J Exp Criminol (2012) 8:331–341
DOI 10.1007/s11292-012-9154-z

Deploying randomized field experiments in the service of evidence-based crime policy

Robert Boruch

Published online: 22 May 2012

Keywords Randomized controlled trials · Management · Operations · Coalitions · Registries

Introduction

In Samuel Johnson's *Idler*, the zealous pundit declared that "Criticism is a study by which men grow more important and formidable at very small expense." In what follows, I abide by the implied counsel, recognizing of course that Johnson was not himself a shy or continent critic.

Rather, the aim here is to recognize the industry and ideas represented in a series of papers on randomized controlled trials presented at the 11th Annual Jerry Lee Symposium on Crime Prevention. The focus, as in the papers, is on the deployment of the trials in the field. My intent is also to enlarge on the lessons that are educed by the papers' authors.

The theme and its pedigree

Current academic journals direct little serious attention to how randomized controlled trials are managed in the field and how course corrections are made when inclement weather—unexpected difficulty—is met. In fact, articles published in such journals are usually short on reportage about field operations, notably in failing to characterize the obstacles, mis-steps, and mishaps in attempts to produce dependable evidence in randomized trials and in other kinds of studies. This lacuna is despite the fact that standards for design of the trial and for reporting on its results are nowadays well articulated. The papers in this edition of the *Journal of Experimental Criminology* help remarkably to advance our understanding of how to deploy trials and the obstacles we may encounter.

This paper was developed in reaction to presentations at the 11th Annual Jerry Lee Crime Prevention Symposium, University of Maryland, May 2, 2011. It was an honor to comment and a delight to learn from the work.

R. Boruch (✉)
University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104, USA
e-mail: robertb@gse.upenn.edu

It is nonetheless possible to identify a few precedents for the theme, if not a clear pedigree. For instance, a 2005 issue of the *Annals of the American Academy of Political and Social Sciences* covered operations in place-randomized trials in education, health, housing, and criminology. Weisburd's (2005) contribution to the volume, on police hot spots trials, is exemplary in describing the handling of ethics (cops' and researchers'), systemic stability and instability, understanding of interventions in each arm of the trial, and time, timing, and staging, among other operational issues.

The field researcher who is attentive to contemporary history can occasionally find reports that handle such matters. In welfare economics, for instance, authors of reports on the Seattle and Denver Income Maintenance Experiments (SIME/DIME) and New Jersey Negative Income Tax Experiment did indeed describe how those massive trials were managed (Christophersen 1983; Kershaw and Fair 1976). One of the lessons was this: don't rely on Princeton graduate students in economics to explain tax structure to low income families. In related work, Riecken et al (1974) dedicated serious attention to management of randomized experiments in one chapter, but few readers appear to have paid attention to it. Boruch (1997) and contributors to Boruch and Wothke (1985) built on the Riecken volume mainly with respect to the challenge of actualizing random assignment. More recently, Appendix 2 of Sherman (1992), written by Sherman, cops and field operations people, is a neat example of reporting. In the social services sector, Alexander and Solomon (2006) asked researchers to write about the innards and soft underbelly of conducting scientific field research, including controlled trials, in difficult settings. Some did.

There are new and important things to be learned, however, from the authors of papers in this edition of the *Journal of Experimental Criminology.* Papers of this sort are, as yet, rare and valuable.

Environmental stability and instability

The institutional settings in which crime prevention trials are deployed are at once stable, meaning that there is plenty of inertia, and also unstable. That is, changes in the settings occur regardless of the experiment. One might call this "ambient instability" or "churn." Construe the phrase or word as including environmental, contextual, institutional, and people changes that may or may not be predictable and that affect the conduct of an effort to understand when and how to conduct a controlled randomized trial.

Strang's (2012) trial on restorative justice, for instance, was embedded in a system in which four police chiefs were emplaced during the five-year course of the trial. MacKenzie (2012) tells us that incarcerated control group members did not receive expected services in her boot camp trial, partly on account of wait-lists and other "structural factors. " Kilburn (2012) reports a litany of surprises in her trial on a home visit service program, including misperceptions and apperceptions of lottery-based allocation and unexpected declinations of offers of the new service that was targeted in her study. This is despite her jurisdiction's frequent use of such lotteries in different arenas (selection of children for preschool, for instance) and people's willingness to avail themselves of or to refer clients to the ostensibly desirable services in those arenas.

Cook et al. (2012) were surprised by their correction system's failure to have the promised services ready, leading to a year's delay in the randomized trial, by operational interruptions attributable to personnel leave and changes, and by unexpected constrictions in the flow of clients to the trial. On top of this, measurement difficulty emerged in that social security numbers, used to enable linkage of research records on participants with the participants' earnings records, were not available for some prisoners. A worldwide recession, anticipated by few mortals, reduced job openings, a nontrivial matter for the reentry program whose effects Cook et al. wanted to estimate.

In Farabee et al.'s (2012) study, demand for the intervention services targeted in the trial were declined "for unknown reasons…" And access to the services in the intervention group was constrained by delays in renovating a building, necessitating the delivery of service at a not easily accessible site.

Talk about the slings and arrows of outrageous fortune. It may be of some small comfort to these authors that they are not alone in being so afflicted. The history of welfare experiments cited earlier, is a case in point. More recently, education experimenters have encountered school systems in which the ambient instability in people's positions (teachers changing jobs and courses, leaving etc.) is over 45 % from one year to the next (Porter et al. 2011). Some welfare to work and training experiments have had to confront unexpected local recessions, factory close-downs, and so on, all of which put ceilings on any possible effects that employment programs could have on post program wage rates. Randomized trials in developing countries have been aborted, or delayed, or modified substantially on account of macro-level changes in government (shifting regimes and revolutions can be distracting), environment (think tsunami), and for other reasons.

Let us put down as a first lesson this. Unexpected changes in the systems in which controlled randomized trials are deployed are inevitable. Further, the unexpected must be expected, even if the specific nature of destabilization or instability cannot be predicted. In fact, the papers here give some nice illustrations of how this abstract lesson has been handled.

Lesson two, as Strang and Kilburn (this issue). say and that MacKenzie (2012) and Hawken (2012) imply, is that someone must be in place in the field so as to monitor and adjust to the unexpected, and that this person has to have close ties to other people in the field. Indeed, the Jerry Lee Symposium discussion revealed that Lawrence Sherman, who serves on the Board of Trustees of the sponsoring organization (the Smith Richardson Foundation), exercised remarkable prescience by arguing for funding to support such a field person to attend to local operations in each trial even though the grant applicants themselves had not requested such funding.

Pipelines, reconnaissance, and training wheels

Expecting the unexpected on account of ambient instability, the churn, and capitalizing on what can reasonably be expected, can be planned at least up to a point. That point's location is usually ambiguous. But we may rest assured, so to speak, that it is out there somewhere.

Pipelines and reconnaissance

Most of the authors of the papers reviewed here are admirably conscientious in reporting on the numerical flow of clients through their field trial's system. As Strang, and Hawken (this issue), recognize, for instance, doing so is part of good reporting practices as suggested by CONSORT guidelines for experimentalists in the health sector. It is also worth recognizing, as Sherman et al. (1992) did during the 1990's Spouse Assault Replication Project, that monitoring and frequent reporting on the pipeline was essential for management, i.e. tracking the sample's engagement and buildup and detecting problems in these processes. Roman et al. (2012) did so also in mapping the pipeline description onto steps in the screening process and giving reasons for leakage. Their recommendation that funders receive quarterly reports on this seems sensible.

Understanding whether and how to enlarge on the idea is important. For instance, it is sensible and sometimes feasible to extend the idea of pipeline or flow charts to periods prior to the trial's initiation. In particular, all authors did reconnaissance to anticipate the number of individuals who might be available for random assignment to the trial. This due diligence can be amplified for eager experimentalists by laying out the sources and character of numerical evidence on people most likely to be eligible prior to the trial. Further, it is conceivable that cousins of such data, on the mobility or turnover or instability among staff who are relevant to the control condition and to the not yet launched intervention. might be laid out in advance as part of the reconnaissance and management. Recall the earlier remarks on ambient instability.

Embedded in the experience and lessons of the authors of these papers is a reminder of two aphorisms. The first aphorism is this. "As soon as the grant is awarded and the contract is signed, the size of the available target population drops in half." The point is that one will encounter unforeseeable problems in the take-up/recruitment of trial participants and that the experimenter needs to do his or her best to anticipate this through early reconnaissance, run in trials, and so on. Moreover, the members of the collation involved in the trial may vastly overestimate the number of cases that would be eleigible for the trial. The second pertinent aphorism, enunciated by a biostatistician and colleague, Lincoln Moses, in a personal communication to this author in 1985. was as follows. "Do the statistical power analysis and then get as many as you can." Regardless of the limits to this simple but wise counsel, how can anyone take serious issue with a person named Lincoln Moses?

Pipeline data acquired prior to the trial can also help in searching for and exploiting covariates that would increase the statistical power of a trial. Cook et al., for example, remind us of the import of covariates in increasing one's capacity to detect and intervention's effects when the experiment's sample size is modest or is reduced by factors beyond the experimentalists' control. Hawken, for example, regrets that she did not get relevant baseline data early in her trial. The opportunity to access dependable administrative records on the participants in crime prevention trials prior to, during, and after a trial is invaluable. Experimentalists in other social sectors often do not have this advantage. More important, the opportunity to combine short time series data with the conventional analysis of variance of outcomes of a trial, so as to enhance statistical power and put the trial into a larger context, is tantalizing. Further, Roman et al. (2012) suggest building a quasi-experimental design alongside the experimental design as a fall back option, another tantalizing idea.

Training wheels

Kilburn engaged in two pilot periods prior to the formal initiation of the trial on the home visitor program that was being tested. In particular, she and her colleagues appear to have waited for the new service delivery, the intervention, to stabilize. and then vetted the research design during a different but overlapping periods. Then, they proceeded with the trial. Others, such as Strang, capitalized on independent experience in mounting similar trials in other places, replication being an astonishing feat in itself, to do roughly the same thing. In a different arena, engineering, this pre-experiment period is often called a "run-stage" by engineers who try out different approaches to controlling and testing purported improvements to a complex system.

In the experimenters' vernacular, this strategy could be labeled as a two-stage or two-cohort randomized trial. Call the first, and relatively brief, stage Alpha. In this stage, the experiment's design and the new intervention to be tested in deployed in situ. The expectation in Alpha is that the exercise will encounter difficulties that could not be anticipated at the outset. One makes all the mistakes one can make during Alpha, recognizing the outcome data that is collected may be useless, or at least ambiguous, in estimating the intervention's effects. New kinds of information, collected as part of Alpha, will inform the Beta stage. Alpha efforts are the training wheels.

The Beta stage capitalizes on the lessons learned in the Alpha stage, so as to forego, dodge, or anticipate and plan how to handle unhappy surprises. This Beta stage would typically involve large samples and longer follow-up periods in the trial. This Beta stage is the one that produces dependable evidence on the intervention's effects, at least in the sites in which the trial is undertaken well.

The general idea has early origins beyond engineering and the papers presented here. The Sherman et al. (1992) randomized experiments on mandatory arrest for misdemeanor domestic violence, for instance, employed the idea, as did others involved in the Spouse Assault Replication Project. In education and more recently, Porter et al. (2011) used a two-cohort approach in cluster randomized trials on purported improvements to science education in nearly 200 schools to good effect.

One of the larger implications of the strategy pertains to foundation and agency funding of such trials in the U.S. and other countries. For instance, some government *contracts* for controlled trials on an intervention, in education for instance, include a provision for reconnaissance and a run in stage. Few private foundation or federal *grants* appear to do so. Grant applicants, possessed perhaps of inordinate optimism (and who among us is not without guilt on this account?), do not ask for the time and money required.

The Alpha-Beta strategy has merit for the funder and the experimentalist when the need to test the intervention in a randomized field trial is clear but environmental factors that may affect the trial are not. This scenario is common in work in developing countries, for instance. The unknowns are daunting. It is not uncommon either in developed countries where local variations in ambient stability can be substantial.

Coalitions, partnerships, and other relations

The idea of engaging "partners" in the design and execution of a randomized trial has served well in developing and testing early childhood programs for at risk children

(Fantuzzo et al. 2006). In this sector, warm and fuzzy is combined with scientific rigor. This pleasing conflation is exercised at and tailored to local levels over time, and builds trust into the idea of partnerships. Actualizing the idea in trials on early childhood programs is not uniform, of course.

The idea of "coalitions" that Strang introduces in her report is arguably more appropriate in the argumentative contexts of crime prevention, enforcement of law, administration of justice, and rehabilitation. The MOUs that she describes undergird tentative agreement of parties whose aims can differ appreciably. These are in the presumed common interest of generating better evidence on how to do better.

As Strang implies, however, the bigger idea is building social capital among the people involved in the trial which itself is a social system. Hawken's emphasis on frequent meetings and newsletters to service the interests of judges and probation officers further illustrates the idea. Kilburn also got deep in these weeds. Implicit in their papers is the message: nothing happens without a tentative trust among the players, and it takes time, effort, integrity, and good people to develop that trust. This gets well beyond MOUs and other more or less formal agreements.

Of course, this trust and the development of social capital depend on other things. In particular, incentives and their sustainability are important, at times. How all this can be worked into a theory about how the intervention is supposed to work with what targets in what complex system of delivery of service and in what context is not clear. Taxman, nonetheless, reminds us that forms of theory in this sector are worth considering seriously.

Much of what these authors confronted is in contrast to what Cook et al. had to do in their trial. The difference between the Strang, Kilburn, and Hawken scenarios versus the Cook et al. scenario begs the question of how independent, how "sequestered" in Kilburn's words, should the experimenter be from the other operations in deploying the intervention to be tested? For instance, Cook et al.'s reentry trial involves an arrangement in which the experimentalists have little substantial contact with service providers. Their duty lay mainly with acquisition of relevant records and competent data analysis. To their credit, they paid attention to other matters. On account of this kind of independence, however, they did not have to develop much trust or social capital beyond the ambit of their agreement with high level authorities.

There are parallels to this difference in approaches, and examples of alternative arrangements, in other sectors. Jeanne Poduska (2011) and her colleagues, for instance, have built three generations of high quality randomized trials in Baltimore, Maryland, schools and are attempting the same in Houston, Texas. Theirs has been a Sol Alinsky orientation to engaging the community so as to build and maintain relationships that permit good experiments and produce dependable incremental advances in knowledge. This is in a spirit that is parallel to Strang's, though her effort crosses more jurisdictions.

All the experimenters whose papers are reviewed here refer directly or indirectly to operational changes in practice that the new intervention being tested in the experiment engenders. Hawken, for instance, discusses the fact that judges had to change the way they did business and probation officers had to take on increased duties. Roman et al. (2012) emphasize the difficulty of changing processes to identify people who were eligible for the reentry program they sought to test. Many aspiring experimentalists are likely to value such information and to having more precise information on the work load increase or the burden of change in practice. How to properly characterize the work-loads in this context is not clear. Simply toting up increased hours is one

way. Trying to capture in words the level of change in practice is another. The topic carries implications for scaling up and for replicating the intervention in other sites. Insofar as the increased burdens cost money, the topic has implications for policy and cost-effectiveness analyses.

Equipoise is not for everybody

Each of the trials reported here passed muster with Institutional Review Boards (IRBs) which are responsible for vetting the ethical propriety of research. In particular, each of the trials appears to meet a basic ethical standard for its mounting, notably equipoise. In medicine, the standard of clinical equipoise can be defined as a "state of honest, professional disagreement in the community of expert practitioners about the relative merits of the treatment alternatives" (Binik et al. 2011). A similar standard applies in the crime prevention sector though it may not be expressed this way. Instead, it might be expressed more plainly among experts in a statement: "We do not know whether A will work any better than B."

The equipoise standard is usually based on research experts' judgments. How non-researchers view this standard is another matter. Indeed, Kilburn's study illustrates how service providers may implicitly disagree with the intervention alternatives that are offered to clients, and instead believe one or the other is the intervention of choice. Similarly, in Strang's experiments, the judges or cops may favor one intervention over another. Further, in the Kilburn case, they may then pile on another justification, saying that randomization to the control condition, rather than to the new intervention deemed desirable would be harmful to the client. Their speculation is that a particular random assignment will degrade the morale of the participants assigned or result in a client's disappointment that is unacceptable in the view of the service provider.

That people who might be involved in randomized trials will, at times, favor one intervention over another before the trial is a fact of life. For them, there is no equipoise. And if they have the discretion to do so, they will exercise it in the interest of dodging random assignment or subverting it. The phenomenon has an interesting history, as exemplified in reports on the subversion of the Lanakshire milk trial (Tippet 1952) by teachers, of experiments on enriched oxygen environments for premature infants (Silverman 1980) by nurses, and others.

One obvious a priori approach to handing the matter is through clear specification of eligibility criteria for peoples' or institutions' inclusion in the experiment, as in the case of medical trials and others. That is not sufficient in criminology, and elsewhere, when the agreement on criteria is tentative and an authority or service provider may a posteriori take exception to a tentative agreement. Early approaches to handling this issue include provision for a "trap door" in which the service provider who can exercise discretion is told prior to the trial that they can choose X % of eligible candidates prior to randomization for assignment to A or B. Such nonrandom choices prior to the trial are excluded from subsequent "intent to treat (ITT)" analysis of final data. Though it may then take longer to accumulate the necessary sample size, on this account, and though such exclusions may restrict the generalizability of the resultant data, the experiment's comparison is still fair in the sense of yielding unbiased estimates of relative effect. And certainly, the nature of the exclusions can be

examined with respect to the implications for estimates of effects of the intervention tested, albeit more equivocally than the trial itself engenders. It is well understood from criminological studies and others that exercising discretion so as to exclude individuals from the group to which they were randomly assigned after the randomization fouls up the experiment.

It is sensible to document, post facto or in real time, the particular reasons for exercising discretionary exceptions when people are able to articulate reasons. Large scale trials in the health sector have done so occasionally when the discretion is exercised by an individual who is given the choice of participating in the trial or not. An aversion to one or the other of the interventions being offered, for instance, is not an uncommon reason for declining to participate in a trial at times, see Table 6.1 in Boruch (1997) and others on variation in participation rates. But there is not yet an empirical and broad compilation of reasons why referral sources exercise this discretion and for whom, and why otherwise willing and eligible people opt out, for any discipline with which this writer is familiar.

How do we explain random assignment, and to whom?

The authors of the reports in this issue of the *Journal of Experimental Criminology* had to explain the idea of random assignment to people whose permission, authority, or imprimatur was essential for getting the randomized controlled trial off the ground. Some of these listeners, gatekeepers at the top, for instance, understood something about the fairness of a randomized controlled trial. Otherwise there would have been no trial.

Once the trial is approved, once the aircraft is off the ground, keeping the thing in the air requires more brains and skills. In their reports on explaining to others and re-explaining to still others, Kilburn, Hawken, and Strang make the point well. These air crew commanders, to push the metaphor, had to be well positioned and neurotically conscientious about the airspace, including the turbulence that they encountered.

It is in the experimenters' interest to understand exactly how colleagues explained or flopped in explaining at the service provider (judge, cop, welfare worker, etc.) levels. Exactly what was included in the scripts prepared? What kinds of Q and A information was provided orally and in print or in other ways to those providers? And how was this modified or adapted to suit the setting or the people who had to re-explain?

Having the scripts, the Q and A material, and so on is good on scientific grounds. It is part of the scientific log for the trial at hand. Making the material available is also good science in the sense of data sharing, permitting others to learn, and facilitating decent replication of the study . Our contemporary capacity to publish such material in electronic form, big reports with all gruesome details, is substantial, apart from the need and capacity to publish more briefly in hard copy.

Exploiting contemporary electronic and hard copy print capacities is not likely to suffice in the future. For instance, actually watching service providers argue about randomization, or watching them discuss the tensions that they feel between the scientific interests and the interest in helping (favoring a particular intervention over another) or not, yields a different kind of understanding than one can get from print alone. Similarly, watching the individuals who might be randomly assigned to A

versus B argue with one another about the merits and concerns about random assignment is different from recording their responses to questions about the process and reporting in print.

There are some interesting precedents for exploiting audio-visual media in the context of explaining randomized trials to different kinds of people. One of these is a videotape developed for the Rockefeller Foundation's randomized experiments designed to discern the effects of substantial employment and training programs for single parents in six sites in the U.S. It is called "Irrefutable Evidence," and a DVD is now lodged in the Foundation's archives. The DVD's title should not deter the properly skeptical criminologist.

More to the point, the ways in which service providers themselves explained the randomization process to potential applicants to the program, the tensions that they express about the "heart" preferences versus the "brain" preferences in participating in the trials are instructive. The videos and other audio-visual media get beyond what a good writer could do in print. Kilburn referred to the same tensions, but in different ways, in her oral presentation at the Jerry Lee Symposium.

A second precedent is a videotape called "Trials on Trial," developed in the context of recruitment of people for trials on HIV vaccines in the Philadelphia area. Community people in Philadelphia were brought together with experts from the Centers for Disease Control (CDC), a ground level service provider, and the MDs from the University of Pennsylvania's Center for Studies of Addiction to discuss how HIV vaccine trials would be done. A popular and unusually thoughtful Philadelphia TV Host moderated the forum. The product is a remarkable reminder, expressed by community people, that: (1) authority is mistrusted by some vulnerable people and has to be taken into account in trials that involve them, and (2) both experts and ground level service providers carry weight in explaining lottery allocation in different ways.

Nowadays, digital recording of everything is possible at moderate to low cost. The implication is that we can record and make accessible the ways in which people view the notion of randomization, and more important the ways in which the idea can be explained at different levels.

Registries of randomized trials and beyond

The explosive growth of information technologies is in everyone's interest, including criminologists' What is less clear is how to develop evidenced-based policy that facilitates information sharing without swamping the boat, overloading our cortical circuitry, or otherwise being blown away by the high velocity and hot digital winds.

Registering trials of the kind reported here, as Strang and Hawken suggest, is a good thing to do for obvious reasons. In the U.S., the vehicles for doing so include NIH's ClinicalTrials.gov (2011) website and the Institute for Education Sciences' What Works Clearinghouse (2011) website. In the health-related sectors, Canada and the UK have had open trials registers for nearly a decade following on yet another decade of arguments about them. See the work by Chalmers et al. (1992) and Moher and Bernstein (2004) on early origins. For criminological researchers, the Cambridge

University Institute of Criminology's REX-POST Register (2011) of Experiments in Policy Strategy and Tactics is an important initiative.

None of these registers handle directly the reports that are generated as a consequence of the trials. Nor do they handle the micro-data that are generated. Policies that govern the sharing of data from publicly sponsored experiments are likely to change this rapidly. The National Institutes of Health, for instance, has issued formal requirements that researchers with grants over a certain size make micro-records from the research available for secondary analysis. Among others, the American Educational Research Association and the American Psychological Association have developed codes of ethics that include researcher responsibility for sharing the data resulting from the studies reported in peer reviewed journals.

Conclusions

The papers reviewed here help substantially to understand how randomized controlled trials have been deployed in the criminological arena, what problems were encountered and how they were handled, and where the gaps in handling are. The authors' candor is remarkable. I admire.

References

Alexander, L. B., & Solomon, P. (2006). *The Research Process in the Human Services: Behind the Sciences*. Belmont: Thomson-Brooks/Cole.

Binik, A., Weijer, C., McRae, A. D., Grimshaw, J. M., Boruch, R., Brehaut, J. C., Donner, A., Eccles, M. P., Saginur, R., Taljaard, M., & Zwarenstein, M. (2011). Does clinical equipoise apply to cluster randomized trials in health research?*Trials, 12*, 118.

Boruch, R. F. (1997). *Randomized Experiments for Planning and Evaluation: A Practical Guide*. Thousand Oaks: Sage.

Boruch, R. F. & Wothke, W. (Eds) (1985). Randomization and Field Experimentation. *New Directions for Program Evaluation*, Number 28.

Cambridge University, Institute of Criminology REX-POST (2011) http://www.cam.ac.uk/research/experiments/rex-post.

Chalmers, I., Dickersin, K., & Chalmers, T. C. (1992). Getting to Grips with Archie Cochrane's Agenda: All Randomized Controlled Trials should be Registered and Reported. *BMJ, 305*, 786–787.

Christophersen, G. (1983). *Final Report of the Seattle-Denver Income Maintenance Experiment. Volume 2. Administration*. Princeton: Mathematica Policy Research.

ClinicalTrials.gov (2011) http://clinicaltrials.gov.

Cook, P. J., O'Brien, M., Braga, A., & Ludwig, J. (2012) Lessons from a Randomized (Un) controlled Field Trial. *Journal of Experimental Criminology.* (in press)

Fantuzzo, J. F., McWayne, C., & Childs, S. (2006). Scientist-Community Collaborations: A Dynamic Tension between Rights and Responsibilities. In J. E. Trimble & C. B. Fisher (Eds.), *The Handbook of Ethical Research with Ethnocultural Populations and Communities* (pp. 27–50). Thousand Oaks: Sage.

Farabee, D., Shang, S. X., & Wright, B. (2012) Managing an On-the –Spot Lottery in Re-Entry Services. *Journal of Experimental Criminology.* (in press)

Hawken, A. (2012). Lessons from a Field Experiment Involving Involuntary Subjects 3,000 Miles Away. *Journal of Experimental Criminology.* doi:1007/s11292-012-9156-x.

Kershaw, D., & Fair, J. (1976). *The New Jersey Income Maintenance Experiment. Volume 1. Operations, Surveys, and Administration*. New York: Academic.

Kilburn, M. R. (2012). Lessons from the Randomized Trial Evaluation of a New Parent Program: When the Community Sees the Glass as Half Empty, and the Evaluators See the Glass as Half Full. *Journal of Experimental Criminology.* (in press)

MacKenzie, D. L. (2012) Challenges of Conducting Field Experiments in Correctional Settings: Boot Camp Prison Study as an Example. *Journal of Experimental Criminology.* doi:10.1007/s11292-012-9149-9.

Moher, D., & Bernstein, A. (2004). Registering CIHR-funded Randomized Controlled Trials: A Global Public Good. *Canadian Medical Association Journal (CMAJ), 171*, 750–751.

Poduska, J. (2011). Forging and Sustaining Partnerships to Support Child Mental Health Prevention and Services Research. *Administration and Policy in Mental Health and Mental Health Services Research.* (in press)

Porter, A., Merlino, J., Desimone, L., Boruch, R., et al. (2011). *Interim Report on the 21st Century Trial on Improving Science Education in Middle Schools*. Washington DC: Presented at the IES Conference on Experiments in Science and Mathematics in the Schools.

Riecken, H. W., Boruch, R. F., Campbell, D. T., Caplan, N.,Glennan, T. K., Pratt, J., Rees, A., & Williams, W. (1974). *SocialExperimentation: a method for planning and evaluating socialintervention*. New York: Academic

Roman, C. G., Fontaine, J., Fallon, J., Anderson, J., & Rearer, C. (2012). Defending and Managing the Pipeline: lessons for running a randomized experiment in a correctional institution. *Journal of Experimental Criminology.* doi:10.1007/s11292-012-9155-3.

Sherman, L. W., Schmidt, J., & Rogan, D. (1992). *Policing Domestic Violence: Experiments and Dilemmas*. New York: Free Press.

Silverman, W. A. (1980). *Retrolental Fibroplasia: A Modern Parable*. New York: Grune and Stratton.

Strang, H. (2012). From Parties to Pipeline: Relations and Recruitment in Randomized Trials. *Journal of Experimental Criminology* (in press)

Tippett, L. H. (1952). *Methods of Statistics*. London: Williams and Norgate.

Weisburd, D. (2005). Hot Spots Policing Experiments and Criminal Justice Research: Lessons from the Field. *The Annals of the American Academy of Political and Social Science, 599*, 220–245.

What Works Clearinghouse (2011) http://ies/ncee/wwc.

Robert Boruch is University Trustee Chair Professor at the Graduate School of Education with a secondary appointment at the Wharton School at the University of Pennsylvania. A founder of the international Campbell Collaboration, he is a Fellow of the Academy for Experimental Criminology, the American Statistical Association, the American Academy of Arts and Sciences, and the American Educational Research Association.