costly. Statistical power refers to our ability to discern the relative effectiveness of the two literacy programs. This power depends, of course, on how literacy is measured. It also depends on how many literacy centers are randomly allocated to one or the other literacy program and on how many students there are in each program. The »statistical power« refers to our ability to detect a difference in the effects of the interventions if indeed there is a difference.

How many centers might be required in this experiment to assure that its statistical power is about .80? Assume, as is likely, that the true difference between the programs is small (.10) and fix the statistical threshold (alpha) at .05. If all the students within schools were independent, about 400 students for each plan would have to be sampled to discern the effect of the treatments under these conditions.

When the similarity among students within a school is substantial, a larger sample size will be necessary to assure that real differences between the intervention is detected. Assuming a low similarity rate (intra-class correlation) of .05, one might then use 85 schools with a sample of 10 students each, for each treatment (program) in a formal test. Or, one may use 44 schools with 40 students each.

In the opinion of LaPrelle et al. (1992), their trial on community-based substance use prevention in citywide programs was underpowered. Four treatments in an experiment were spread over ten communities. Their thoughtful post-trial analysis suggested that about 40 communities per

group would have been required to detect an important difference in the effectiveness of smoking prevention programs.

Place-based randomized trials have relied successfully on at least three tactics to assure adequate statistical power. First, entities that are independent should be screened for eligibility and a reasonable level of homogeneity. Second, the entities should be matched and then randomized. A third tactic is implicit: engage as many entities as possible in the trial.

## 4.2 Measurement Systems and Theory

By a theory of »what should happen,« we mean laying out the way that the programs being compared are each expected to engage and affect the entities. That is, the logic of how the thing is supposed to work needs to be made plain. More to the point, the theory guides us in selecting what should be measured and, at its most sophisticated, whether and how well it might be measured.

Consider the multi-site Wagenaar et al. (1997) trial. It was designed to understand whether a community-based program could reduce the use of alcohol by underage youth. Mobilization of communities was regarded as theoretically important to creating alcohol use policy. Observations then were made of community power structures and the attitudes of students and youth. Analyses were undertaken of media coverage. Changes in community practice were also measured on the supposition that these would follow community mobilization. Among other efforts, this

stage included surveys of retail alcohol outlets to determine if indeed they failed to ask proof of the age of customers whose appearances were youthful. This was done because, in theory, decreasing youth access to alcohol would result in fewer alcohol-related traffic accidents. Further, the latter were assessed using state and local record systems.

### 4.3 Engaging Sites and other Entities

Engaging sites, administrative units, and other entities in a randomized field trial requires considerable skill. Walker et al (2000) provide an exceptionally detailed description of strategies for recruiting U. K. Hospitals into randomized trials. They focus attention on identifying stakeholders and gatekeepers, informing them, approaching gatekeepers to engage the hospital, negotiating the terms of engagement, conducting the study, and providing feedback of different kinds to gatekeepers and stakeholders. The process is time consuming an challenging. To judge from researchers success in mounting such trials. The strategies are worth serious consideration.

Consider next, Ellickson's (1994) paper on the conduct of Project ALERT, which involved 30 schools being randomly assigned to ALERT or to a control condition. Its object was to determine how well the ALERT project worked to prevent substance abuse among children and how long the project's effects last. Recruiting entire schools into a RFT must recognize natural limits on their capacity to participate. Ellickson (1994) reported that eleven schools out of about 60 schools that were

invited to participate declined to do. One school, for instance, could not participate on account of a court order demanding considerable resource allocation on racial equity. Four of the eleven schools declined to participate because they already had prevention programs in place. The reasons for other declinations concerned their capacity, e.g., inability to assure community support for engaging in the experiment.

### 4.4 Temporal and Structural Stability

We expect sites not to change much over a short period of time. Nonetheless, the stability of certain characteristics of sites may be low or trends may reverse direction. Bauman et al. (1991), for example, found high positive correlation over a two-year period ( $r = .77$ ) for adolescents' reported rates of recent smoking in a sample of 10 cities. The researchers found a negative correlation ( $r = -.31$ ) for adolescents' rates of experimentation with smoking in the same cities. Reasons for this finding are unclear. The instability is clear.

One normally assumes that the places or other entities that are targeted for a program will be structurally stable over the study's course. A school in year 1, for instance, is expected to be a school in year 2. To judge from experience, it is prudent to anticipate some change. For example, the Midwestern Prevention Project involved randomly assigning schools to different conditions. Pentz (1994) reported that 8 of the initial 50 targeted middle schools and high schools »closed or consolidated with

other schools over the first three years of the study» (p. 44). Further, feeder schools changed as a consequence of changes in busing patterns and the creation of magnet schools that drew students from areas outside the original catchment area schools.

Similar problems have occurred elsewhere. In the Irish Standardized Testing experiment, after matching and randomly assigning schools based on census data, the researchers found that many important school characteristics had changed (Kellaghan, Madaus, and Airasian, 1982). Tennessee's experiment on school incentives encountered difficulties because schools were closed or consolidated with other schools (Bickman, 1985). All this engenders complex problems in designing randomized trials and in their analysis.

#### 4.5 Regional Variation

To produce a good estimate of the effect of smoking prevention programs, Bauman et al. (1991) focused attention on only one geographic region. Despite this attempt to work in a homogeneous context, the experiment was underpowered. That is, the sample of organizations within the region may have been too small to discern a real effect of programs because there was considerable variation within the region. For instance, the rates of recent smoking among adolescents across ten cities in one region reported by Bauman et al. (1991) were in the range 2-7% in 1985 and 13-20% in 1987. Rates of smoking in 1987 among 1985 nonsmokers were in the range of 3-14% across the cities.

Stratification or blocking by region in

a place-based trial makes sense. But the definitions of region and the implications of a choice have not been investigated deeply. In any event, reconnaissance prior to mounting a randomized experiment—pilot tests and analyses of extant data—are warranted.

#### 4.6 Unbalanced Groups and Restricted Randomization

Consider a randomized trial in which a sample of communities that is provided with increased literacy resources is compared to a sample of communities that has been allocated to a waiting list, i.e., have not yet been given the resources. The number of communities involved in such a study must often be relatively small, say 20 to 40, in each of the groups. For the analyst, this raises a concern that the two groups that are randomly composed will not be »equivalent« at the outset. That is, there is an imbalance between the groups that is attributable to chance. This »unhappy random configuration« will complicate comparisons. One approach used to reduce the problem in multi-site RFTs is restricted randomization.

In restricted randomization, some configurations of the random allocation of sites to different treatments are defined as undesirable a priori. That is, all possible randomized configurations under a particular experiment's design are laid out beforehand. The »unhappy« ones are then eliminated from eligibility. A random selection is then made from the remaining eligible configurations. For the applied researcher, constraining the randomization

options to sensible configurations prevents badly unbalanced groups of institutions from being assigned to different program variations. For instance, Ellickson and Bell (1992) linked »unlike schools from districts into pairs and randomly (assigned) the pairs to the experimental conditions...« to achieve balance (p. 85).

The implication is that when a small number of sites are the units of allocation in randomized trials, we can enumerate all possible allocations of sites in advance of the trial. Further, we can eliminate the possible allocations that are strange, out-of-line, and so on. Having eliminated the allocations that are out-of-line, we can randomly select a configuration, allocate institutions in accord with it, and develop a comparison of programs that is fair.

#### 4.7 Implementation Fidelity and Measurement

It makes no sense to estimate the effect of a new program unless one can verify that the program activities occur and can be described. »Implementation fidelity« here refers to the degree to which a new program

is actually delivered to target individuals. Its measurement refers to observing indicators of fidelity. We need to determine whether administrative actions are taken, information systems are emplaced, and so on. Learning that actions are indeed taken is a prerequisite for any impact evaluation.

Trials that attempt to evaluate interventions that involve »integration« or »coordination« of services across many agencies within an organization or community present special problems. Developing a coherent definition of integration and measurable indicators of integration is not easy. Consider studies of ACCESS' effect on the homeless and mentally ill, for instance. The various jurisdictional units may differ on: whether and how they employ interagency coalitions; interagency teams for service delivery; interagency management systems; interagency agreements and memorandums of understanding; finding arrangements; eligibility standards; and co-location of services (Randolph et al., 1997). Learning how to observe any of these reliably and to assure fidelity in implementation and its measurement is demanding.

### Bibliography<sup>2</sup>

Aplasca, M., Siegel, D., Mandel, J. S., Santana, R., Paul, J., Hudes, E. S., Monzon, O. T, and Hearst, N. (1995) Results of a Model AIDS Prevention

Program for High School Students in the Philippines. AIDS, Supplement 1, 7-13. (\*)

Basen-Engquist, K., Parcel, G. S., Harrist, R., Kirby, D., Coyle, K., Banspach, S., and Rugg, D. (1997) The Safer Choices Project: Methodological Issues in School Based Health Promotion Intervention Research. Journal of School Health, 67(9), 365-371. (\*)

Bauman, K. E., LaPrelle, J., Brown, J. D., Koch, G.

<sup>2</sup> The references in the bibliography that are marked with an asterisk (\*) report on trials that involve places, organizations, and groups or other entities as the units of random allocation in randomized trials.

- C. and Padgett, C. A. (1991) The Influence of Three Mass Media Campaigns on Variables Related to Adolescent Cigarette Smoking: Results of a Field Experiment. *American Journal of Public Health*, 1991, 81, 597-604. (\*)
- Bickman, L. (1985) Randomized Field Experiments in Education. *New Directions for Program Evaluation*, 28, pp. 39-54. (\*)
- Bloom, H., Bos, J. and Lee, S. W., (1998) Using Cluster Random Assignment to Measure Program Impacts: Statistical Implications for the Evaluation of Education Programs. New York: New York University, Robert F. Wagner School of Public Service (Research Report). (\*)
- Boruch, R. F. (1993(a)) Multi-site Tests in the Civil and Criminal Justice Arena. Invited Presentation, Annual Meeting of the American Society of Criminology (October 30, 1993) Phoenix, Arizona. Available from: Author. University of Pennsylvania, Philadelphia, PA 19104. (\*)
- Boruch, R. F. (1993(b)) Multi-site Evaluation and the Children's Initiative. Paper prepared for the Pew Charitable Trusts, Philadelphia, PA. Available from: Author. University of Pennsylvania, Philadelphia, PA 19104. (\*)
- Boruch, R. F. (1997) Randomized Experiments for Planning and Evaluation: A Practical Guide. Thousand Oaks, CA: Sage.
- Boruch, R. F. and Foley, E. (2000) The Honestly Experimental Society: Sites and Other Entities as the Units of Allocation and Analysis in Randomized Trials. In L. Bickman (Ed.) *Validity and Experimentation: Donald Campbell's Legacy Volume 1*. Thousand Oaks, CA, London, New Delhi: Sage Publications.
- Botvin, G. J., Baker, E., Dusenburg, L., Botvin, E. M., and Diaz, T. (1995) Long Term Follow-up Results of a Randomized Drug-Abuse Prevention Trial in a white Middle Class Population. *Journal of the American Medical Association*, 273, 1106-1112. (\*)
- Campbell, D. T. (1969) Reforms as Experiments. *American Psychologist*, 24(4), 408-429. (\*)
- Campbell, D. T. (1988) The Experimenting Society. Chapter 11 in S. Overman (Ed.) *Methodology and Epistemology for Social Science: Selected Papers by Donald T. Campbell*. Chicago: University of Chicago Press, pp. 290-314.
- Campbell, D. T. and Stanley, J. C. (1963) Experimental and Quasi-experimental Designs for Research Teaching. In N. L. Gage (Ed) *Handbook of Research on Teaching*. Chicago, IL: Rand McNally, pp 171-246.
- Coulson, J. E. (1978) National Evaluation of the Emergency School Aid Act (ESAA): A Review of Methodological Issues. *Journal of Educational Statistics*, 3(3), 1-60. (\*)
- Coyle, S. L., Boruch, R. F., and Turner, C. F. (Eds.) (1991) *Evaluating AIDS Prevention Programs (Expanded Edition)*. Washington, DC: National Academy of Sciences. (\*)
- Coyle, K., Kirby, D., Purcel, G., Basen-Engquist, K., Banspach, S, Rugg, D., and Well, M. (1996) Safer Choices: A Multicomponent School Based HIV/STD and Pregnancy Prevention Program for Adolescents. *Journal of School Health*, 66(3), 89-84. (\*)
- Dean, J., Seare, B., Galda, K., and Heyneman S. P. (1981) Improving Elementary Mathematics Education in Nicaragua: An Experimental study of the Impact of Textbooks and Radio on Achievement. *Journal of Education Psychology*, 73(4), 556-567. (\*)
- Donner, A. and Klar, N. (2000) *Design and Analysis of Cluster Randomized Trials in Health Research*. New York: Oxford University Press.
- Ellickson, P. L. (1994) Getting and Keeping Schools and Kids for Evaluation Studies. *Journal of Community Psychology (Monograph Series: CSAP Special Issue)*, pp. 102-116. (\*)
- Ellickson, P. L. & Bell, R. M. (1992) Challenges to Social Experiments: A Drug Prevention Example. *Journal of Research in Crime and Delinquency*, 29(1), pp. 79-101. (\*)
- Ellickson, P. L. & Bell, R. M. (1990) Drug Prevention in Junior High: A Multi-site Longitudinal Test. *Science*, 247, pp. 1299-1305. (\*)
- Ennett, S. T., Tobler, N. S., Ringwalt, C. L., and Flewelling, R. L. (1994) How effective is Drug Abuse Resistance Education? A Meta-analysis of Project DARE's Outcome Evaluations.

Boruch, Foley & Grimshaw: Estimating the Effects of Interventions...

- American Journal of Public Health, 84(9), 1394-1401. (\*)
- Fairweather, G. W., Sanders, D. H., & Tornatsky, L. G. (1974) *Creating Change in Mental Health Organizations*. New York: Pergamon. (\*)
- Fairweather, G. W. and Tornatzky, L. G. (1977) *Experimental Methods for Social Policy Research*. New York: Pergamon Press. (\*)
- Farrington, D. P. (1997) *Evaluating a Community Crime Prevention Program*. *Evaluation*, 3. (\*)
- Flay, B. R., Ryan, K. B., Best, J. A., Brown, K. S., Kersell, M. W., d'Avernas, J. R. & Zanna, M. P. (1985) Are Social-psychological Smoking Prevention Programs Effective? The Waterloo Study. *Journal of Behavioral Medicine*, 8(1), pp. 37-59. (\*)
- Freedman, R. & Takashita, J. T. (1969) *Family Planning in Taiwan: An Experiment in Social Change*. Princeton, NJ: Princeton University Press. (\*)
- Gay, K. E. M. (1996) *Collaborative School-based Research: The Creation and Implementation of an HIV/AIDS Prevention Curriculum for Middle School Students*. PhD Dissertation, University of Pennsylvania, Philadelphia, PA. (\*)
- Glaser, E. M., Coffey, H. A., and others (1967) *Utilization of Applicable Research and Demonstration Results*. Los Angeles, CA: Human Interaction Research Institute. (\*)
- Gosnell, H. F. (1927) *Getting Out the Vote: An Experiment in the Stimulation of Voting*. Chicago: University of Chicago Press. (\*)
- Green, S. B. and Washington, A. E. (1991) *Evaluation of Behavioral Interventions for Prevention and Control of Sexually Transmitted Diseases*. In: J. N. Wasserheit, S. O., Aral, K. K., Holmes, and P. J. Hitchcock (Eds.) *Research Issues in Human Behavior and Sexually Transmitted Diseases in the AIDS Era*. Washington, D.C.: American Society for Microbiology, pp. 345-352.
- Grossman, D. C., Neckerman, H. J., Koepsall, T. D., Liu, P., Asher, K. N., Beland, K., Frey, K., and Rivara, F. P. (1997) Effectiveness of a Violence Prevention Curriculum among Children in Elementary School: A Randomized Controlled Trial. *Journal of the American Medical Association*, 277(20), 1605-1611, (\*)
- Hansen, W. B. & Graham, J. W. (1991) Preventing alcohol, marijuana and cigarette use among adolescents; Peer pressure resistance training versus establishing conservative norms. *Preventive Medicine*, 20, 414-430.
- Hornik, R. (1991) *Alternative Models of Behavior Change*. In J. N. Wasserheit, S. O., Aral, K. K. Holmes, and P. J. Hitchcock (Eds.) *Research Issues in Human Behavior and Sexually Transmitted Diseases in the AIDS Era*. Washington, D.C.: American Society for Microbiology, pp. 201-218.
- Hornik, R. C., Ingle, H.T., Mayo, J. K., McAnany, E. G., and Schramm, W. (1972) *Television and Education Reform in El Salvador*. (Report No. 14) Stanford University, Institute for Communication Research. (\*)
- Jamison, D., Searle, B., & Suppes, P. (1980) *Radio Mathematics in Nicaragua*. Stanford, CA: Stanford University Press. (\*)
- Jason, L. A., Weine, A. M., Johnson, J. H., Donner, K. E., Kuraski, K. S., & Sohlberg, L. (1993a). The school transitions project: A comprehensive preventive intervention. *Journal of Emotional and Behavioral Disorders*, 1(1), pp. 65-70. (\*)
- Jason, L. A., Weine, A. M., Johnson, J. H., Sohlberg, Filippelli, Turner, E., & Lardon, C. (1992) *Helping Transfer Students: Strategies for Educational and Social Readjustment*. San Francisco: Jossey-Bass. (\*)
- Jason, L., Johnson, J. H., Danner, K. E., Taylor, S., and Krasaki, K. S. (1993b) *A Comprehensive, Preventive, Parent-Based Intervention for High Risk Transfer Students*. *Prevention in Human Services*, 10(2), 27-37. (\*)
- Kaftarian, S. J. & Hansen, W. B. (1994) (Eds.) *Community Partnership Program: Center for Substance Abuse Prevention*. CSAP Special Issue/Monograph Series. *Journal of Community Psychology*. (\*)
- Kellaghan, T., Madaus, G. F., Airasian, P. W. (1982) *The Effects of Standardized Testing*. Boston/The Hague/London: Kluwer-Nijhoff. (\*)



- Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. E. (1974) The Kansas City Preventive Patrol Experiment: A Summary Report. Washington, DC: Police Foundation. (\*)
- Kelly, J.A., Lawrence, J.S., Diaz, Y. E. and others. (1991) HIV Risk Behavior reduction Following Intervention with Key Opinion Leaders: An Experimental Analysis. *American Journal of Public Health*, 81, 168-171. (\*)
- Killen, J.D., Telch, M.J., Robinson, T.N., Maccoby, N., Taylor, C., & Farquar, J. W. (1988) Cardiovascular Disease Risk Reduction for Tenth Graders: A Multiple Factor School-based Approach. *Journal of the American Medical Association*, 260(12), pp. 1728-1733. (\*)
- Kirby, D., Korpi, M., Adivi, C. and Weismann, J. (1997a) An Impact Evaluation of Project SNAPP: An AIDs Prevention and Pregnancy Middle School Program. *AIDS Education and Prevention*, 9(Supplement A), 44-61. (\*)
- Kirby, D., Korpi, M., Barth, R. P., and Cagampang, H. H. (1997b) The Impact of Postponing Sexual Involvement Curriculum among Youths in California. *Family Planning Perspectives*, 29, 100-108. (\*)
- Kuusi, Pekka (1957) (WestPhaler, A. Translator). *Alcohol Sales in Rural Finland*. Volume 3 Publication of the Finish Foundation for Alcohol Studies. Stockholm, Sweden: Almqvist and Wiksell.
- LaPrelle, J., Bauman, K. E. & Koch, G. G. (1992) High intercommunity variation in adolescent cigarette smoking in a 10-community field experiment. *Evaluation Review*, 16(2), pp. 115-130. (\*)
- Leviton, L., Valdiserri, R., Lyter, D., Callahan, C., Kingsley, L., Huggins, J., and Rinalde, C. R. (1990) Preventing HIV Infection in Gay and Bisexual Men: Experimental Evaluation of Attitudes Changes from Two Risk Reduction Experiments. *AIDS Education and Prevention*, 2(2), 95-108. (\*)
- Lorian, R. P. (1994) Epilogue: Evaluating the Community Partnership Program. Reflections on a Name. *Journal of Community Psychology* (Monograph Series: CSAP Special Issues), pp. 201-205. (\*)
- McKay, H., McKay, A., Sinnestera, L., Gomez, H., and Lloreda, P. (1978) Improving Cognitive Ability in Chronically Deprived Children. *Science*, 200(4), 270-278. (\*)
- Moskowitz, J. et al. (1984) The Effects of Drug Education and Follow-up. *Journal of Alcohol and Drug Education*, 3, pp. 45-49. (\*)
- Murray, D. (1998) *Design and Analysis of Group Randomized Trials*. Oxford and New York: Oxford University Press.
- Murray, D. M., McKinlay, S. M., Martin, D., Donner, A. P., Dwyer, J. H., Raudenbush, S. W., & Graubard, B. I. (1994). Design and Analysis Issues in Community Trials. *Evaluation Review*, 18(4), pp. 493-514. (\*)
- Murray, D. M. and Wolfinger, R. D. (1994) Analysis Issues in the Evaluation of Community Trials: Progress Toward Solutions in SAS/STAT Mixed. *Journal of Community Psychology* (Monograph Series: CSAP Special Issue), pp. 140-154. (\*)
- Murray, D., Moskowitz, J. M., and Dent, C. W. (1996) Design and Analysis Issues in Community-Based Drug Abuse Prevention. *American Behavioral Scientist*, 39(7), 853-867. (\*)
- Pentz, M.A. (1994) Adaptive Evaluation Strategies for Estimating the Effects of Community Based Drug Abuse Prevention Programs. *Journal of Community Psychology* (Monograph Series CSAP Special Issue), pp. 5-25. (\*)
- Perry, C., Parcel, G. S., Stone, E., Nader, P., McKinlay, S. M., Leupker, R. V., and Webber, L. S. (1992) The Child and Adolescent Trial for Cardiovascular Health (CATCH): An Overview of Intervention Program and Evaluation Methods. *Cardiovascular Risk Factors*, 2(1), pp. 36-43. (\*)
- Peterson, P. L., Hawkins, J. D., & Catalano, R. F. (1992) Evaluating Comprehensive Community Drug Risk Reduction Interventions. *Evaluation Review*, 16(6), pp. 579-602. (\*)
- Randolph, F., Basinsky, M., Leginski, W., Parker, L., and Goldman, H. H. (1997) Creating Integrated Service Systems for Homeless Persons with Mental Illness: The Access Program. *Psychiatric Services*, 48(3), 369-373. (\*)

- Reichardt, C. S. & Rindskopf, D. (1978) Randomization and Educational Evaluation: The ESAA Evaluation. *Journal of Educational Statistics*, 3(1), 61-68. (\*)
- Reiss, A. J. & Roth, J. A. (Eds.) (1993) *Understanding and Preventing Violence*. Washington, DC: National Academy of Sciences Press.
- Riccio, J. A. (1998) *A Research Framework for Evaluating Jobs-Plus, A Saturation and Place-Based Employment Initiative for Public Housing Residents* (Working Paper). New York, Manpower Demonstration Research Corporation. (\*)
- Riecken, H. W., Boruch, R. F., Campbell, D. T., Caplan, N., Glennan, T. C., Pratt, J. W., Rees, A., & Williams, W. (1974) *Social Experimentation: A Method for Planning and Evaluating Social Programs*. New York: Academic Press. (\*)
- Rosenbaum, D. P., Ringwalt, C., Curtin, T. R., Wilkinson, D., Davis, B., & Taranowski, C. (1991) *Second Year Evaluation of D.A.R.E. in Illinois*. (Available from: D. P. Rosenbaum Center for Research in Law and Justice, University of Illinois at Chicago, Chicago, Illinois 60607). (\*)
- Schaps, E., Moskowitz, J., Condon, J., & Malvin, J. (1982) A Process and Outcome Evaluation of a Drug Education Course. *Journal of Drug Education*, 12, pp. 245-454. (\*)
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002) *Experimental and Quasi-experimental Designs for generalized Causal Inference*. New York: Houghton Mifflin.
- Sherman, L. and Weisburd, D. (1995) General Deterrent Effects of Police Patrol in Crime »Hot Spots«: A Randomized Controlled Trial. *Justice Quarterly*, 12(4), 625-648. (\*)
- Simpson, J. M., Klar, N. and Donner, A. (1995) Accounting for Cluster Randomization: A Review of Primary Prevention Trials, 1990 through 1993. *American Journal of Public Health*, 85(10), 1378-1383. (\*)
- Soumerai, S. B., McLaughlin, T. J., Gurwitz, J. H., Guadagnoli, E., Hauptman, P. J., Borbas, C., Morris, N., McLaughlin, B., Gao, X., Willison, D. J., Asinger, R. and Gobel, F. (1998) Effect of Local Medical Opinion Leaders on Quality of Care for Acute Myocardial Infarction. *Journal of the American Medical Association*, 279(17), 1358-1363. (\*)
- Soydan, H. (1998) (Issue Editor) *Evaluation Research and Social Work*. *Scandinavian Journal of Social Welfare*, 7 (2).
- Wagenaar A. C., Murray, D. M., Wolfson, M., Forster, J. L., & Finnegan, J. R. (1994) Communities mobilizing for Change on Alcohol: Design of a Randomized Community Trial. *Journal of Community Psychology* (Monograph Series/CSAP Special Issue), pp. 79-101. (\*)
- Wagenaar, A. C., Murray, D. M., Gehan, J. P., Wolfson, M., Forster, J. L., Toomey, T. L., Perry, C. L., and Jones-Webb, R. (1997) *Communities Mobilizing for Change on Alcohol (CMCA): Outcomes from a Randomized Trial*. Report. University of Minnesota. Submitted. (\*)
- Walker, A. E., Campbell, M. K., Grimshaw, J. M., and the TEMPEST Group (2000) A Recruitment Strategy for Cluster Randomized Trials in Secondary Care settings. *Journal of Evaluation in Clinical Care Settings*, 6(2), 185-192.
- Wasserheit, J. N., Aral, S. O., Holmes, K. K., and Hitchcock, P. J. (Eds.) (1991) *Research Issues in Human Behavior and Sexually Transmitted Diseases in the AIDS Era*. Washington, D.C.: American Society for Microbiology. (\*)
- Weisberg, H. (1978) How Much does ESAA Really Accelerate Academic Growth. *Journal of Educational Statistics*, 3(1), 69-78. (\*)
- Weisburd, D., Sherman, L., & Petrosino, A. J. (1990) *Registry of Randomized Criminal Justice Experiments in Sanctions*. Washington, DC: National Criminal Justice Reference Service (SRO 19000-00/129725).
- Woodruff, S. I. (1997) Random Effects Models for Analyzing Clustered Data from a Nutrition Education Intervention. *Evaluation Review*, 21 (6), 688-697. (\*)

## Summary

Randomized trials have yielded good evidence about which programs work better, for whom, and how long in medicine, criminology, welfare reform, education and other sectors. Trials that involve the random assignment of places such as communities, housing projects, organizations, neighborhoods, schools or other entities, to different interventions so as to generate fair comparison are not yet common. But they can be justified for theoretical, statistical, policy, political and ethical reasons.

The theoretical rationale for place-based trials is that programs work when organizational elements in a place concert, e.g., community-wide programs. A basic statistical rationale for focusing on places or institutions as the units of random allocation in a trial is that conventional statistical analyses of the effect of programs can be wrong when analyses are based on individuals rather than on institutions.

The policy and political rationale for focusing on organizations and other sites as the units for study is that organizations are the immediate target for a government agency and foundation action. Individuals are not. The ethical and cultural rationale is that, at times, the random allocation of organizations to alternative regimens, in the interest of a fair comparison, is more acceptable and desirable than random

assignment of individuals.

The feasibility of using places, and other entities as units in controlled randomized trials is demonstrable. Entities have been allocated at random to different interventions in trials on fertility control methods, welfare enhancement, education reform, law enforcement, health-risk reduction programs and others. The units of random allocation have been neighborhoods, factories, classrooms and schools, hospitals, saloons, and so on.

There are difficulties in executing such trials, of course. Able administrators, researchers, civil servants, and foundation people have met the challenges at times. Statisticians and methodologists who understand the design of place based randomized trials can tailor the trials design at times so as to meet the challenges.

Regardless of the difficulties, the future of place-based randomized trials is promising. They are being run more frequently. Place-based trials have been mounted in diverse areas such as education, crime and delinquency, mental health, employment, health risk reduction and welfare. They are an important tool in generating evidence about which programs work and for whom, which do not work, and which programs are promising.