

QUICK READ SYNOPSIS

Place Randomized Trials:
Experimental Tests of Public Policy

Special Editor: ROBERT BORUCH
University of Pennsylvania

Volume 599, May 2005

Prepared by Herb Fayer (Consultant)

DOI: 10.1177/0002716205275835

Using Place-Based Random Assignment and
Comparative Interrupted Time-Series Analysis to
Evaluate the Jobs-Plus Employment Program
for Public Housing Residents

Howard S. Bloom and James A. Riccio, MDRC

Background

This article describes a place-based research demonstration program to promote and sustain employment among residents in public housing.

- Because all residents were free to participate, it was not possible to randomly assign individuals to the program or a control group.
- Instead, the impact analysis is based on a design that selected matched groups of two or three public housing developments in each participating city and randomly assigned these developments to the program and or control group.
- In addition, an eleven-year comparative interrupted time-series analysis is being used to strengthen the design in order to provide credible estimates of program impacts.

NOTE: The program, Jobs-Plus Community Revitalization Initiative, is a place-based saturation-level employment demonstration tested in six cities. Instead of attempting to achieve a variety of community changes simultaneously, it focused on one goal: improving employment-related outcomes.

- The Initiatives* With a goal of changing individuals and also of transforming the housing developments in which they live, the designers looked for guidance to a growing number of community change initiatives.
- Although their goals and tactics differ in details, these initiatives tend to share a common set of “community-building” principles that stress local control, collaborative decision making, resident empowerment, building on existing resident and community assets, and strengthening the capacity of residents and local institutions to promote and sustain community changes.
 - The designers included a special component, which they called “community support for work,” that helped residents to become sources of work promotion, encouragement, information, advice, and support to each other.
 - They also saw the value of enlisting community stakeholders in designing, funding, and operating the project.
 - This was a three-component intervention:
 - employment-related activities and support services,
 - financial incentives to work, and
 - community support for strengthening residents’ work-focused social capital.
- NOTE: Based on the above, Jobs-Plus was planned to be an unusually comprehensive and intensive community-focused employment intervention.
- Saturation Approach* Jobs-Plus is also distinctive because of its attempt to implement all program components at saturation levels. It was targeted toward all working-age residents living in selected public housing developments.
- Thus, at the very least, all such residents were to be exposed to new work-promoting messages from staff and neighbors.
 - The families who participated could benefit from the new financial incentives and a diverse array of services and supports.
 - Saturation level is fundamental to the program’s theory of change.
 - Targeting the intervention toward the entire working-age population was expected to produce a critical mass of residents whose experiences would generate momentum for change.
 - As these workers grew in number, it was expected that their visibility and influence would be enhanced.
- Collaborative Process* The demonstrations’ planners decided not to attempt to make detailed design choices centrally.
- They chose to leave these decisions to local collaboratives.
 - Each collaborative was expected to include a broad group including the public housing authority, resident reps, the welfare department, and the workforce development system under the Workforce Investment Act—there were also lots of other agencies and providers.
- Measuring Impacts* A central feature of the Jobs-Plus evaluation design is its focus on impacts from two different perspectives:
- Specific individual public housing residents—how did the program affect the future earnings, employment, and welfare receipt of its target individuals, even if they moved away?
 - Specific public housing developments—How did the program affect levels of earnings, employment, and welfare receipt within its target developments, given that different people were living there at different times?
- NOTE: The distinction between these two perspectives is key to any evaluation of a place-based initiative because sample members can move into or out of its target area.

Program Impacts

In the field of employment and training research, random assignment experiments are now regarded as the best way to estimate program impacts.

- However, Jobs-Plus was not a program to which individuals or households could be assigned randomly—all able-bodied working-age adults in selected housing developments could participate.
- The random choice was thus among matched groups of two or three candidate housing developments.
- The remaining developments served as controls.
- Ethical concerns were handled by choosing groups to participate by lottery. Some residents who were interviewed noted that this was fairer than having the local public housing authority make the selection.

NOTE: It will be possible to estimate impacts by comparing outcomes of residents in the program group to those for residents in the control group. It will be possible to pool these estimates across sites by taking their average.

Statistical Power

Just how much statistical power is lost when moving from individual-level to group-level random assignment is an empirical issue that reflects three factors.

- The degree to which individual outcomes vary across groups.
- The number of groups being randomly assigned.
- The extent to which the variance of the outcome within and across groups is reduced by statistical controls for preexisting individual-level or group-level characteristics.

Lessons Learned

The experience with Jobs-Plus provided valuable lessons about using place-based random assignment to evaluate a comprehensive community initiative.

- It was possible to put such a research design into effect.
- It was necessary and possible to keep the research design in place for many years.
- A sustained and intensive effort was required to keep the research design in place and to maintain its integrity.
- When feasible, embedding a comparative interrupted time-series analysis within a place-based random assignment design can improve estimates of program impacts appreciably.

HIV Prevention among Women in Low-Income Housing Developments: Issues and Intervention Outcomes in a Place-Based Randomized Controlled Trial

Kathleen J. Sikkema, Yale University School of Medicine

Background

The scope and urgency of the HIV epidemic requires the development and evaluation of behavior change intervention strategies directed toward communities at risk for increased incidence of new infections.

- Large-scale trials that could reach greater numbers of people are needed.
- HIV prevention outcomes need behavior change supported by peer group norms consistent with risk reduction for best long-term impact.

- For better ways to intervene, we need to not only change the behavior of individuals but also to change social networks and communities to reinforce risk avoidance efforts of population members.
- Diffusion* There has been a resurgence of social diffusion theory as a potential means to induce changes in communities and population member behavior.
- Mechanisms by which diffusion produces behavior change include
 - modeling and observational learning,
 - personal influence,
 - induction of beliefs that the innovative change brings benefits, and
 - creation of new social norms.
 - How efficiently innovations diffuse depends on the nature of the innovation, characteristics of the early adopters, and characteristics of the population.
 - For HIV prevention, the innovative trends include increased communication between sexual partners about AIDS and health concerns, adoption of safer practices, and deferral of sexual activity under risky circumstances when an individual does not want to have sex.
- Modeling* Modeling processes influence behavioral learning and performance in a wide variety of areas.
- Peers who are liked, admired, and viewed as competent exert a strong influence on observer behavior.
 - Perceptions about whether others in one's social group adopt, encourage, or support avoidance steps such as condom use can influence one's own sexual practices.
- Gay Men Trials* The initial study focused on men in gay bars in sixteen small U.S. cities.
- Based on social diffusion theory, popular opinion leaders (POLs) were recruited.
 - They were trained to model and endorse safer sexual behavior.
 - There were eight trial groups and eight controls that received only educational information and condoms.
 - After a year, risk behavior decreased significantly in the intervention cities.
- NOTE: This study is believed to have been the first randomized, controlled trial of a community-level intervention to reduce HIV sexual risk behaviors.
- Female Trials* The researchers extended the prevention models trials to poor, inner-city minority women living in low-income housing developments.
- Many women had low levels of risk sensitization, weak behavior change attitudes, limited use of condoms, and weak perceived norms concerning risk avoidance.
 - The intervention needed to include AIDS preventive skills, attitudes, intentions, and efficacy beliefs and to provide normative and social supports—they also needed to identify social interaction and social influence patterns in the community.
 - The intervention was done over a twelve-month period.
 - Opinion leaders helped develop the risk reduction workshop.
 - These women formed Women's Health Councils (WHCs) and invited women to participate in the two-month workshop.
 - Their goal was to reach all women tenants and to strengthen risk-avoidance behavior, attitudes, and normative perceptions about risk reduction.
 - Community acceptance of the trials was enhanced by the offer of HIV risk reduction workshops to women in the control groups at the end of the study.

- After one year, there was a definite reduction in risk behavior in the intervention group compared to the control group. Several WHCs continued to offer community assistance including funding for additional HIV prevention efforts and programs for adolescents.
- Other Projects* The effectiveness and appeal of this community-level approach has resulted in a variety of ongoing research projects, including among adolescents.
- Dissemination of Findings* It is important to rapidly develop strategies for disseminating research findings and transferring evidence-based HIV interventions to those providing services in community-based and nongovernmental organizations.
- Further research is needed to identify innovative strategies to effectively transfer findings from research into prevention services with support of their implementation in at-risk communities.

Cluster Randomized Trials of Professional and Organizational Behavior Change Interventions in Health Care Settings

Jeremy Grimshaw, University of Ottawa; Martin Eccles,
University of Newcastle upon Tyne; Marion Campbell, University of Aberdeen;
Diana Elbourne, London School of Hygiene and Tropical Medicine

- Background* This article discusses the practical and ethical issues in the design, conduct, and analysis of cluster randomized trials.
- Cluster randomized trials are commonly used in health care. They raise ethical and methodological issues that have rarely been addressed.
 - In spite of positive new evidence to improve patient care, patients, health services, and professionals are very slow to adopt it.
 - Typical approaches to promote findings have emphasized journals, lectures, and seminars, but providers and managers have little time to read literature or attend meetings.
 - They also encounter a range of barriers in areas of finance, organizational facilities, peer group standards, individual skills, and information overload.
- NOTE: It is therefore not surprising that the uptake of research findings into health care appears slow and haphazard.
- Implementation Research* Implementation research looks at methods to promote the uptake of research findings into routine clinical practice by testing approaches to change professional and organizational behavior.
- We should expect the same strength of evidence when considering which strategies to use to improve uptake of research findings as we do when considering which antibiotic to use for each patient.
 - The same arguments used to justify randomized controlled trials of clinical interventions are as salient to the evaluation of dissemination and implementation strategies:
 - The effects of interventions are likely to be modest.
 - The potential for bias is substantial.

- Our poor understanding of competing explanations for an observed effect makes it difficult to adjust for these in nonrandomized designs.
 - Finally, most health care professionals and organizations have limited resources to support implementation initiatives.
- Design Issues* Schwartz and Lellouch made a distinction between explanatory and pragmatic studies.
- Explanatory studies aim to test whether an intervention is beneficial under ideal conditions.
 - Pragmatic studies aim to test whether an intervention is likely to be effective in routine practice by comparing the new procedure against the current regimen.
- NOTE: Implementation research aims to develop a generalizable evidence base to support the choice of professional and organizational change strategies.
- Level of Randomization* Implementation researchers need to consider at what level to randomize units.
- This often involves considering the trade-off between contamination and feasibility.
 - Consider a trial to evaluate professional behavior change strategies within hospital settings:
 - Potential levels of randomization (from “higher” to “lower”) include the hospital, the clinical service or directorate, the ward, and the individual clinician.
 - Randomization at the level of the hospital will minimize the risk of contamination (e.g., members of the control group exchanging information with members of the treatment group) but dramatically increase the number of hospitals required in the study.
 - This could have substantial logistical and financial implications, limiting the study’s feasibility.
 - In contrast, randomization at the level of the individual ward will decrease the number of hospitals required but potentially increase the risk of contamination to an unacceptable level because of regular contact by professionals working within the same ward environment.
 - In our two case studies, we could have randomized at the level of the town, the health care center (where more than one family practice might be based), the family practice (most United Kingdom family practitioners work in group practices), or the individual family practitioners.
 - We chose to randomize at the level of the family practice, reasoning that the risk of contamination within a practice was high but that the risk of contamination across practices sharing the same premises was sufficiently low based upon our own clinical experiences and experiences of previous trials that had not demonstrated contamination at this level.
- Sample Size Issues* Within individual patient randomized trials, responses of patients are considered to be independent from each other.
- Patients within any one cluster are often more likely to respond in a similar manner and can no longer be assumed to act independently.
 - This lack of independence in turn leads to a loss of statistical power in comparison with a patient randomized trial.
 - To achieve the equivalent power of a patient randomized trial, standard sample size calculations need to be inflated by a factor based upon the degree of clustering.

- Pretest Data* Relatively few clusters are allocated to control and study groups in cluster trials in implementation research.
- This causes an increased danger of imbalance in baseline performance of both groups.
 - By adopting a pre-post design, the imbalance can be examined.
 - Adjusting for baseline performance in the analysis stage can also increase statistical power if the premeasure is expected to be a good predictor of postperformance.
 - In implementation research studies, baseline measures of performance are also useful in that they provide an initial estimate of the magnitude of the problem.
- Intervention Rationale* Experimental interventions may be selected for a variety of reasons, including
- the judgment of the researcher based upon a formal or informal assessment of the barriers to adopting an evidence-based practice,
 - empirical evidence about the effectiveness of the intervention under similar (or different) conditions,
 - theoretical considerations, and
 - a proposed change in policy.
- Control Interventions* Experiments may include a no-intervention control or an intervention control.
- A no-intervention control will provide the best evidence about the likely counterfactual in the absence of the experimental intervention—this may not be informative if there is a well-established existing policy or intervention.
- Ethical Issues* Researchers have not paid great attention to ethical issues arising in cluster randomized trials where professionals provide data about the care of individual patients.
- It is helpful to distinguish between ethical consent for study participation and consent for data collection.
 - Individual consent is not the case for cluster-cluster trials, which are typical in implementation research—under such circumstances, professionals' usual ethical responsibilities should override the effects of anything the professional may consider harmful for a patient.
- Data Collection Issues* Commonly, researchers have measured performance by asking the professional what they have done or intend to do in a specific set of circumstances.
- There is evidence that self-reports of activity tend to overestimate actual performance.
 - There is a danger that the intervention may sensitize the professional in the experimental group about desired practice, potentially leading to an imbalance in the degree to which the experimental and control groups report their behavior.
 - Given these concerns, researchers should measure actual performance and not rely on self-report.
- Analytical Issues* There are three general approaches to the analysis of cluster randomized trials:
- analysis at cluster level,
 - the adjustment of standard tests, and
 - advanced statistical techniques using data recorded at both the individual and cluster level.

NOTE: The most appropriate analysis option will depend on a number of factors, including

- the unit of inference,
- the study design,
- whether the researchers wish to adjust for other relevant variables at the individual or cluster level,
- the type and distribution of outcome measure,
- the number of clusters randomized,
- the size of cluster and variability of cluster size, and
- statistical resources available to the research team.

Economic Evaluation

The most informative evaluations to policy makers are those that incorporate concurrent economic evaluations of the relative efficiency of different implementation strategies in addition to their relative effectiveness.

- The evaluation can be used to judge whether costs are reduced and benefits increased, in which case the decision is straightforward, or if both costs and benefits have increased and if the benefits are worth the extra costs.

Process Evaluations

Randomized trials provide little information about the likely causal mechanisms of interventions and modifying factors.

- Process evaluations do provide insight into why an intervention was successful.

Reporting Issues

The need for clear reporting of randomized trials has been widely recognized.

- This has been highlighted through the publication of the CONSORT statement, which outlines the common standards for reporting of field trials.
- The CONSORT statement has been instrumental in improving the standards of reporting in clinical trials and has been widely adopted by medical journals.

Cluster Randomized Trials for the Evaluation of Strategies Designed to Promote Evidence-Based Practice in Perinatal and Neonatal Medicine

Laura C. Leviton, Robert Wood Johnson Foundation;
Jeffrey D. Horbar, University of Vermont

Background

Despite a large and growing body of evidence, the overuse, underuse, and misuse of therapeutic interventions are commonplace.

- New strategies designed to promote evidence-based practice must be identified, tested, and implemented.
- Using the illustrations of two cluster randomized trials, the authors discuss some of the theoretical, methodological, and practical issues of using these designs in a medical setting.
- Hospitals, long-term care settings, group practices, and health insurance plans all offer potential as the unit of assignment and intervention.
- There are many likely causes for the gap between evidence and practice.
 - Most obvious is that evidence continues to develop while practice changes less rapidly.

- Print, electronic, and organizational strategies do not change practices very much.
 -
- Interventions* Organizational interventions offer some promise.
- Decision and administrative supports enable the physician to change practice—things like hiring staff to counsel patients on self-management of chronic illness or a reminder in the medical chart.
 - Several strategies mix the influences for change, such as local opinion leaders pushing innovations.
- Rapid Cycle* A promising development in translating evidence into practice is the rapid cycle, collaborative improvement process.
- Quality teams first identify and plan an overall measurable aim.
 - The teams develop a list of ideas.
 - Through trial and learning cycles that introduce and test relatively small changes, adoption takes place and then another change is added.
- Corticosteroid Trial* The aim of this trial was to test an intervention to encourage obstetricians and maternal-fetal specialists to adopt a practice guideline for using corticosteroid therapy when women go into premature labor.
- Despite strong evidence of this therapy's effectiveness, only 26 percent of eligible cases got the therapy.
 - The study became a test of dissemination strategies—the usual methods would be contrasted with a more active effort.
 - The American College of Obstetricians and Gynecologists endorsed the therapy; this helps effect change more rapidly.
 - Treatment hospitals were exposed to active, low-cost dissemination:
 - designation of a lead physician and nurse at each hospital,
 - grand rounds by an eminent physician associated with the study,
 - group discussions on uses of the therapy,
 - chart reminders, and
 - performance feedback at the hospital level.
- Corticosteroid Trial Results* The intervention gave a statistically significant boost to the use of the therapy.
- The results were useful to show some conditions under which marginal increases in uptake of evidence-based practice could be achieved for relatively low cost.
 - Although changes were already under way in the use of the therapy, the intervention speeded these changes.
 - It confirmed the usefulness of a multifaceted approach.
- Surfactant Trial* The aim of this trial was to evaluate a coordinated, multifaceted intervention designed to close the gap between research and practice in the use of surfactant therapy for preterm infants. There were three components:
- Hospitals in the intervention group received confidential feedback in their surfactant use.
 - The researchers generated a review of the evidence on early surfactant therapy for use in a workshop.
 - The workshop focused on evidence-based quality improvement.
- Surfactant Trial Results* After the intervention, a significantly greater proportion of infants in the intervention hospitals received surfactant therapy in the delivery room.
- The median time an infant received the first dose was twenty-one minutes in intervention hospitals versus seventy-eight minutes in control hospitals.

*The Case for
Cluster
Randomized
Studies*

- There were no significant differences, however, in the primary patient-level outcomes, infant mortality, and pneumothorax.
- Two “sea changes” in practice may have led the researchers to expect too large an effect on mortality.
 - Surfactant use time had already started to decrease.
 - Increased use of antenatal corticosteroids has led to a reduction in the mortality rate of preterm infants since the studies were conducted.

One way to frame the outcome measures for these studies is to track uptake of the practice at specific time points.

- Can intervention speed up the adoption of a practice?
- Taking into account features of an organization can help gain insight for more powerful interventions.
- The empirical argument for cluster random assignment is that medical practices vary substantially across organizations and localities, and powerful forces beyond individual beliefs and knowledge enable or prevent practice change.
 - Within a practice, change is best accomplished when systems and processes are retooled to enable change.
 - Outside the practice, forces affecting change include the standard of local care, the influence of peers, hospital admitting privileges, liability concerns, and the dominant health insurance plans.
 - Practitioners are embedded in organizations and networks that have strong influence on their behavior.
 - Secular trends can reduce or even wash out the effects of an intervention.
- Preexisting research networks offer economies of scale (in reduced cost and improved quality) in conducting cluster randomized studies.

Historical Review of School-Based Randomized Trials for Evaluating Problem Behavior Prevention Programs

Brian R. Flay, University of Illinois at Chicago;
Linda M. Collins, Pennsylvania State University

Background

The design and statistical methodologies used in school-based intervention research have advanced greatly in the past twenty years. Methods have improved for

- randomization of whole schools and the resistance to randomization,
- choice of appropriate control groups,
- solutions when randomization breaks down,
- limiting and handling of variation in integrity of the intervention received,
- limiting biases introduced by data collection,
- awareness of the effects of intensive and long-term data collection,
- limiting and analysis of subject attrition and other missing data,
- parental consent approaches,
- design and analysis issues when only small numbers of schools are available or can be afforded,
- the choice of the unit of analysis,
- phases of research,
- optimizing and extending the reach of interventions, and

Efficacy Trials

- differential effects in subpopulations.

NOTE: School-based prevention research still faces many significant methodological challenges in the above areas.

Most prevention studies have been efficacy trials.

- They have been conducted within the framework of models derived from the FDA's approach to the development of new drugs.
- Many researchers have bemoaned the lack of effectiveness trials, or other kinds of studies to assess the effects of proven interventions under real-world conditions—the need is for more randomized trials.

Efficacy and Effectiveness

Before the 1980s, there was little recognition of the desirability of a carefully designed sequence of studies to inform the development, testing, and adoption of effective prevention programs.

- Flay and Best advocated a careful distinction between formative and summative research, with the idea that interventions developed for early phases of research would not necessarily translate into effective programs in the real world.
- They also suggested the need for a thoughtful sequence of phases of research during the development of prevention programs.

Conclusions

There are six major conclusions.

- Sequence planning is important.
- Time is important—the ultimate effects of importance to society of an intervention occur over the long term.
- Keeping up with and remaining open to methodological advances is important.
- Publication of all results is important.
- Accumulation of knowledge is important—the true measure of advancement in knowledge is not the individual study but the accumulated findings from reviews and meta-analyses.
- The devil is in the details—methodologically sound school-based research is not easy; it requires lots of institutional relationships and on-the-ground work relating to school administrators and teachers.

Place-Based Randomized Trials to Test the Effects on Instructional Practices of a Mathematics/Science Professional Development Program for Teachers

Andrew C. Porter, Vanderbilt University;
Rolf K. Blank, Council of Chief State School Officers;
John L. Smithson, University of Wisconsin–Madison;
Eric Osthoff, University of Wisconsin–Madison

Background

A professional development model was designed for and is being tested in about fifty U.S. middle schools in five large urban districts.

- The focus of the article is on learning about doing place-based randomized trials to test the efficacy of education programs (treatment is not yet complete and study results are not available).

- Half the study schools in each district receive the intervention and half are controls.
- Each school forms a math/science leadership team including at least one administrator.
- Teams receive professional development workshops, then work with all math and science teachers in their school to teach them to use data on their practices to improve their effectiveness.

Reasons for the Study

There are several reasons to study a professional development program using place-based randomized trials.

- Over the past decade, a great deal of research has established the characteristics of effective professional development.
- We now have several tools for studying teachers' decisions about what to teach.
- There is now a great deal of interest in data-based decision making in education to help improve teacher practices.
- There is also interest in and support for doing randomized experiments in education.

NOTE: The focus of this article is on the design and implementation of the place-based randomized trial.

Independent Variable

The independent variable is the treatment versus the control condition—in education, there is no strict control.

- Control school teachers continued to participate in regular professional development.
- The teachers in intervention schools participated in professional development other than the treatment.
- The data were collected not only on the quality of the treatment implementation but also on the other professional development experiences of both the treatment and the control teachers.

NOTE: The professional development is aimed at helping teachers reflect on their individual and collective practices, and from that they are to decide how their instruction might be strengthened.

Dependent Variables

The lack of prescriptiveness for classroom practice makes defining valid and sensitive dependent variables difficult.

- One dependent variable is the degree of alignment between the content of each teacher's instruction and the content of the state or district test used for accountability purposes.
 - Uniquely targeted dependent variables were formulated for each school, focusing on the types of changes in instruction that each school decided were most needed.
 - The problem with these targeted variables is that they are aligned with the intended changes in one school rather than the intended changes across all schools.
- Ideally, gains in student achievement would have been another dependent variable, but due to a three-year limit on the program by the National Science Foundation, there are not strong enough effects to measure student achievement.

Data on Enacted Curriculum

The professional development model for the Data on Enacted Curriculum (DEC) project is based on

- standards-based improvement of instruction;
- continuous improvement of practice using data and formative evaluation;

- school-based collaboration and networking to foster the sharing of ideas, models, and strategies improvement;
- district contact persons to involve district-level instructional support staff by inviting them to participate in all workshops; and
- treatment steps that include
 - baseline surveys of instructional practice and student achievement,
 - a two-day professional development workshop,
 - follow-up technical assistance in schools,
 - a professional development follow-up workshop, and
 - evaluation of progress and refocusing of assistance.

NOTE: A major intent of the DEC treatment is to involve teams in the processes and techniques for using their data to highlight important questions and to discover tentative causal factors. Another goal is to develop the capacity of leadership teams to engage a larger group of their own staff in dialogue about their data and inquiry into their own teaching and learning.

Participation Factors

Five factors have affected the extent and quality of school participation in the treatment:

- time for meetings of the school leadership teams;
- stability of the teams;
- schoolwide use of the DEC treatment—it is difficult to ensure the quality of treatment due to variable team training and possible conflicts with school cultural norms;
- principal participation; and
- district priorities and policies.

Critiques

Campbell and Stanley provided a powerful template for critiquing experiments in education.

- They identified internal validity—the extent to which an experiment provides unbiased estimates of treatment main effects.
- They also identified external validity—the extent to which the results from an experiment can be generalized.
- To the above can be added issues of precision—the extent to which the effect size is estimated with a small standard error.

In this experiment,

- The internal validity is strong, though not perfect and the external validity has several limits:
 - The focus is on urban schools with students from low-income families.
 - The work is in middle school math and science.
 - Schools were volunteers.
 - A few schools dropped out.
 - The timing of the study and the context in which it was conducted—under a huge emphasis on standards-based reform and a push in math and science achievement.
- The three-year limitation had negative effects on external validity.
 - A delayed test of effects was not possible.
 - Effects on student achievement in such a short period are unlikely.
- The designers of the intervention were the designers of the study—a third-party evaluator would be preferred.

Insights

Since experiments have been rare in education, it is worth sharing some insights.

- A fundamental issue in considering an experiment in education is knowing if an experiment is warranted.

- First, there must be a promising intervention.
- Second, the treatment needs to be reasonably straightforward.
- Third, the treatment should be pilot-tested to get an indication of its feasibility and promise.
- Another issue is the hypothesis being tested.
 - To what should the treatment be compared? Selecting a comparison group is one of the most important decisions.
- The degree of treatment implementation is dependent to a considerable extent on the vagaries of the school districts and states.

Three Key Challenges

To bring the treatment to scale, should it be found effective, has three key challenges.

- A way would have to be found to produce more trainers on a par with the one used in this experiment.
- There is a need to build an infrastructure for “selling” the treatment.
- A way needs to be found to maintain the fidelity and integrity of the treatment as it is scaled up.

NOTE: These challenges were not addressed in this experiment but need to be once the effectiveness of the experiment is shown. Questions about taking the treatment to scale, which are in some ways analogous to concerns about external validity, must be addressed if an intervention is ultimately to be found worthy.

Emergent Principles for the Design, Implementation, and Analysis of Cluster-Based Experiments in Social Science

Thomas D. Cook, Northwestern University

Background

All things being equal, it is much more difficult to do an experiment that simultaneously deals with units at two or more levels—this is usually the individual and some higher-order level like a community or school.

- Social scientists have not yet acquired much experience in implementing cluster-based experiments, and the knowledge to be gained by emulating medicine’s more evolved tradition of multisite randomized clinical trials is limited.
 - In the medical tradition, individuals are assigned to treatments from within sites rather than in between-site designs.
 - Thus, sites serve to increase power and generalization, not to be units of assignment per se.

NOTE: Social scientists will have to learn most of their lessons about cluster-based experiments in the crucible of their own experience. Solutions will also have to come from reflections on such experiences as well as from developments in statistical theory.

Cluster-Level Research

Why assign at the cluster level?

- Most concepts to be studied like school governance, culture, climate, norms, teams, and networks cannot be reduced to individual behavior—the

*Random
Assignment*

hope is to create a new culture whose norms will affect the behavior of current (and also future) teachers and students.

- The unit of assignment level is chosen either because of theory or to protect against a potential source of bias.
- An argument for cluster-based assignment has to do with desired impact. The hope is that individual change will be greater if it is achieved through group- rather than individual-level processes.
- There is a political reason for assigning clusters rather than individuals. Assignment to different statuses inevitably creates potentially large inequality.

Why the random assignment of larger units?

- To state that social structures are hierarchically ordered only scratches the surface of the myriad forms of this ordering.
- To add to the complexity, social structures are not fixed in time.
 - Students can change classes or leave school.

NOTE: Fortunately, well-implemented cluster-based random assignment simplifies structural complexity in ways that nonexperiments cannot. Without random assignment, any school-based causal study would have to struggle to rule out the possibility that various interdependent structural realities function as causal confounds.

Principles

Here are five principles of improved cluster-level design:

- *Principle 1:* Know the size of unconditional and conditional intraclass correlations, what determines them, and how they affect statistical power and hence sample size estimation.
- *Principle 2:* Assign units to treatments at the lowest level possible as long as this does not change the research question, for lower units entail many important advantages.
- *Principle 3:* Minimize interunit communication, though it will often be especially unclear how much to expect in designs where treatment assignment is within-school or within-neighborhood.
- *Principle 4:* Avoid black box experiments despite their policy relevance; instead, explore implementation and causal mediating processes.
- *Principle 5:* Explicitly acknowledge the great heterogeneity in treatment implementation and analyze its consequences confessing to the lower inferential quality of such conclusions.

Randomization and Social Program Evaluation: The Case of Progresá

Susan W. Parker, Centro de Investigación y Docencia Económicas;
Graciela M. Teruel, Universidad Iberoamericana

Background

This article analyzes the development of Progresá, the principal antipoverty strategy of the Mexican government.

- The program is aimed at increasing families' investment in human capital as defined by education, health, and nutrition and includes cash transfers linked to regular school attendance and family clinic visits. The program also

offers in-kind benefits including basic health care and nutritional supplements for children up to age five as well as pregnant and lactating women.

- It has served as a model for similar programs in several other Latin American countries.

NOTE: The program was subject to a rigorous evaluation effort in rural areas that included an experimental research design and impact studies conducted by an outside institution, the International Food Policy Research Institute (IFPRI).

Program Benefits

Progresa combines education, health, and nutrition in one program to enhance the effectiveness.

- Cash benefits are conditional on children attending school and the family completing a schedule of regular clinic visits.
- Proxy means tests carried out for targeting.

NOTE: All the monetary benefits are given to a woman of the family based on research that women use more resources for children than do men.

Human Capital

The design of Progresa focuses on the lack of human capital as a central cause of poverty and hopes to reduce poverty by increasing investment in human capital, particularly that of children.

- It assumes that returns to education, health, and nutrition are likely to be high for its beneficiary population.
- The caveat is the little available information on returns to schooling and health in poor areas where Progresa operates.

Progress Evaluation

The Progresa evaluation includes a randomized design.

- The strategy was to select a sample of Progresa communities and randomize communities between the treatment and the control groups:
 - A potential limitation of randomizing at the community level was the likelihood of observing less randomness at the individual level.
 - A sample of 506 communities with 24,077 households in seven states was used.
- The experiment lasted only a year and a half, which makes longer-term impacts more difficult to analyze.
- Control group could not be maintained longer because of the rapid growth of the Program.

Evaluation Results

The results of the evaluation were extremely positive.

- In the words of the coordinator of the IFPRI evaluation, “The results show that after only three years, poor children have increased their school enrollment, have more balanced diets, are receiving more medical attention, and are learning that the future can be very different than the past.”
- The only noted negative impact in the IFPRI evaluation was some increasing tension between Progresa beneficiaries and nonbeneficiaries in some communities.

Impacts on Public Policy

Some of the impacts of the evaluation include:

- The evaluation played an important role in ensuring that the program was not eliminated with the change of government.
- The year after the evaluation was released, Congress issued a new law requiring all social programs to carry out external evaluations of their impacts every year.
- Other Latin American countries have implemented similar programs and also implemented the model of external evaluation.

NOTE: Whereas the experimental design has been a critical factor in the program's initial success, it appears that estimation of the longer-term impacts will require nonexperimental estimators or structural estimation.

Hot Spots Policing Experiments and Criminal Justice Research: Lessons from the Field

David Weisburd, Hebrew University and University of Maryland

Background

There is widespread acceptance that randomized experiments provide more valid answers to policy questions than do nonexperimental studies.

- Experiments provide a superior method for assessing the effectiveness of a given intervention.
- Yet randomized studies are the oddity rather than the norm in spite of the fact that these experiments have been successfully carried out across a wide array of criminal justice settings. Explanations often given:
 - There are widespread ethical problems to overcome.
 - Experiments are difficult to implement.
 - Implementation imposes so many limitations on criminal justice practice that they are not likely to have much policy relevance.

NOTE: Place-based randomized trials in the area of hot spots policing appear at odds with the above assumptions—hot spot policing has become a core strategy and the research played an important role in its adoption.

Situational Causes of Crime

Criminologists recognize that the situational opportunities provided at the level of “place” can affect the occurrence of crime.

- Criminal events require not only a “motivated offender” but the presence of a “suitable target” and the absence of a “capable guardian.”
 - Crime rates can be affected by changing the nature of the targets or of the guardianship regardless of the motivations that individual offenders bring to the crime situation.
- One natural outgrowth of these perspectives was that the place where crimes occur would become an important research focus.
 - Important research was done in Minneapolis and Jersey City that radically changed how criminologists understood the crime equation.
 - The findings showed that place crime can be reduced without displacing crime to other areas.

NOTE: It is important to recognize that randomized experiments must be conducted if criminal justice is to draw valid policy conclusions about what works. Ethical concerns, enforcement, and cost are factors that deter researchers from using this method.

Ethical Concerns

It may be possible to avoid many ethical dilemmas by randomly allocating at the organizational or place level, rather than by individuals.

- This was seen in both the Minneapolis and Jersey City studies.
 - No individuals were contacted for information, and all data came from police records and observation of the sites.
 - Accepted police practices were used.
 - There was no placebo—control sites also received some form of criminal justice treatment.

<i>Importance of Monitoring</i>	<p>Although monitoring may be difficult, the experience of the hot spots studies is that monitoring may be facilitated by focusing on clearly defined places and the application of treatments in a visible social environment.</p> <ul style="list-style-type: none"> • The importance of monitoring in both studies, Minneapolis and Jersey City, was confirmed by identification of breakdowns in the application of treatment. • It is important to integrate clinical work and research work in criminal justice to get the practitioner to believe in the importance of implementing a randomized study. • It is also important to consider coercive pressure from higher-ups to maintain treatment fidelity.
<i>Concerns</i>	<p>The fact that a limited number of cases can be included in many cluster randomized trials raises concerns regarding the ability of randomization to provide for equivalent groups.</p> <ul style="list-style-type: none"> • Hot spots were grouped into statistical blocks made up of smaller components within the study—for example, the fifty-two hot spots were divided into four groups based on reported arrests and emergency calls to police. • It appears that in spite of the successes using randomized trials in hot spot studies, there is reluctance to use these trials in other areas due probably to cost and difficulties in gaining approvals from funders who are not predisposed to this science.
<i>Conclusions</i>	<p>There are eight specific lessons regarding the implementation and development of place-based randomized trials and experimental methods more generally:</p> <ul style="list-style-type: none"> • Crisis in the legitimacy of conventional practices may encourage randomized controlled experiments. • There must be a predisposition toward randomized trials in political processes that lead to funding. • There is a need for governmental support. • Random allocation of places can lead to fewer ethical objections. • Monitoring of the treatment fidelity is essential. • Strong hierarchical controls within the institution administering treatment facilitates implementation. • The more complex the treatment, the more the need for greater coercive controls of the treatment integrity. • Block randomization provides a method for overcoming some problems related to restrictions in sample size.

Introducing New Contraceptives in Rural China: A Field Experiment

Herbert L. Smith, University of Pennsylvania

<i>Background</i>	<p>This article presents an overview of work during the 1990s on a family planning experiment in four counties in rural China.</p> <ul style="list-style-type: none"> • The project featured three interventions: <ul style="list-style-type: none"> • provision of better contraceptives; • better training of family planning workers; and
-------------------	--

- offering a choice of contraception method—an innovation in the Chinese family planning system.
- The project was designed as a randomized experiment.
- The experimental unit was not the individual woman but rather the lowest organizational unit feasible for the design.
 - Within each county, six townships were selected at random.
 - Two of the six were randomly designated as controls.

NOTE: The decision to use twenty-four townships, with treatment groups in a two-to-one ratio to control groups, reflected the cost-benefit compromise that attends social research.

Study Settings

It became evident that interesting demographic and programmatic distinctions existed both within and between the counties in the study.

- The authors became interested in the politics, economics, and social organization of the system, including variations within the study sites.
- Although the formal design took no account of the measurement of these phenomena, field visits and notes came to rely heavily on gleaning “side information” about local process, to better understand the data that were dominated by effects not attributable to the experimental interventions.
- Visits to the sites were variously to
 - meet local officials and select study sites,
 - monitor the general course of the research,
 - participate in training of interviewers,
 - interview family planning workers and their clients,
 - set up a record-keeping system,
 - maintain a demographic and contraceptive surveillance system,
 - examine related health system records, and
 - observe and evaluate the extent to which innovations were maintained and expanded beyond the project sites.

Experimental Design

In experimental terms, the treatments were weak and ill defined; many investigators did not understand the experiment as designed; contamination was rife; and “history” intervened.

- In concept, the design was probably not a bad one—it was based on the distinction between the unit of experimentation and the unit of analysis.
 - It is advisable to use randomization at the level at which units are most naturally manipulated.
 - When the unit of observation is “lower” than the unit of randomization or assignment of treatment, then for many purposes the data need to be aggregated to measure it at the level of assignment.
- Unlike the typical individual-level encouragement design, this project randomized the treatment at the lowest level that a policy could conceivably be implemented.
- Variations for each annual difference between treatment and control townships have two components:
 - Differences between townships with treatment types, with fifteen and seven degrees of freedom, respectively.
 - Sampling variability in the estimate of each township ratio, exacerbated by clustering (by village) which tends to inflate design effects.
- Even though the sample size at the level of the unit of observation is fairly large, differences between treatment and control townships from the Baseline Survey are well within the margin of error accruing to randomization and sampling.

- One of the hallmarks of the project was the recognition that population and family planning processes varied widely across study counties as a function of latent differences in the organization, motivation, and political centrality of the respective county planning administrations.
- Statistics* When the data are examined, the benefits of randomization are substantially attenuated.
- As for the variation between treatment and control groups within a county, this is founded on two control townships and four treatment townships.
 - With so few degrees of freedom, randomization does not afford much statistical power—we are back at something like comparative case studies.
- Conclusion* The project was an attempt to examine the effects of policy shift by manipulation of aggregated subjects—the rural townships.
- Hindsight suggests that even at this level there is the possibility of serious interference between units (contamination).
 - The policy intervention was weak as measured against a host of external factors influencing demographic change in these areas.
 - There is the problem of incomplete randomization as a function of the small number of experimental units, a problem that is exacerbated when treatment effects interact with population characteristics.
 - It would be folly to argue that the New Contraceptives in Rural China study, although quite successful as a demonstration project, was a compelling counterexemplar to skepticism regarding randomization at the level at which policy is being implemented.