PREFACE

Better Evaluation for Evidence-Based Policy: Place Randomized Trials in Education, Criminology, Welfare, and Health

By
ROBERT BORUCH

Phrases such as "evidence-based policy" are seductive and are used promiscuously in some quarters. Here, we take the phrase and its intent seriously, as others have. In the international sector, for instance, the Campbell Collaboration was created to generate systematic reviews of dependable evidence on the effects of policies, programs, and practices. This is in the arenas of crime and justice, education, and social services. The Cochrane Collaboration, Campbell's older sibling, was created to achieve a similar aim in health care. Their reviews of evidence are intended to assist people in making decisions.

Executing the Campbell Collaboration's and the Cochrane Collaboration's mission requires specifying what high-quality evidence is. The Campbell Collaboration's focus is similar to that of the Cochrane Collaboration. It is on randomized controlled trials that are designed to evaluate the relative effects of different interventions. See http://www.campbellcollaboration.org and http://www.cochrane.org.

One important and emerging vehicle for generating dependable evidence falls under the rubric of place randomized trials. In fact, one activity that led to this *Annals* volume was a U.S. trial in which entire housing developments were randomly assigned to the Jobs Plus program or to control conditions to generate high quality of evidence on the program's effect on wage rates, employment, and other outcomes. The Rockefeller Foundation and other organizations contributed resources for this trial. More pertinent here, the Rockefeller Foundation also pro-

Robert Boruch is University Trustee Chair Professor at the University of Pennsylvania.

NOTE: Research and development work on this topic has been supported by the Rockefeller Foundation and was organized under the auspices of the Campbell Collaboration. Dorothy de Moya, executive officer of the Collaboration, organized the Bellagio meetings and assisted in the organization of the New York meetings on this topic.

DOI: 10.1177/0002716205275610

vided support for a broadened view of place randomized trials under Campbell Collaboration auspices.

The Campbell Collaboration initiative was actualized partly through commissioning papers that appear in this *Annals* volume. The papers were vetted in conferences organized by the Campbell Collaboration and at which authors presented briefings on their papers in Bellagio, Italy, and in New York. The Campbell initiative also involved developing an internet accessible register of randomized trials in which places, groups, clusters of individuals, and entities are the target of studies to understand what works.

The articles in this issue of the *Annals* are diverse in some respects. They also have some remarkable commonalities. This introduction covers both and lays out some important dimensions of the topic of place randomized trials. It also identifies new issues and questions that ought to be addressed in further research and some recent efforts to enhance the quality of reporting and in designing such trials.

Recent History: Evidence-Based Policy and Randomized Trials

A variety of organizations have undertaken substantial efforts to understand whether and how randomized trials can be used to inform evidence-based policy.

During 2000 to 2005, for instance, the World Bank's Operations Evaluation Division laid on workshops (at Carleton University) and conferences (Washington, D.C.) in which randomized trials were considered seriously as a vehicle for generating better evidence. The most pertinent of recent proceedings of the conference are given in Pitman, Feinstein, and Ingram (2005); see especially the paper by Rawlings (2005) on randomized trials in Central and South American countries.

During 2000 to 2005, the U.S. Institute of Education Sciences (IES) was created and began to invest money in place randomized trials to understand how to improve children's learning, character education, and teacher development. During the same period, various federal agencies in the United States began to cooperate with one another in developing standards of evidence and producing better evidence. The U.S. Office of Management and Budget, for instance, has invested intellectual and organization resources in harmonizing the language used to describe research designs, including randomized trials.

The Organization for Economic Cooperation and Development (OECD), until recently, had invested mainly in passive surveys. The OECD-IES Conference on Evidence suggests a shift toward better evidence based on fair comparisons, especially randomized trials.

Within countries, the push toward better evidence has led to both quarrels and to defensible evidence from randomized trials. The U.K. Cabinet Office's evaluation unit (Davies 2004) has fostered the development of new randomized trials to evaluate programs for low-income families, for instance. In the health sector, at

least one special issue of *Statistics in Medicine* is dedicated toward cluster (place) randomized trials (Campbell, Donner, and Elbourne 2001).

In the crime arena, the U.S. National Academy of Sciences has reiterated collective ignorance about information on firearms and their use in crime in the United States. The report also reiterates the need for controlled trials, including place randomized trials, on what works to reduce the misuse of firearms (Wellford, Pepper, and Petrie 2005).

Flay and Collins, in this *Annals* volume, give a recent history on school-based prevention research. Sikkema also does so for one strand of work on AIDS prevention work. Earlier, Boruch and Foley (2000) identified influential groups with serious interests in evidence and that have contributed notably to the development of research policy on trials in which places, organizations, and other charters are the target of randomization.

Places That Are Randomized: Theory and the Units of Randomization

The rudimentary theory underlying all of these *Annals* articles is that one must change entire places to enhance the well-being of people in those places or to decrease the risks that confront them. The "places" vary considerably in character.

In the health care sector, for instance, Grimshaw and his colleagues randomly allocated medical family practices to different interventions. Leviton and Horbar randomly allocated willing and eligible tertiary care facilities and neonatal intensive care units within hospitals to different regiments to learn whether better practices could be introduced and whether the practices would have a discernible effect on outcomes of patients.

Sikkema and her colleagues randomly allocated eligible and willing housing developments to different regimens to learn whether a particular opinion leader-based intervention, tested earlier in other trials and sustained by coherent theory, produced a detectable and substantial effect on women's health behavior. Although the particular focus was AIDS, the work has implications for other arenas.

In the Jobs-Plus trials, Bloom and Riccio and their colleagues also randomized housing developments but for a different purpose: understanding the effects of this form of increasing social capital on wage and employment rates, and other outcomes. Sikkema depends on a long history of research on the effects of diffusion strategies, including opinion leader approaches. Bloom and Riccio depend more on what could be loosely construed as a theory of developing human capital. A theory of diffusion of innovation and change is implicit, rather than explicit, in their work.

Flay appears to be the first social scientist to have succeeded in executing a well-designed trial that involved randomizing entire schools to different health risk reduction interventions. This was in Waterloo, Canada. He and his colleagues in

the United States and Canada have depended on and have advanced the theory of change to understand how to develop the interventions and to test a theory that posits that changing individual behavior depends on changing school conditions including group processes within the school. Cook's article in this volume depends on similar theory that is implicit. Porter et al. provide more concrete information on training the trainer approach in trails on a teacher professional development program.

Weisburd's study focused on high crime geographic areas within cities—hot spots—as opposed to specific institutions such as housing developments or schools or hospital units. The broad theory posits that redistributing police resources to hot spots will reduce crime. This is in counterpoint to a theory that says it will not have an effect and, furthermore, that focusing police resources on hot spots will lead to the migration of criminal activity to neighboring areas.

Entire villages and other geopolitical jurisdictions were randomly assigned to different interventions in the studies in Mexico that Parker and Teruel describe and a complex study in China that Smith describes. In Mexico's Progresa trial, the theory was that changing village behavior, notably through financial incentives that would increase the rate at which children stayed in schools rather than working in the fields, would lead to better outcomes at the village level on account of local economies. In the fertility control trials in China that Smith describes, the implicit theory is that people will change if given the opportunity to do so but that the change depends on local circumstances and place-based influences.

Each of these articles educates us. Each invites further questions. For instance, how can we develop better theory about what should happen at various levels of the units being considered in a place randomized trial: province or county, and city or village within province, and institution or housing development within village or city, and catchment area or hot spot within city? Each article also invites a question about how to learn more easily about other trials of this sort, involving yet other units of allocation and analysis.

Relationships and Agreements

How have people gotten place randomized trials off the ground? Part of the answer to this question hinges on the development of agreements between the trialist's team and the prospective participants in the place based trial. "Participants" here mean individuals or groups whose cooperation is essential in deploying both the intervention and the trial.

In their article on prevention research trials, for example, Flay and Collins emphasize written agreements signed by people who are authorized to represent their organizations' interest in testing ways to improve. In their case, this includes the school principal and the school superintendent. Given the mobility (turnover) of school administrators and teachers in some parts of the United States, at least, one or the other signatory might disappear before the end of the trial. Having both

kinds of school administrators as signatories to an agreement helps reduce the obstacles to running a fair trial to completion. Elsewhere, Robert Slavin and his colleagues take a similar but different tack in their tests of Success for All. They demand that a large fraction of teachers in a school agree on testing his program at a point in time. For Kellam and Van Horn (1997), developing long-term relationships is essential to generating evidence over long periods of time. Porter et al. in this volume remind us about the practical matter that agreements are essential but substantial changes in the places (in leadership for instance) may abrogate agreements and that this prospect has to be considered in such trials.

[O]ne must change entire places to enhance the well-being of people in those places or to decrease the risks that confront them.

In some areas of the health sector, agreements have been able to depend on an overarching theme and organization. Leviton and Horbar's work in this volume of the *Annals*, for instance, depends partly on the Vermont Oxford Network. Health care institutions commit in advance to be willing to engage in research, including perhaps randomized trials at times in the interest of contributing to the accumulation of knowledge about what works. At least part of the Grimshaw et al. effort described here also depends on formal networks of people and institutions whose interest lies in collaborating on studies that help us understand what works better.

The Bloom and Riccio trials on Jobs-Plus depended on identifying entities and people who wanted to develop a relationship in the interest of a place randomized trial and for reasons that were more important. The process of identifying prospective partners, inviting proposals, and reaching agreement on willingness and capacity to participate that the authors describe is instructive. The relationships were formalized through written agreements at the level of agencies within city, such as departments responsible for public housing, as well as with federal and other entities. Very complex, and requiring time and industry.

Weisburd emphasizes the need to develop personal relationships that lead to trust and willingness to experiment on innovations that may work better than conventional practice. His description of the Jersey City effort and others is refreshing.

Relationships involve reputation, of course, a reputation that must be developed at different levels to get a place randomized trial off the ground. The principal authors of each of the articles in this *Annals* volume, and the institutions to which they belong, are graced with high reputations. The reputations reflect authors' high productivity, stamina, and willingness to recognize their own ignorance and to

reduce their own and others' ignorance. Their reputations are high and, more important, give them access to relationships that permit good research to be done.

Relationships and agreements have been important in mounting high-quality trials. These authors' handling of the topic invites attention to questions for the future: How do we develop better contracts and agreements with organizations, public and private, that permit us to generate better evidence about the effects of an innovation? And how do we make the specific contents of agreements and changes in these agreements more accessible to other trialists and the people who might collaborate with trialists so that they can learn and contribute?

Incentives and Justifications

For Weisburd, an expert in criminological research, an important condition for mounting a randomized trial on police patrolling crime hot spots is that conventional practice, policy, or program was under attack. Cops are local theorists, and they disagree about what might work. Crime experts have also disagreed about what might work in high-crime areas. Weisburd's position accords with ethics of trials in the medical arena. Neither the Salk vaccine trials on polio nor trials on streptomycin for treating tuberculosis would have been mounted had the purported effects of conventional treatments not been suspect and under attack, for instance.

For Leviton and Horbar, the "quality chasm" between what constitutes good health care based on dependable evidence and contemporary medical procedures in hospital units constitutes the justification for running a trial. Grimshaw and colleagues' work in medical practices is similarly justified.

In considering the prospects of the Progresa/Oppotunidad income support program in Mexico that Parker and Teruel describe, economists disagreed about whether an incentive program such as Progresa would keep Mexican children away from working in the agricultural fields and increase the likelihood that they would stay in schools. The disagreements were important because they made plain the uncertainty. A place randomized trial was mounted to appreciably reduce this uncertainty.

The United States has a long history of attempting to reduce poverty, much of which depends on community-based efforts. To judge from the Jobs-Plus effort described by Bloom and Riccio, the incentive for people in local departments of housing and other government agencies to participate in the trial include their interest in reducing the problem.

In trials of a teacher development program that Porter et al. describe here, justifications for the research and development effort lay in U.S. interests in improving mathematics and science education in the middle-school grades, especially in large urban school districts. Justifications lay also in the fact that, prior to the Porter et al. work, no sizeable controlled trials on the effects of any professional development had ever been run.

The scientific justification for such trials is of course the assurance that, if the trial is carried out properly, (1) there are no systematic differences between groups

of places randomized and (2) a legitimate statistical statement of one's confidence in the results can be made. This justification holds for all the articles in this volume. Furthermore, as Weisburd, Grimshaw et al., and Smith point out, the simplicity and transparency of the idea of fair comparison through a randomized trial has strong appeal for policy people and decision makers who cannot understand and do not trust complex model-based analyses of data from nonrandomized studies.

Empirical evidence on the vulnerability of nonrandomized trials in comparison to the strength of randomized trials has been building since at least the 1950s. None of the articles in this *Annals* volume refer directly to this body of literature simply because the articles have different aims. See Boruch (2005) for a brief list of other studies on such comparisons and examples that range from pigs to astronomy and econometrics. Virtually all such methodological studies involve data from randomized trials in which individuals rather than places are the units of random allocation and analysis and related quasi-experiments. The biases in estimating an intervention's effect based on the quasi-experiments can be very large. Evidence on whether and by how much estimates of the effect of a nonrandomized trial differ from those of a randomized trial, when individuals are the units of random assignment and analysis (as in most medical trials), are important.

It remains to be seen whether similar methodological studies on aggregate-level analyses (cluster, place, group) yield similar results, that is, serious biases, but it is reasonable to expect biases here also. Bertrand, Duflo, and Mullainathan (2002), for instance, focused on biases in estimates of the standard error of effects assuming no effect at all using "differences in differences" methods that are conventional in some economic analyses. They found Type I error rates that were nine times the error rate presumed (.05) in using conventional statistical tests. This was partly on account of serial correlation. More methodological research, however, needs to be done on the quasi-experimental approaches to estimating effects at the place (aggregate) level to understand when the biases in estimates of effect appear, when the biases in estimates of their standard errors appear, and how large these are, relative to estimates based on place randomized trials.

The incentives and justifications that are identified in these articles are important in the near term. In the long term, it would be good to understand what the incentives are and to make these explicit at different levels and in different areas: policy (crime, health, education, and social services), institution (government and nongovernment agencies), and individual service providers.

Deploying the Intervention: Implementation, Dimensionalization, and Measurement

Incentives are essential for assuring that places, and the influential people in them, are willing to participate in a randomized trial. Understanding how to deploy a new program or practice in each place requires more than willingness, of course. It also demands experience.

Achieving this understanding is no easy matter, one that is apart from the challenge of executing the randomized trial in which the places are embedded. Jobs Plus, for instance, required research teams to engage and guide coordination of the local housing authority, welfare department, workforce development agency, and

In the long term, it would be good to understand what the incentives are and to make these explicit at different levels.

public housing residents within each site. The challenge of getting these agencies to work together more closely than any had in the past was piled on top of other challenges to implementation including getting housing development residents involved as full partners despite inexperience and distrust, integrating services across providers (to meet housing needs and encourage employment in complex welfare environments), and generating job search and acquisition processes that fit the place.

For school-based trials, the implementation challenges are different insofar as school bureaucracies differ from housing department bureaucracies, labor departments, and so on. Nonetheless, consider the way that Porter et al. dimensionalize one class of implementation. They do so by developing indices of people's participation in meetings and teams that were major ingredients for change. For the DEC (Data on Enacted Curriculum) trials, one such index set includes counting training sessions in which at least one team member participated and computing the average number of participants per session, the proportion of sessions to which local leaders (school principals) contributed, and the consistency of people's participation over time. This also included qualitative reconnaissance on factors that impede participation, such as the limited time that schools allow for teachers to meet during work hours.

The challenges to Sikkema, in deploying an AIDS risk reduction program to women in U.S. housing developments, is similar to others in some respects but differs in others. She and her team developed three different intervention models prior to the trial and, using focus groups, reconnoitered the virtues and vulnerabilities of each model relative to local standards of acceptability. It involved identifying opinion leader cadres, selecting them, and providing workshops to support them. Sikkema, as others, relies on basic count data to index level of participation in workshops and community events. As in the Jobs-Plus effort, learning how to negotiate agreements with different influential entities in housing developments, and documenting this, was an essential part of implementation. More interesting, Sikkema

relies on explicit and tentative theory of change, as does Flay. The theory is implicit, rather than explicit, in articles by other authors in this *Annals* volume.

All of the authors rely to a greater or lesser extent on nonnumerical information to understand the deployment of the programs being tested, as well as on numerical information. Quarrels between the qualitative and quantitative tribes have no substantial role here, partly because place-based trials cannot be black box studies. Indeed, Bloom and his colleagues have won an award from the Association for Policy Analysis and Management for the Jobs-Plus trials partly because of the interesting coupling of methods for research.

These articles are informative but necessarily brief about deploying programs. The brevity invites broad questions about how the authors' experience *in detail* can be shared with others, for example, Web-based journals, or reports without page limits, workshops, and so on. It invites more scientific attention to the question of how one can dimensionalize the implementation of programs across different sectors and within them. They invite attention to the engineering questions of how to measure implementation level better and how to enhance it in different places.

Technical Vernacular and Technical Resources

Here, we use the phrase "place randomized trial" because it seems easy for some readers, scientists and public administrators and otherwise, to understand. The vernacular in the statistical and social science literature that is related includes phrases such as "cluster randomized trials," "group randomized trials," "saturation trials," and others.

The important technical references include Donner and Klar (2001) on cluster randomized trials. This is a fine and blessedly brief book that considers technical approaches and examples from different countries, mainly in the health arena. The important references include Murray's (1998) book on group randomized trials, which is more detailed and considers examples from the United States. Bloom's (forthcoming) monograph is excellent on account of its clarity and use of examples in the United States. Bryk and Raudenbush's (1992) text directs detailed attention to analysis of multilevel statistical data that may or may not have been generated by cluster/group/place randomized trials.

Following the Campbell Collaboration conferences on place randomized trials in 2001 and 2002, the William T. Grant Foundation initiated efforts to enhance the technical capacity of researchers to design such trials and to analyze results. Raudenbush and his colleagues have developed software for estimating the statistical power of a trial under various assumptions. This is given at the foundation's Web site, http://wtgrantfdn.org. The foundation also supported a pilot consulting service, through its Web site on design of the trials, and convened a special institute on the topic at the University of Michigan. Bloom, a contributor to this volume, and Raudenbush, a participant in one of the Campbell conferences on the topic, co-led the institute.

A big technical issue in design of the trial includes ensuring that the size of the sample of places or institutions that are randomized is large enough to permit one to detect relative effects of interventions with confidence and that other features of the design, especially covariates, can be exploited to increase statistical power. Bloom and Riccio, Leviton and Horbar, and Grimshaw et al. consider the matter of sample size briefly in this *Annals* volume. See the references cited earlier on technical issues, and see Boruch, Wortman, and Cordray (1981) for another handling of the topic. Bloom and Riccio advance the state of the art in one important respect: coupling conventional comparison of randomized places with time series analyses of data from the places. This approach permits one to also take into account the changing composition of individuals within the places (housing developments).

How Do People Find Out about Trials? International Registers of Place Randomized Trials, Standards of Reporting, and Other Resources

Until recently, people could go to no single source to learn about place (cluster, group) randomized trials that have been completed or that are under way. The authors of studies in this *Annals* volume, for instance, cover undertaken work in China, Canada, England, the United States, and Mexico. It is work that is reported piecemeal, in diverse scientific journals, Web sites, and limited circulation reports.

This diversity in publication venues makes searching the literature difficult. For instance, relying on Web-based searches such as ERIC and PsychInfo is imprudent, at best, relative to hand searches of the research literature, at least in regard to locating randomized trials. The industrious hand searcher can uncover three times as many randomized trials in a full-text search relative to depending on current search engines (Turner et al. 2003).

The difficulty in identifying randomized trials has been reduced with the creation of the international Cochrane Collaboration in health care in 1993 and the international Campbell Collaboration in 2000. Both organizations, which rely heavily on voluntary participation, have developed registers on randomized trials and possibly randomized trials that can be accessed through the relevant Web sites. Both rely heavily on hand searches (full-text readings) of peer-reviewed journals rather than on conventional machine-based searches.

For instance, the Campbell Collaboration Social, Psychological, Educational, and Criminological Trials Register (C-SPECTR) contains about thirteen thousand entries on randomized and possibly randomized trials, and includes more than two hundred references to place randomized trials. They are coded as "CRT" in indexed or nonindexed fields in the register, so searching for them is easy. Most references are supported by abstracts, many of which are standardized. In C2 SPECTR, one can find entries on place randomized trials involving schools and classrooms in Kenya, El Salvador, and India; brothels in Thailand; factories in Rus-

sia; barrio segments in Colombia; neighborhoods areas in Taiwan; and bars in England and Canada, among others.

Standards for reporting on randomized trials that involve individuals as the unit of random allocation and analysis are a relatively recent initiative. In the health arena for instance, the CONSORT statement gives guidance on the ingredients that a good report on a randomized trial in the health arena should contain (Moher et al. 2001). The CONSORT statement has been extended to include standards for reporting on place/group/cluster randomized trials. This extension, given in Campbell, Elbourne, and Altman (2004), depends on the ingredients of other kinds of trials, of course, but pays special attention to important topics embedded in a place randomized trial. This includes, for instance, consideration of the rationale for this research design and the fact that there are at least two levels of inference: the cluster (place) level and the individual (person within place) level. There are usually two levels of sampling and sample sizes associated with each level, different levels of attrition and measurement, and so on.

Most contemporary standards for reporting include no requirement that microrecords from the trials be made available for secondary analysis. Such reanalysis is often essential to verify the appropriateness and correctness of original analyses of the data, to test new hypotheses or explore new relationships, to entertain different assumptions about the structure underlying the data, and for other reasons (Boruch, Wortman, and Cordray 1981). As the state of the art in running place randomized trials advances, and as it becomes obvious that resultant data are a valuable resource and deserve secondary analysis, the standards for reporting trials are likely to change. Furthermore, readier Internet access to the data generated in such trials is likely to lead to changes in standards for reporting meta-analyses (Moher et al. 1999) as well as to improvements in the conduct of the trials and to productive exploitation of the resultant data.

Organizations that attempt to evaluate the published statistical evidence on relative effects of large-scale interventions, and that publish the results of their reviews, are also a resource. Many, not all, put randomized trials, including place randomized trials, at a high priority. These organizations identify the trials, review them, and make results accessible through the Internet and in other ways. The Campbell and Cochrane Collaborations illustrate international efforts. In the United States, the Institute of Education Sciences' What Works Clearinghouse focuses heavily on trials in education (2005, http://whatworks.ed.gov), as does the Office of Justice Programs for drug abuse and violence prevention (http://www.colorado.edu/cspv/blueprints). Federal efforts to bring uniformity to standards of evidence and that attend to the trustworthiness of randomized trials are reflected in the products of the Federal Collaboration on What Works (2005) and the Agency for Healthcare Research and Quality (2005, http://ahrq.gov), among others. The Coalition for Evidence Based Policy (2005, http://www.excelgov.org) provides routine updates on such efforts and has been a contributor to them. None thus far attends to or depends on access to microrecords for reanalysis of data from trials. But this may change for reasons given earlier.

Concluding Remarks

The shift in science policy in many countries, toward fair comparisons and toward randomized trials on the effects of an intervention, has had a choppy history at times. Progress has been relatively steady in the health sector. This is partly on account of precedents such as the Salk vaccine and streptomycin trials.

In the social sector, progress has been choppier and less obvious. But the progress depends on some of the same ingredients as the health sector. Good precedents are also effective, for instance. The Tennessee Class Size Trial in the United States, the Waterloo Trial in Canada, Progresa in Mexico, and the police hot spots trials in Jersey City were remarkable in setting precedent.

One of the challenges to progress lies in the need to enhance people's capacity to learn about place randomized trials and how to do them. This issue of the *Annals* was developed with Rockefeller Foundation support and under Campbell Collaboration auspices to meet part of the challenge. The authors of articles in this *Annals* issue are experienced and, more important, have been able to mount such trials with reasonable success. The augmentation and public access to information on place (cluster) randomized trials in the Campbell Collaboration's register is another important product.

Evidence-based policy requires resources. The resources required for place randomized trials described here were supplied, at times, by private foundations. Some foundations have assisted notably or have served as the catalyst in running place randomized trials. The larger and sustained resources, of course, come from countries in which government and the people it represents are interested in evidence about what works, what does not, what is inconclusive, and what is harmful. When the countries, or their governments, are unwilling or unable to care about work issues, we will make no progress. When the countries and the government do care, we will enhance people's lives.

References

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

Bloom, H. Forthcoming. *Learning more from social experiments: Evolving analytic approaches*. New York: Russell Sage Foundation.

Boruch, R. F. 2005. Comments on the papers by Rawlings and Duflo and Kremer. In *Evaluating development effectiveness*, ed. G. K. Pitman, O. N. Feinstein, and G. K. Ingram 232-39. New Brunswick, NJ: Transaction Books.

Boruch, R. F., and E. Foley. 2000. The Honestly Experimental Society: Sites and other entities as the units of allocation and analysis in randomized trials. In *Validity and social experimentation: Donald Campbell's legacy*, ed. L. Bickman 193-238. Thousand Oaks, CA: Sage.

Boruch, R. F., P. M. Wortman, and D. Cordray, eds. 1981. *Reanalyzing program evaluations: Policies and practices for secondary analysis of social and educational programs.* San Francisco: Jossey-Bass.

Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical linear models.* Newbury Park, CA: Sage.

Campbell, M. J., A. Donner, and D. Elbourne, eds. 2001. Design and analysis of cluster randomized trials. *Statistics in Medicine: Special Issue* 20 (1): 329-496.

Campbell, M. K., D. R. Elbourne, and D. G. Altman. 2004. CONSORT statement: Extension to cluster randomized trials. *BMJ* 328:702-8.

Davies P. 2004. Evidence-based policy. The Jerry Lee Lecture, annual colloquium of the Campbell Collaboration, Washington DC.

Donner, A., and N. Klar. 2001. *Design and analysis of cluster randomized trials in health research*. London: Arnold.

Federal Collaboration on What Works. 2005. *Draft report of the working group: The what works repository*. Washington DC: Office of Justice Programs.

Kellem, S. G., and Y. V. Van Horn. 1997. Life course development, community epidemiology, and preventive trial. *American Journal of Community Psychology.* 25:177-87.

Moher, D., D. J. Cook, S. Eastwood, I. Olkin, D. Rennie, and D. Stroup, for the QUORUM Group. 1999. Improving the quality of reports of meta-analysis of randomized controlled trials: The quorum statement. *Lancet* 354:1896-1900.

Moher, D., K. F. Shulz, M. Egger, F. Davidoff, D. Elbourne. 2001. The CONSORT statement: Revised recommendations for improving the quality of reports on parallel group randomized trials. *Lancet* 357:1191-2004.

Murray, D. M. 1998. *Design and analysis of group randomized trials.* New York: Oxford University Press.

Pitman, G. K., O. N. Feinstein, and G. K. Ingram, eds. 2005. *Evaluating development effectiveness*. Vol. 7 of World Bank Series on Evaluation and Development. New Brunswick, NJ: Transaction Books.

Rawlings, L. 2005. Operational reflections on evaluating development programs. In *Evaluating development effectiveness*, ed. G. K. Pitman, O. N. Feinstein, and G. K. Ingram 193-204. New Brunswick, NJ: Transaction Books.

Turner, H., R. Boruch, A. Petrosino, J. Lavenberg, D. de Moya, and H. Rothstein. 2003. Populating an international web-based randomized trials register in the social, behavioral, criminological, and education sciences. *Annals of the American Academy of Political and Social Science* 589:203-25.

Wellford, C. F., J. V. Pepper, and C. V. Petrie, eds. 2005. *Firearms and violence: A critical review*. Washington, DC: National Research Council.