American Behavioral Scientist

Estimating the Effects of Interventions That are Deployed in Many Places: Place-Randomized Trials

Robert Boruch, Henry May, Herbert Turner, Julia Lavenberg, Anthony Petrosino, Dorothy de Moya, Jeremy Grimshaw and Ellen Foley

The online version of this article can be found at:
http://abs.sagepub.com/cgi/content/abstract/47/5/608

Published by:

$SAGE Publications

http://www.sagepublications.com

Additional services and information for *American Behavioral Scientist* can be found at:

Email Alerts: http://abs.sagepub.com/cgi/alerts

Subscriptions: http://abs.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 39 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://abs.sagepub.com/cgi/content/refs/47/5/608

Estimating the Effects of Interventions That Are Deployed in Many Places

Place-Randomized Trials

ROBERT BORUCH
HENRY MAY
HERBERT TURNER
JULIA LAVENBERG
University of Pennsylvania

ANTHONY PETROSINO
Campbell Collaboration

DOROTHY DE MOYA
University of Pennsylvania

JEREMY GRIMSHAW
Campbell Collaboration

ELLEN FOLEY
Brown Universtiy

Place-randomized trials have been mounted in a variety of countries to estimate the relative effects of interventions that are intended to ameliorate problems or improve conditions in organizations and geopolitical jurisdictions. This article presents studies in which villages, police hot spots, housing developments, hospital units, schools, and other entities are the units of random allocation. The challenges to such work, approaches to meeting them, and the value added of such trials are outlined. The scientific value added includes better evidence on what works at the macro level. Web-oriented registers of such trials are being developed by the Campbell Collaboration.

Keywords: *randomized cluster trials; group randomized trials; experiment; macroexperiment*

1. INTRODUCTION

A place-randomized trial is a study in which a number of places are randomly assigned to two or more interventions so as to learn which intervention works best. The places may be villages or neighborhoods, schools or juvenile facilities,

housing developments, or other organizations. On account of the random assignment, the places that are assigned to the different interventions will not differ systematically at the outset. This equivalence permits a fair comparison of outcomes; that is, one can make an unbiased estimate of the relative effects of the interventions and a statistical statement of one's confidence in the results.

Trials in which individuals are randomly assigned to different interventions are common in medical, psychological, and other research. Random allocation of units such as places and entities is less common. Nonetheless, as Donald T. Campbell suggested in his vatic article on reforms as experiments, "For most social reforms, larger administrative units will be involved, such as classrooms, schools, cities, counties or states. We need to develop the political postures and ideologies that make randomization at this level possible" (Campbell, 1969, pg. 408; 1988, pg. 290). When Campbell wrote, place-randomized trials were rare. In what follows, we build on Campbell's insight and on others' work.

1.1 DEFINITIONS

The unit of allocation here refers to who or what is randomly assigned to different interventions in a trial. Conventional textbooks in psychology and medical trials typically handle experiments in which individuals are the units of allocation. Here, we focus on places, sites, administrative units, or entities rather than on individuals. We refer to place-randomized trials in this article. More common and general terms include cluster-randomized (Donner & Klar, 2000) and group-randomized trials (Murray, 1998). The latter could involve clusters of people in families or youth organizations, for example.

Our using the phrase "place randomized" may seem to contribute little to the social science vernacular. We think it justified, however, because the phrase is more easily apprehended in policy contexts. It also may help to arrange better our thinking about the design, execution, and analysis of such trials.

The units of analysis are those for which data are available and used to generate statistical analyses. Juvenile facilities may be the units of random allocation in a randomized trial that compares two facility-wide approaches to reducing recidivism in a multifacility trial. The units of analysis may then be the facilities. Or, the units may be both the facilities and the juveniles within the facilities.

1.2 THE CONTENTS OF THIS ARTICLE

In what follows, we explain assumptions that underlie interest in place-randomized trials, explain their rationale, and describe how they add value to understanding about whether new social interventions work. Furthermore, we identify difficulties in their use. The examples come from trials in education, welfare, criminology, and health, mounted in the United States, Mexico, China, the United Kingdom, and a half dozen other countries. This is partly to demonstrate that such trials can be carried out in a variety of settings. Learning how to identify place-randomized trials in the context of complex contemporary information systems is a challenge. We focus on the international Campbell Collaboration (http://www.campbellcollaboration.org), the Cochrane Collaboration in health care (cochrane.org), and the U.S. Department of Education's What Works Clearinghouse (http://w-w-c.org) as vehicles for identifying such trials.

1.3 ASSUMPTIONS

Our first assumption is that governments, private foundations, and others are interested at times in estimating the relative effect of new social and education. Put another way, we assume that the public is interested in answering the questions, "What works better, for whom, and for how long? And how do we know?" Contemporary governmental emphasis on evidence-based policy illustrates this theme, as does some private foundation efforts.

A second assumption is that a defensible estimate of an intervention's effect depends on determining how sites or other entities would behave in the absence of the intervention. 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, from a set of juvenile facilities that are eligible and willing to participate in a trial, some facilities might be randomly assigned to a new intervention program. The outcomes at these facilities would then be compared to the outcomes at eligible facilities that had been randomly assigned to continue operating under their existing programs.

The random assignment ensures that the two groups of facilities do not differ systematically, apart from the influence of the intervention program under study. A randomized control group provides a fair estimate of outcomes in the absence of the intervention. Alternatives to randomized trials, including time series and their vulnerabilities, are described by Campbell and Stanley (1963) and Shadish, Cook, and Campbell (2002). Here, we assume that time series data and other quasi-experimental comparisons are insufficient to produce unbiased estimates of a program's effect.

A third assumption is that the future of impact evaluation in 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. Place-randomized experiments have been undertaken to do this at times. Illustrations follow.

RATIONALE AND VALUE ADDED: WHY USE PLACES AS THE UNITS?

Why should we consider places or other entities as the units of random assignment to programs in evaluating the effect of a program? What is the value added by place-randomized trials beyond the understanding that is achieved in trials that randomly assign individuals to different interventions? The value added is a function of program theory, law, ethics and culture, policy, the counsel of advisory groups, and statistical theory.

2.1 PROGRAM THEORY

Here, theory means declaring how an intervention is supposed to have the effects that someone expects that the intervention will have. In other words, the theoretician proposes a "logic model" or "path model" to explain what should happen when a program is implemented. Many contemporary theories of social change argue that a program will work if it is delivered by organizational elements acting in concert. In other words, individuals will change their behavior partly because their environment (place) changes and partly because the intervention affects them at the individual level.

One of the most thorough histories of theory and research in the area of school-based prevention trials is given by Flay (2002). The various stages he identifies include programs based first on simple provision of information, then on approaches that appeal to motivation, values, affect, and (most recently) approaches based on skills training and the correction of the incorrect normative beliefs of youth by deploying interventions at the school level.

In another education arena, Kellam and Anthon (1998), for instance, depended on a theory of change and epidemiological data to create an intervention that could be deployed at the classroom and school level so as to reduce children's problem behavior. Their test of the Good Behavior Game involved the random allocation of children to a set of classrooms to this regimen and another set of classrooms to Mastery Learning. The Good Behavior regimen caused changes in children's norms and reductions in the likelihood of tobacco use for boys, but not for girls.

The theory that institutions ought to be the vehicles for intervention and that institutional factors will influence the effectiveness of the intervention is not new. Related theory drove randomized trials during the 1970s on how to improve hospital services for treating certain forms of mental illness. The latter included entire hospital's involvement in workshops or demonstration projects as opposed to merely sending brochures to physicians (Fairweather, Sanders, &

Tornatsky, 1974; Fairweather & Tornatsky, 1977). More recent efforts to change hospital practice, so as to improve patient care, are based on theories of change, diffusion of innovation, and how opinion leaders and local networks work. Leviton and Horbar (2002), for instance, discussed these theories as a basis for designing interventions to change perinatal and neonatal care practices in the United States and Canada.

2.2 LAW, ETHICS, AND CULTURE AND THE VALUE ADDED

One reason why places or 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 place may not be acceptable on legal, ethical, or cultural grounds. Random allocation of entire places or sites to alternative programs might be acceptable on these grounds.

The value added by place-randomized trials then is that a fair comparison can be made of interventions that are deployed at the place level. No comparison at all may be possible at the level of randomizing at the individual level.

Similarly, each of the 80 or so juvenile facilities in Sweden, for example, may object to random allocation of their clients to different programs to discern which program is most effective in reducing recidivism. Other ethical values in the local facility may take precedence, such as 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).

Political values reflect societal ethics, which are, at times, indistinguishable from political reality. In the health sector, for instance, Donner and Klar (2000) reiterate others' expert opinion that it was not politically feasible to assign some individuals within each village in Indonesia to Vitamin A supplements while others received no treatments. But it was politically feasible to randomly assign 450 villages to supplements and control conditions.

2.3 POLICY AND POLITICS, AND VALUE ADDED

Governments often sponsor new programs that directly affect sites, organizations, or other places rather than directly affecting individuals. These programs require entities to take particular actions, create transactions, and so on. A study of the effects of such a program must recognize entities as the immediate targets in a trial, although the individuals within sites may be the ultimate targets for service.

For example, a sample of 500 villages in Mexico's Progresa project were randomly assigned to an income transfer program and to control conditions so as to determine whether the program kept children out of the agricultural fields and in the schools, which it did (Parker & Tervel, 2002). The rationale for assigning

entire villages to alternative regimens, so as to produce unbiased estimates of relative effects, was partly a function of Mexican social policy.

High-quality trials on the effect of vouchers in the context of public schools versus private school education have been undertaken by Peterson et al. (2002). These trials were to understand whether voucher systems can enhance children's achievement. Families or children were assigned access to vouchers or to normal access to public schools. The random assignment in these trials ensures fair comparison at the child or family level. Such trials cannot provide unequivocal evidence about the effects of vouchers at the school, school district, or state level. To produce fair comparisons about the effects of vouchers on entire schools and the school people, districts, or state requires that entire schools, districts, or states participate in fair comparison studies.

2.4 STATISTICAL THEORY AND ANALYSIS, AND VALUE ADDED

Many contemporary statistical methods of analysis rely on the assumption that an observation on any given individual or entity in a trial 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 nonindependence. Donner and Klar (2000) and Murray (1998) give details.

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 be influenced by other gang members' responses even though the program directs attention to individual gang members. A child's grade on a test of ability to work in teams presumably will not be independent of interactions with other children on the same team. Because physicians within a medical practice are not independent of one another, groups (or the independent units) are the targets for random allocation in trials on implementing changes in practice (Grimshaw, Eccles, Campbell, & Elbourne, 2002; Leviton & Horbar, 2002).

The value added by place-randomized trials lies in the idea that it is not individuals who ought to be randomly assigned to programs, and it is not individual-level data that must be used to estimate the program's effect. Rather, the random allocation and analysis focuses first on groups, such as classrooms, or organizations, such as schools or villages. Analysis may then focus secondarily on individuals within each group or entity.

Donner and Klar (2000) give a statistically nuanced rationale that hinges on the idea of independence and impinges on other categories of rationale that are given here. Newer statistical methods have been developed that can, at times, accommodate dependence among units. These include multilevel approaches that recognize relatedness of individuals (say) within places and take advantage of randomization of places (see also Grimshaw et al., 2002).

2.5 THE COUNSEL OF ADVISORY GROUPS ON RESEARCH AND EVALUATION POLICY

Groups that advise government agencies have offered direct and indirect recommendations about randomized trials that use places as the units of allocation and analysis. In education, the National Academy of Sciences Committee on Scientific Principles in Education Research recognized as important the conduct of randomized trials in which schools, school districts, and classrooms were randomly assigned to different regimens (Shavelson & Towne, 2001).

At the request of the U.S. Congress, Sherman et al. (1997) reviewed studies of the effects of crime prevention programs. The report recognized the importance of randomized trials in estimating effects. This report's follow-ups, such as Eck (2002) and Sherman, Farrington, Welsh, and MacKenzie (2002), took seriously the idea that places such as neighborhoods, schools, premises, hot spots, and other locations ought to be the target of attention for crime prevention and law enforcement. Earlier, the National Research Council's Panel on the Understanding and Control of Violent Behavior (Reiss & Roth, 1993) called attention to multicommunity approaches and tests of the interventions so as to improve causal understanding.

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 randomized trials to learn the effectiveness of the services (Coyle, Boruch, & Turner, 1991). Multidisciplinary conferences on sexually transmitted diseases (STDs), sponsored by the National Institute on Allergy and Infectious Diseases, have led to the observation that clinical practices, churches, and other organizations, as well as communities, might properly serve as the units in randomized trials (Green & Washington, 1991).

In considering approaches to preventing abuse of controlled substances, the participants in the "Communities That Care" Evaluation Design Conference (Peterson, Hawkins, & Catalano, 1992) emphasized the idea that rigorous evaluation of a comprehensive community intervention requires an experimental design whereby communities are randomly assigned to experimental and control conditions.

England's Joseph Rowntree Foundation has been influenced by similar concerns (Farrington, 1997). 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).

At times, some individuals and teams of researchers anticipate the counsel of officially constituted advisory groups and conferences. Flay and Best (1982) and Flay (2002) did so, for instance, in identifying ways that prevention research could be improved, and their advice and concerns remain pertinent.

3. EXAMPLES AND FEASIBILITY

People often do not realize that it is possible to randomly allocate places, or other entities, to different interventions so as to permit fair estimates of the effects of one intervention versus another. The illustrations that follow are organized around the kinds of places that are randomized: schools, communities and geopolitical areas, private organizations, and public housing. Within this rudimentary framework, we make distinctions based on the outcomes of interest, for example, children's educational achievement and adult health behavior.

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 substance abuse in studies by Botvin, Baker, Dusenberg, Botvin, and Diaz (1995); Moskowitz (1984); Murray, Moskowitz, and Dent (1996); and Schaps, Moskowitz, Condon, and Malvin (1982). 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, Tobler, Ringwalt, and Flewelling (1994).

In efforts to evaluate a theory-driven program to reduce alcohol use by underage youth, Wagenaar, Murray, Wolfson, Forster, and Finnegan (1994) mounted a field trial involving 15 school districts randomly assigned to a special community-based prevention program and to a control condition. Schools have been the units in two smoking prevention experiments. The Television, School and Family Smoking Prevention Project used multiattribute balancing to randomly assign 35 Los Angeles–area schools to different media-based smoking prevention campaigns. In tests of schoolwide cardiovascular risk reduction programs, schools have been randomly assigned to such programs in four states (Hansen & Graham, 1991; Killen et al., 1988; Perry et al., 1992).

Randomized trials have been mounted to understand what kinds of programs might be deployed in education settings so as to enhance children's understanding of high-risk sexual behavior and how to avoid it. In the United States, 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 California, Kirby, Korpi, Adivi, and Weismann (1997) 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 in 50 schools were randomized in one component (Kirby, Korpi, Barth, & Cagampang, 1997). A different stream of health-related work has concerned nutrition education, for example, Woodruff's (1997) classroom-based trials in community colleges. 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 6 months.

Mobile societies have 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; Jason, Johnson, Danner, Taylor, & Krasaki, 1993) focused on children who transferred into new schools and who were, as a consequence, vulnerable. One project involved randomly assigning members of 10 matched pairs of schools to an innovative treatment program or to a control condition to determine whether their special transition program worked.

Until the late 1990s, no comprehensive school reform programs had been tested in place-randomized trials. Cook et al. (2001a, 2001b) undertook trials on Comer's theory-based approach in two locations. Results from the Chicago trial suggest positive effects on children's educational achievement. In Maryland, effects were not discernible, perhaps because the program was not implemented well in the schools that had agreed to deploy the program. Poor implementation of the intervention also may have led to null findings in Ramey et al.'s (2002) trials on school-based Head Start Transition programs. Related problems appear to have occurred in trials during the 1970s in which eligible schools were randomly assigned to special funding for programs for minority youth and reducing racial isolation (Coulson, 1978; Reichardt & Rindskopf, 1978; Weisberg, 1978).

Early examples to test different approaches in different countries to enhancing children's education or reducing their vulnerability deserve recognition. 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, Seare, Galda, & Heyneman, 1981; Jamison, Searle, & Suppes, 1980). A related randomized trial in El Salvador disintegrated; Hornik, Ingle, Mayo, McAnany, and Schramm (1972) gave an admirably candid description. In the Philippines, Aplasca et al. (1995) targeted classrooms in

schools to test a program for reducing the risk of AIDS. In Kenya, Glewwe et al. (2000) tested the effects of flip charts on children's achievement in a trial in which schools were randomly assigned to charts and teacher guides or to a controlled condition.

3.2 COMMUNITIES AND GEOPOLITICAL ENTITIES AS THE UNITS OF RANDOM ASSIGNMENT

Localities such as communities, neighborhoods, villages, and other geopolitical regions have been used as units in place-randomized trials. The following illustrations pertain to health-related initiatives, education programs, and crime prevention. They concern trials mounted in the United States, Taiwan, Colombia, Mexico, and China.

Communities have been the units of allocation in evaluations of several 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 10 communities to one of three treatments and to a control group. The intent was to change entire communities in the interest of reducing tobacco use. The Community Intervention Trial for Smoking Cessation (COMMIT) assigned 11 matched pairs of communities to its treatment and comparison groups (Freedman, Green, & Byar, 1990, cited in Peterson et al., 1992). In media-based smoking prevention campaigns, standard metropolitan statistical areas (SMSAs) have been allocated randomly to the campaigns or to control conditions (Bauman, LaPrelle, Brown, Koch, & Padgett, 1991; see Biglan et al., 2000; Wagenaar et al., 2002, for recent trials in this area).

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 also have been used as units in randomized studies of HIV risk prevention tactics (Kelly et al., 1991).

Education and nutrition 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. This was to determine that the program was effective relative to randomly assigned control areas (McKay, McKay, Sinnestera, Gomez, & Lloreda, 1978).

More recently, Parker and Tervel (2002) reported on a randomized trial involving about 500 villages in Mexico to estimate the effects of an income transfer program and other services on children's educational progress and health. The program had no discernible educational effect on children younger than age 10 but did have the effect of keeping more boys older than 10 in school and out of the agricultural work force. The program's effect on girls is being explored.

Some randomized trials have been mounted because integrating multiple services at the community level is believed 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, Basinsky, Leginski, Parker, & Goldman, 1997). About 50 agencies within each jurisdiction cooperated in the study.

Finally, consider research on crime prevention. In the Kansas City patrol experiment, 15 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, & Brown, 1974). Twenty years later, Sherman and Weisburd (1995) and Weisburd (1995) executed better trials in Minneapolis and Jersey City. The researchers identified "hot spots," that is, local areas of predictably high crime, and randomly allocated some of these areas to more intensive police patrol and the remainder to a normal patrol activity. They learned that such patrols reduced crime in the hot spots without any spillover of crime in adjacent areas. Weisburd (2002) provided an excellent review of these trials, earlier work, and challenges that were met in generating the evidence.

3.3 ORGANIZATIONS AS UNITS OF RANDOM ASSIGNMENT

Organizations here includes factories, hospitals and medical practices, nonprofit and for-profit franchise operations, and private or commercial properties. The illustrations concern health, professional education, organizational change, and crime. The trials have been undertaken in Zimbabwe, Russia, the United States, Canada, Thailand, and elsewhere.

In some countries, a sensible way to reduce the health risks of individuals is through certain private organizations or public ones. The National Institute of Allergies and Infectious Diseases, for example, invested in a test of factory-based peer education that 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, Klar, & Donner, 1995). Donner and Klar (2000) refer to two other trials using factories as units in Russia. One concerned prevention of coronary heart disease and the other concerned detection of breast cancer. Sex establishments in Thailand also have been randomly allocated to a program involving male and female condoms so as to determine whether STDs could be reduced (Fontanet et al., 1998).

The gap between conventional medical practice and evidence that supports different and better practice is important. Leviton and Horbar (2002) reported on two major trials, one that targeted nearly 30 tertiary care hospitals and one that targeted more than 100 neonatal intensive care units in the United States and

Canada. The hospitals and units were randomly allocated to interventions that were designed to increase their adherence to changing practice based on high-quality evidence in treating women in premature labor and infants who were born prematurely. The intervention for others led to increases in hospital adoption of better practice; the substantial variation in adoption rate is still being investigated (Leviton, Goldenberg, Baker, & Freda, 1999). As of this writing, the analysis of the effects of a related but different approach to engaging neonatal care units in better treatment is underway. (See Soumerai et al., 1998, for an interesting example involving 40 community hospitals' engagement in a trial on introducing better medical handling of victims of acute myocardial infarction.)

In an early crime prevention trial, 120 franchise convenience stores were reported to have been randomly assigned to a specialized robbery prevention program or to a control condition. Mounted in the 1970s, the results suggest that this approach, invented and tested by the Western Behavioral Sciences Institute, reduced robberies from nearly 60% to just more than 40%. The summary is given in *National Criminal Justice Research Abstracts*; we are seeking the original report. In a more recent trial, drug nuisance properties in Oakland were randomly assigned to a civil remedies approach to crime prevention and to control conditions. These properties (localities) include commercial ones, private homes, and rental properties. The effect on calls to request public service depended on type of property, with private residences undergoing a decline and commercial properties an increase (Mazarole, Price, & Roehl, 2000).

Nonprofit service organizations have, at times, committed resources to randomized trials. In an early place-randomized trial, for instance, Good Will Industries in the United States 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.

3.4 PUBLIC HOUSING DEVELOPMENTS AS THE UNITS OF RANDOM ASSIGNMENT

In the United States, public housing developments (buildings or sets of buildings) serve low-income families. In some cities, people who live in some of these housing developments are at risk of criminal victimization, poor health service, and scant opportunities for employment, training, social support, and child care. The illustrations that follow are pertinent to these different kinds of outcomes.

To date, the most ambitious economic trial involving housing developments is Jobs Plus (Bloom & Riccio, 2002). Its aim was to enhance social capital, indexed through employment rates and other outcomes, by providing training and employment services, financial incentives, and technical support to everyone in the poverty-level housing developments. Seven cities, initially selected from more than 50, were the locations for randomly assigning members of pairs

of developments to either Jobs Plus or to a control condition. The study's interim results and working papers are given by Bloom and Riccio (2002) and can be found at http://mdrc.org.

Eighteen housing projects in four cities were the units of random allocation in a trial of an opinion leader approach and support services to reducing the risk of HIV among women (Sikkema et al., 2000). Results suggest that protected sexual activity, including use of condoms, increased remarkably and unprotected activity decreased as a result of the approach. Sikkema's (2002) report on more recent work involves new trials in which adolescents in housing developments are the ultimate targets, with housing developments again being the units of random assignment.

In the crime prevention arena, the Davis and Medina-Ariza (2001) randomized trial involved 60 housing projects to evaluate a program for preventing elder abuse. Early results suggested an increase in reporting of abuse, but in the 12-month follow-up, differences in reports between control and treated projects could not be discerned. Davis and Taylor's (1997) study also engaged housing projects in a place-randomized trial on a family violence prevention program.

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 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. Roughly speaking, statistical power refers to our ability to discern the relative effectiveness of the two literacy programs in a randomized trial if indeed there is a difference in effectiveness. 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 ensure that its statistical power is about .80? Assume that the true difference between the programs is small (having a standardized effect size of .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 ensure that real differences between the

interventions are detected. Assuming a low similarity rate (intraclass 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.

LaPrelle et al. (1992) found that their trial on community-based substance use prevention in citywide programs was underpowered. Four treatments in an experiment were spread throughout 10 communities. Their posttrial 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 ensure 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: Do a statistical power analysis and then engage as many entities as possible in the trial. Free power analysis software is available at http://www.ssicentral.com, based on work by Raudenbush (1997), Raudenbush and Liu (2000), and Raudenbush and Liu (2001). Newer approaches, involving combining times series data with cross-sectional data on sites and cohort studies, are given in Bloom and Riccio (2002).

4.2 THEORY AND MEASUREMENT SYSTEMS

Rudimentary theory of how an intervention is supposed to work, of course, guides us in selecting what variables should be measured. The measurement system determines how and how well the variable will be measured.

Consider the multisite 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, media coverage, and the attitudes of students and youth. This study surveyed retail alcohol outlets to determine if indeed outlets failed to request 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 based on state and local record systems.

Some trials of strategies to change medical practice depend on diffusion theory to guide the work, for example, Leviton and Horbar (2002). Other trials depend on heuristics, a simple logic, or implicit theory and properly ignore the vapid theory often promulgated in the social sciences, for example, Grimshaw et al. (2002). In either case, administrative and medical records that reflect practice, and adoption of better practice, typically serve as the outcome measures.

The quality of measurement in place-randomized trials is as important as it is in other kinds of trials. Institutional context and pressures to achieve certain

goals can, of course, lead to under- or overreporting. Smith (2002), for example, describes a tendency to underreport "all kinds of pregnancies" (p. 22) as well as births and abortions in Chinese trials on fertility control interventions in multiple townships and villages within township.

4.3 ENGAGING PLACES

Engaging places, such as hospitals, schools, and other administrative units, and other entities in a randomized trial usually requires considerable skill. Walker, Campbell, Grimshaw, and the TEMPEST Group (2000), for example, provided an exceptionally detailed description of strategies for recruiting U.K. hospitals into randomized trials. They focused 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.

In the Jobs-Plus trials described by Bloom and Riccio (2002), about 50 cities were seriously considered on the basis of their public housing authorities, composition of public housing developments, and so on. Because the trials were privately sponsored (by Rockefeller Foundation), substantial screening could be done, and was done, on the basis of the housing developments' capacity to collaborate with community agencies that are relevant to public housing. Of the seven cities finally selected, one "dropped out" in a potentially positive sense, having located sources of money that would deploy a program that would swamp JobsPlus (see reports at http://mdrc.org).

In school-based prevention trials, Ellickson (1994) reported that 11 schools out of about 60 schools that were invited to participate declined to do so. One school, for instance, could not participate on account of a court order demanding considerable resource allocation on racial equity. Four of the 11 schools declined to participate because they already had prevention programs in place. The reasons for other declinations concerned their capacity, such as inability to ensure community support for engaging in the experiment.

Understanding whether and how to engage places in place-randomized trials, and to maintain the engagement, is not easy. One effort to lay out the state of the art was undertaken by the Campbell Collaboration in 2002 at a Rockefeller Foundation–sponsored conference on the topic (Boruch & de Moya, 2002).

4.4 TEMPORAL AND STRUCTURAL STABILITY OF UNITS AND CONTEXT

We often expect organizations, sites, and other places to change little over a short period of time. Nonetheless, the stability of certain characteristics of places may be low. Trends may reverse direction. Bauman et al. (1991), for example, found high positive correlation over a 2-year period for adolescents' reported rates of recent smoking in a sample of 10 cities. They also found a

modest negative correlation for adolescents' rates of experimentation with smoking in the same cities.

One normally assumes that the places 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). Furthermore, feeder schools changed as a consequence of changes in busing patterns and the creation of magnet schools that drew students from areas outside of 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 since the census had been taken (Kellaghan, Madaus, & Airasian, 1982). Tennessee's experiment on school incentives encountered difficulties because some schools closed or were consolidated with other schools (Bickman, 1985). Such problems have to be anticipated in designing place-randomized trials and in analyzing the trials' results.

Programmatic shifts at the nation level may, of course, dramatically affect a trial. A major Chinese study of different approaches to fertility control at the township level, for instance, appears to have been swamped by national initiatives that led to a decline in the abortion to birth ratio (Smith, 2002).

4.5 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, that is, have not yet been given the resources. The number of communities involved in such a study often may 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 similar at the outset; that is, the groups may be imbalanced on account of chance. This "unhappy random configuration" will make interpretation of data difficult. One approach used to reduce the problem 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. Furthermore, 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.

4.6 IMPLEMENTATION FIDELITY AND MEASUREMENT

It makes no sense to estimate the effect of a new program in a place-randomized trial (or any other kind of trial) unless one can verify that the program activities occur and can be described. We need to know that the program has been deployed by observing and measuring its implementation. Learning how to observe any of these reliably and to assure fidelity in the program's implementation and its measurement is demanding.

Place-randomized trials that attempt to evaluate interventions that involve integration or coordination of services across many agencies and communications among them present special problems. Developing indicators of integration, coordination, and communication 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 colocation of services (Randolph et al., 1997).

To judge from candid reports on place-randomized trials, implementation of the interventions that are supposed to be deployed is far more difficult than implementing the trial itself. In efforts to improve education in the United States, for example, "training the trainer" approaches had to confront challenges at the school level, in leadership teams, and at the teacher level and handle massive changes at the multiple school level (Porter, 2002). Smith's (2002) report on Chinese trials on fertility control efforts documents considerable variation among townships and villages with a township in what approach to fertility control was acceptable and how and how well the local service providers could provide information (Smith, 2002). In trials designed to change medical practice, Leviton and Horbar (2002), Grimshaw et al. (2002), and Sikkema (2002) also report the problem: changing practice is difficult, the change has to be observed, and different approaches to encouraging change have to be tested.

4.7 IDENTIFYING PLACE-RANDOMIZED TRIALS

Identifying place-randomized trials is difficult. This is partly because the names for these kinds of trials and related trials vary in the research literature.

The authors of articles in refereed research journals often do not make the units of random allocation clear in the title of their article, abstract, or keyword listing. Eager readers of a refereed journal or report may discover that the units of random allocation are schools or hospitals, but not until they reach the Methods section of the article. The implication is that anyone who relies on electronic search engines, which usually rely on abstracts, keywords, or titles of articles rather than full text, will fail to uncover many trials and often will be forced to use different terms to identify these trials.

One approach to resolving the problem of identifying place-randomized trials is to build electronic registers of randomized trials based on hand searches of literature and other tactics. The Campbell Collaboration is trying to do this (http://www.campbellcollaboration.org). In particular, the Collaboration's Social, Psychological, Educational, and Criminological Trials Register (C2-SPECTR) includes about 200 entries that concern verified place-randomized trials; they are denominated as CRTs, which stands for cluster-randomized trials (Petrosino, Boruch, Rounding, McDonald, & Chalmers, 2000). Grimshaw et al. (2002) summarized results of a review of 100 such trials on attempts to change in medical practice so as to close the gap between conventional practice and practice that conforms to evidence-based guidelines.

At this writing, we can give evidence on the prevalence and incidence of reports on cluster-randomized, group-randomized, and place-randomized trials based on C2-SPECTR. The evidence is given in Figures 1 and 2. One obvious implication of the charts is that reports on such trials have increased and continue to do so. A second implication is that the incidence of reports has increased each year but progress is choppy.

As of 2002, researchers who mined C2-SPECTR have been asked to contribute to it. This means that people who wanted to identify place-based trials in their arena of interest, and who exploited C2-SPECTR so as to find them, were expected to contribute entries voluntarily to the register.

The Campbell Collaboration also expects its registers of trials to feed into, and to be fed by, other registers. For instance, a What Works Clearinghouse (WWC) (w-w-c.org) that has been funded by the U.S. Department of Education will depend partly on C2-SPECTR. The WWC intends to prepare reviews of topics, approaches, and interventions in the education arena and builds on high quality and transparent standards of evidence and/or parallel efforts to produce information that have been developed by the Campbell Collaboration and others. The Campbell Collaboration in turn expects to exploit the WWC effort to populate C2-SPECTR and other electronic registers on evidence and reviews of evidence. The Campbell Nordic Center based in Copenhagen will want access to C2-SPECTR. In return, the Nordic Center colleagues are expected to help populate C2-SPECTR and other registers.
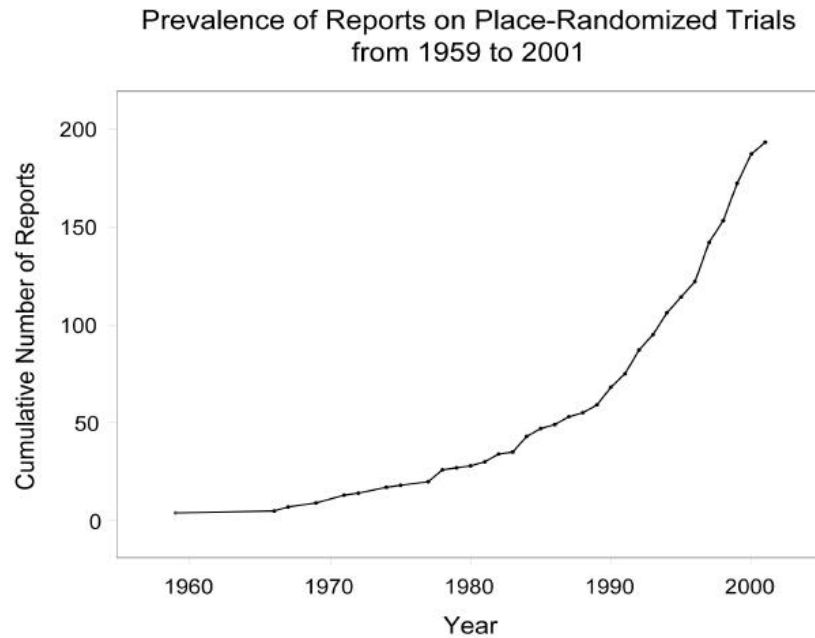
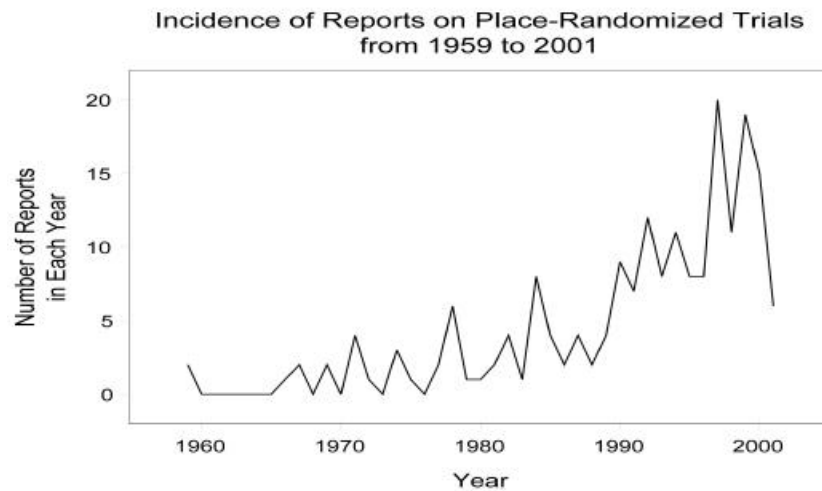**Figure 1: Prevalence of Reports on Place-Randomized Trials From 1959 to 2001**

Figure 2: Incidence of Reports on Place-Randomized Trials From 1959 to 2001

5. SUMMARY

Randomized trials yield scientifically defensible good evidence about which programs work better, for whom, and how long in medicine, criminology, social welfare, education, and other sectors. Some of these trials involve the random assignment of places, such as communities, housing projects, organizations, neighborhoods, and schools or other entities, to different interventions to generate a fair comparison. Place-randomized trials are not common. However, they can be justified for theoretical, statistical, policy, political, and ethical reasons.

The theoretical rationale is that people's well-being can be advanced at times when organizational elements in a place are concerted. A basic statistical rationale for focusing on places as the units of random allocation in a trial is that conventional statistical analyses of the effect of programs and inferences about program effects can be wrong when analyses are based on individuals rather than on institutions.

The policy and political rationale for focusing on places and organizations as the units for study is that organizations are the immediate target for many government agency and foundation actions. Individuals are not. The ethical and cultural rationale for such trials 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 tests of programs in fertility control, welfare enhancement, education reform, law enforcement, health–risk reduction programs, crime prevention, and others. The units of random allocation have been neighborhoods, factories, classrooms and schools, housing projects, hospitals, saloons, private organizations, and so on.

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

Identifying such trials in the literature, to learn about these trials, is also difficult. Efforts undertaken by the Campbell Collaboration to identify all randomized trials for a publicly accessible electronic register have been productive, and they continue. The U.S. Department of Education's What Works Clearinghouse, the Campbell Collaboration, the Cochrane Collaboration, and other efforts will help greatly to identify many different kind of trials in education.

Regardless of the difficulties, the future of such trials is promising. Place-randomized trials are an important way to generate evidence about which programs work and for whom, which do not work, and which programs show potential. Unless we develop fair place-randomized trials, we will have to

depend on ignorant advocates for change, on one hand, and ignorant opponents of change, on the other hand. People deserve better.

REFERENCES

Aplasca, M., Siegel, D., Mandel, J. S., Santana, R., Paul, J., Hudes, E. S., et al. (1995). Results of a model AIDS prevention program for high school students in the Philippines. *AIDS* (Suppl. 1), pp. 7-13.

Bauman, K. E., LaPrelle, J., Brown, J. D., Koch, G. C., & 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*, *81*, 597-604.

Bickman, L. (1985). Randomized field experiments in education. *New Directions for Program Evaluation*, *28*, 39-54.

Biglan, A., Ary, D. V., Smolkowski, K., Duncan, T., & Black, C. (2000). A randomized controlled trial of a community intervention to prevent adolescent tobacco use. *Tobacco Control*, *9*, 24-32.

Bloom, H. S., & Riccio, J. A. (2002). *Using place-based random assignment and comparative interrupted time-series analysis to evaluate jobs plus employment program for public housing residents*. Paper presented at the Campbell Collaboration conference on place-randomized trials, Rockefeller Foundation, Bellagio, Italy [Author: Manpower Demonstration Research Corporation, http://mdrc.org].

Boruch, R. F., & de Moya, D. (2002). *Agenda and background papers: Campbell Collaboration Conference on Place Randomized Trials*. Rockefeller Foundation Conference Center, Bellagio, Italy. Available at http://campbellcollaboration.org

Botvin, G. J., Baker, E., Dusenburg, L., Botvin, E. M., & 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. In S. Overman (Ed.), *Methodology and epistemology for social science: Selected papers by Donald T. Campbell* (Chap. 11, pp. 290-314). Chicago: University of Chicago Press.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.

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., & Turner, C. F. (Eds.). (1991). *Evaluating AIDS prevention programs* (Expanded ed.). Washington, DC: National Academy of Sciences.

Davis, R., & Medina-Ariza, J. (2001, September). *Results from an elder abuse experiment in New York: Research in brief*. Washington, DC: National Institute of Justice.

Davis, R. C., & Taylor, B. G. (1997). A proactive approach to family violence: The results of a randomized experiment. *Criminology*, *35*(2), 307-333.

Dean, J., Seare, B., Galda, K., & 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., & Klar, N. (2000). *Design and analysis of cluster randomized trials in health research*. New York: Oxford University Press.

Eck, J. E. (2002). Preventing crimes at places. In L. M. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 241-204). New York: Routledge.

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), 79-101.

Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Flewelling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of project DARE's outcome evaluations. *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., & Tornatsky, L. G. (1977). *Experimental methods for social policy research*. New York: Pergamon.

Farrington, D. P. (1997). Evaluating a community crime prevention program. *Evaluation*, *3*, 157-173.

Flay, B. R. (2002, November). *School-based randomized trials for evaluating problem behavior prevention programs*. Paper presented at the Rockefeller Foundation Conference on Place-Randomized Trials, Bellagio, Italy [Author: University of Illinois, Chicago].

Flay, B. R., & Best, J. A. (1982). Overcoming design programs in evaluating health behavior problems. *Evaluation and the Health Professions*, *5*, 43-69.

Fontanet, A. L., Saba, J., Chandelyng, V., Sakondhavat, C., Bhiraleus, P., Rugpao, V., et al. (1998). Protection against sexually transmited diseases by granting sex workers in Thailand the choice of using the male or female condom: Results from a randomized cotnrolled trial. *AIDS*, *12*(14), 1851-1859.

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*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Glaser, E. M., Coffey, H. A., et al. (1967). *Utilization of applicable research and demonstration results*. Los Angeles: Human Interaction Research Institute.

Glewee, P., Kremer, M., Moulin, S., Zitkewitz, E. (2000). *Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya* (National Bureau of Economic Research working paper #8018, JEL No. 121, N37). Retrieved from http://www.nber.org.papers

Green, S. B., & 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, & P. J. Hitchcock (Eds.), *Research issues in human behavior and sexually transmitted diseases in the AIDS era* (pp. 345-352). Washington, DC: American Society for Microbiology.

Grimshaw, J., Eccles, M., Campbell, M., & Elbourne, D. (2002). *Cluster randomized trials of professional and organizational behaviour change interventions in health care settings*. Paper presented at the Rockefeller Foundation Conference on Place-Randomized Trials, Bellagio, Italy [Principal author: University of Ottawa].

Grossman, D. C., Neckerman, H. J., Koepsall, T. D., Liu, P., Asher, K. N., Beland, K., et al. (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, G. W. (1991). Preventing alcohol, marijuana and cigarette use among adolescents: Peer pressure resistance training versus establishing conservative norms. *Preventative Medicine*, *20*, 414-430.

Hornik, R. C., Ingle, H. T., Mayo, J. K., McAnany, E. G., & 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., Johnson, J. H., Danner, K. E., Taylor, S., & Krasaki, K. S. (1993). A comprehensive, preventive, parent-based intervention for high-risk transfer students. *Prevention in Human Services*, *10*(2), 27-37.

Jason, L. A., Weine, A. M., Johnson, J. H., Sohlberg, M. L., Filippelli, L., Turner, E., et al. (1992). *Helping transfer students: Strategies for educational and social readjustment*. San Francisco: Jossey-Bass.

Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). *The effects of standardized testing*. Boston: Kluwer-Nijhoff.

Kellam, S. G., and Anthon, J. C. (1998). Targeting early antecedents to prevent tobacco smoking: Findings from an epidemiologically based randomized field trail. *American Journal of Public Health*, *88*(10), 1490-1495.

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., Stevenson, L. U., Hauth, A. C., Brasfield, T. L., et al. (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), 1728-1733.

Kirby, D., Korpi, M., Adivi, C., & Weismann, J. (1997). An impact evaluation of Project SNAPP: An AIDs prevention and pregnancy middle school program. *AIDS Education and Prevention*, *9*(Suppl. A), 44-61.

Kirby, D., Korpi, M., Barth, R. P., & Cagampang, H. H. (1997). The impact of postponing sexual involvement curriculum among youths in California. *Family Planning Perspectives*, *29*, 100-108.

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), 115-130.

Leviton, L. C., Goldenberg, R. L., Baker, C. S., & Freda, M. (1999). Randomized controlled trial of methods to encourage the use of antenatal corticosteroid therapy for fetal maturation. *Journal of the American Medical Association*, *281*(1), 46-52.

Leviton, L. C., & Horbar, J. D. (2002). *Cluster randomized trials for the evaluation of strategies designed to promote evidence-based practice in perinatal and neonatal care*. Paper presented at the Campbell Collaboration–Rockefeller Foundation Conference on Place-Randomized Trials, Bellagio, Italy [Author: Robert Wood Johnson Foundation, Princeton, New Jersey, and University of Vermont Oxford Network].

Mazarole, L., Price, J., & Roehl, J. (2000). Civil remedies and drug control: A randomized trial in Oakland, California. *Evaluation Review*, *24*(2), 212-241.

McKay, H., McKay, A., Sinnestera, L., Gomez, H., & Lloreda, P. (1978). Improving cognitive ability in chronically deprived children. *Science*, *200*(4), 270-278.

Moskowitz, J., (1984). The effects of drug education and follow-up. *Journal of Alcohol and Drug Education*, *3*, 45-49.

Murray, D. (1998). *Design and analysis of group randomized trials*. Oxford, UK: Oxford University Press.

Murray, D. M., McKinlay, S. M., Martin, D., Donner, A. P., Dwyer, J. H., Raudenbush, S. W., et al. (1994). Design and analysis issues in community trials. *Evaluation Review*, *18*(4), 493-514.

Murray, D., Moskowitz, J. M., & Dent, C. W. (1996). Design and analysis issues in community-based drug abuse prevention. *American Behavioral Scientist*, *39*(7), 853-867.

Parker, S. W., & Tervel, G. M. (2002). *Randomization and social program evaluation: The case of progresa*. Paper presented at the Campbell Collaboration–Rockefeller Foundation Conference on Place-Randomized Trials, Bellagio, Italy [Authors: Centro de Investigacion y Docencia Economicas, Mexico City].

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., et al. (1992). The Child and Adolescent Trial for Cardiovascular Health (CATCH): An overview of intervention program and evaluation methods. *Cardiovascular Risk Factors*, *2*(1), 36-43.

Peterson, P. L., Hawkins, J. D., & Catalano, R. F. (1992). Evaluating comprehensive community drug risk reduction interventions. *Evaluation Review*, *16*(6), 579-602.

Peterson, P. E., Howell, W. G., Wolf, P. J., & Campbell, D. E. (2002). *The Education Gap: Vouchers and Urban Schools*. Washington, D.C.: Brookings Institution.

Petrosino, A., Boruch, R., Rounding, K., McDonald, S., & Chalmers, I. (2000). The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions. *Evaluation and Research in Education*, *14*(Nos. 3, 4), 206-219.

Porter, A. (2002). *Memo on experimental design to measure effects of assisting teachers in using data on enacted curriculum to improve instruction in mathematics and science education*. Paper presented at the Campbell Collaboration–Rockefeller Foundation/Campbell Collaboration Conference on Place-Randomized Trials, Bellagio, Italy [Author: University of Wisconsin].

Randolph, F., Basinsky, M., Leginski, W., Parker, L., & Goldman, H. H. (1997). Creating integrated service systems for homeless persons with mental illness: The access program. *Psychiatric Services*, *48*(3), 369-373.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*(2), 173-185.

Raudenbush, S. W., & Liu, X. -F. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*(2), 199-213.

Raudenbush, S. W., & Liu, X. -F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, *6*(4), 387-401.

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.

Riecken, H. W., Boruch, R. F., Campbell, D. T., Caplan, N., Glennan, T. C., Pratt, J. W., et al. (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, IL 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*, 245-454.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.

Shavelson, R., & Towne, L. (Eds.). (2001). *Scientific research in education*. Washington, DC: National Academy of Sciences Press.

Sherman, L., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized controlled trial. *Justice Quarterly*, *12*(40), 625-648.

Sherman L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (2002). *Evidence-based crime prevention*. New York: Routledge.

Sherman, L. W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising?* Washington, DC: Office of Justice Programs and National Institute of Justice; and College Park, Maryland: University of Maryland (NCJ 165366).

Sikkema, K. J. (2002). *HIV prevention among women in low income housing developments: Issues and intervention outcomes in a randomized place based controlled trial*. Paper presented at the

Campbell Collaboration–Rockefeller Foundation Conference on Place-Randomized Trials, Bellagio, Italy [Author: Yale University].

Sikkema, K., Kelly, J., Winnett, R., Solomon, L., Cargill, V., Roffman, R., et al. (2000). Outcomes of a community level HIV prevention intervention for women living in 18 low income housing developments. *American Journal of Public Health*, *90*(1), 57-93.

Simpson, J. M., Klar, N., & 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.

Smith, H. L. (2002). *Introducing new contraceptives in rural China: A field experiment*. Paper presented at the Campbell Collaboration–Rockefeller Foundation-Campbell Collaboration Conference on Place-Randomized Trials, Bellagio, Italy [Author: University of Pennsylvania].

Soumerai, S. B., McLaughlin, T. J., Gurwitz, J. H., Guadgnoli, E., Hauptman, P. J., Borbas, C., et al. (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. (Issue ed.). (1998). Evaluation as discourse in social work research. *Scandinavian Journal of Social Work Research*.

Wagenaar, A. C., Murray, D. M., Gehan, J. P., Wolfson, M., Forster, J. L., Toomey, T. L., et al. (1997). *Communities Mobilizing for Change on Alcohol (CMCA): Outcomes from a randomized trial*. Report submitted for publication, University of Minnesota.

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.

Walker, A. E., Campbell, M. K., Grimshaw, J. M., & 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.

Weisberg, H. (1978). How much does ESAA really accelerate academic growth? *Journal of Educational Statistics*, *3*(1), 69-78.

Weisburd, D. (1995). Policing hot spots: The Jersey City drug market analysis experiment. *Justice Quarterly*, *12*, 711-735.

Weisburd, D. (2002). *Hot spots policing experiments and criminal justice research: Lessons from the field*. Paper presented at the Campbell Collaboration–Rockefeller Foundation Conference on Place-Randomized Trials, Bellagio, Italy [Authors: Hebrew University of Jerusalem and University of Maryland].

Woodruff, S. I. (1997). Random effects models for analyzing clustered data from a nutrition education intervention. *Evaluation Review*, *21*(6), 688-697.

ROBERT BORUCH is the University Trustee Chair Professor at the Graduate School of Education, Statistics Department of the Wharton School, and the Fels Center for Government, all at the University of Pennsylvania. He is cochair of the Campbell Collaboration's Steering Committee.

HENRY MAY recently completed a doctoral program in policy research, evaluation, and measurement at the Graduate School of Education at the University of Pennsylvania. He is a research specialist at Consortium for Policy Research in Evaluation (CPRE) and a lecturer at the Graduate School of Education. His research interests include the application of advanced statistical methods in education research, program evaluation, equity within education systems, and standards-based reform.

HERBERT TURNER is Manager of Research Projects and a member of the Information Retrieval Methods Groups Steering Committee, both with the International Campbell Collaboration. He is also a postdoctoral student in the Graduate School of Education at the University of Pennsylvania.

JULIA LAVENBERG is a University Trustee Chair Fellowship recipient and doctoral student in the Graduate School of Education at the University of Pennsylvania. She serves on the Steering Committee of the Campbell Collaboration's Information Retrieval Methods Group. Her research interests center on the identification of randomized trials in social welfare, education, and criminology; the exploration of the specialized terminology used by social scientists to document their research; and parental involvement in school-based health promotion programs.

ANTHONY PETROSINO is a research consultant based in Bedford, Massachusetts. He is currently coordinator for the Campbell Collaboration's Crime and Justice Group (www.aic. gov.au/campbellcj) and also assists the Study of Decisions in Education at Harvard.

DOROTHY DE MOYA is Executive Officer of the Campbell Collaboration and serves the International Campbell Collaboration Steering Group. She is an adjunct assistant professor at the University of Pennsylvania, School of Nursing. Dr. de Moya coconvened the C2 Rockefeller Bellagio Conference on Place Randomized Trials and contributes to the Department of Education's What Works Clearinghouse. She earned her doctoral degree from the Catholic University of America School of Nursing, Washington, D.C.

JEREMY GRIMSHAW holds a Canada Research Chair in health Knowledge transfer and uptake and is the Director of the Clinical Epidemiology Program of the Ottawa Health Research Institute and the Centre for Best Practices, Institute of Population Health, University of Ottawa, Canada. He trained as a family practitioner and received his PhD in Health Services Research from the University of Aberdeen, UK. He is interested in evaluating methods of improving quality improvement in health care and has conducted many cluster randomized trials of such interventions.

ELLEN FOLEY is a principal associate at the Annenberg Institute for School Reform at Brown University. She earned her BA in political science from Boston College and her MSEd and doctoral degrees in education policy from the University of Pennsylvania.