Are We Getting What We Paid For?
The Need for Randomized Environmental Policy Experiments in Georgia

Water Policy Working Paper #2005-022

Prepared by:

Paul J. Ferraro*

June, 2005

***Paul Ferraro is affiliated with Georgia State University, Andrew Young School of Policy Studies, Department of Economics.** **The author acknowledges the assistance of Kwaw Andam in developing the ideas in this paper. The author gratefully acknowledges financial support for this work received from the Georgia Soil and Water Conservation Commission, Contract No. 480-05-GSU1001; and the U.S. Department of Agriculture, Award Document No. 2003-38869-02007-1.**

Are We Getting What We Paid For?

The Need for Randomized Environmental Policy Experiments in Georgia

Abstract

In the field of environmental policy, the decision to choose one policy over another should be evidence-based. Randomized policy experiments are important tools for generating evidence on the effectiveness of policies. They are an important component of policy design in fields such as poverty assistance, criminal rehabilitation, public education, and public health. In contrast, the use of randomized experiments in the field of environmental policy is nonexistent. In this short paper, I argue that randomized experiments are needed to improve environmental policy in Georgia. They can take place in the context of planned pilot initiatives and thus require little additional money to implement. Because they can be incorporated into the implementation of a field initiative, policy experiments also mitigate concerns that research and program implementation are mutually exclusive. However, the difference between what one can learn from a pilot initiative that uses a randomized design and from one that does not is enormous. We illustrate how one can use a randomized policy experiment in the context of an existing water conservation initiative in Georgia.

Are We Getting What We Paid For?

The Need for Randomized Environmental Policy Experiments in Georgia

I. Introduction

In the 18th century, the notion that medical treatments should be evaluated through randomized trials gained acceptance. Supporters of such trials argued that knowledge about the effectiveness of interventions could not be left to intuition and faith, but must be evidence-based. Today, randomized trials are essential to the evaluation of medical treatments in the world.

In the United States, the application of randomized experiments to the evaluation of policy is also well established. Experiments are an important component of policy design in fields such as poverty assistance, criminal rehabilitation, public education, and public health. In contrast, randomized experiments have not been used in the field of environmental policy.

Program evaluation is fundamentally about making inferences about an unobserved counterfactual event: what would have happened if there had been no intervention? For example, how different would a firm's pollution have been had it not been provided with a subsidized environmental audit? How different would a community's water use have been had it not received water conservation education materials?

Non-randomized evaluations generally take one of two forms. One looks at participants before and after the program. The other compares participants with non-participants. Both forms of evaluation are flawed. In the first case, the absence of a control group makes it impossible to know how participants would have fared in the absence of the program. In the second case, the decision to participate is often related to factors that affect outcomes. For example, firms that participate in a voluntary audit program are often more likely to be actively searching for ways to reduce input use, and thus comparing them to non-participants yields a biased (upwards) estimate of the program's effect.

Randomized trials represent the best evaluation method known to social science. With random assignment, the outcomes for the control group (non-participants) represent what would have happened to those in the treatment group (participants) in the absence of the program.

To improve Georgia's environmental quality, it is essential that we know which of Georgia's existing or proposed environmental policies work. As in medicine, if we want to know the effect of a policy intervention, randomized experiments provide the most accurate answers.

In order to make the discussion about the use of policy experiments concrete, we will focus on water conservation initiatives in Georgia. However, the application of randomized experimental trials is applicable to many more environmental contexts.

II. Evaluating Water Efficiency Programs in Georgia

Georgia faces growing water scarcity. In response, the Department of Natural Resources (DNR) has developed water efficiency programs for businesses, utilities and households to ensure efficient use and reduce water losses. Given the limited funds available for such programs, policymakers and citizens would like to know, "Which types of water efficiency programs are most effective?" The best way to answer such a question is to use randomized field experiment in which a specific program (e.g., adding educational inserts to water users' bills) is randomly assigned to some members of the target population, but not to others.

Most evaluations of water efficiency programs rely on "before" and "after" measures of water use for program participants. However, policymakers cannot infer a program's effect on water use without similar measures of changes in water use by non-participants. Non-participants serve as controls and allow analysts to determine how much of the observed change in water use among participants is due to the program.

For example, an evaluation of a water efficiency program in Florida found that program participants reduced their water use by nearly 40 percent.[1] However, the study found that non-participants had also reduced their water use by more than 30 percent (high rainfall was observed during the study period). Thus, the program seems to account for less than 10 percent of the reduction in the participants' water use.

However, not just any non-participant can serve as a control. The participants and non-participants in the Tampa program were not assigned at random and thus conclusions based on their changes in water use are susceptible to "selection bias." Selection bias occurs when characteristics that influence the outcome variable also influence the probability of participating in the program. For example, say that the Florida program was implemented in areas that were experiencing substantial growth in irrigation use. If one chooses controls from the set of all non-participating areas, one may select areas that are not experiencing high growth in irrigation. Thus, when the analyst compares the changes in water use between program participants and non-participants, the difference yields a biased estimate of the program's effect. For example, Florida program participants may have in fact experienced only a 10% decline in irrigation had there been no program, rather than the 30% estimated decline implied by the poorly chosen controls.

[1] Mulville-Friel, D. and Anderson, D. Justifying Water Conservation through Evaluations. *Florida Water Resources Journal*, August 1996, 18-20.

Randomized policy experiments offer a way to avoid these and other pitfalls of conventional environmental program evaluations. In a randomized framework, the treatment and the control groups are alike, on average, in all respects except the treatment itself (recall that "treatment" refers to the environmental program being evaluated). Thus the control group's outcome represents what the treatment group's outcome would have looked like had it not received the treatment. In other words, the control group provides an accurate estimate of the counterfactual and thus the difference in outcomes between those who participate in the program and those who do not represents the effect of the program.

An experiment is often nothing more than redesigning an existing pilot initiative to assign participation randomly among interested individuals or firms, rather than use some other criteria to assign participation (or to allow all interested individuals to participate). For example, a mentoring program in which senior female professors mentor junior female professors was created with funding from the National Science Foundation. They originally did not plan to run the pilot program as an experiment. However, more junior female faculty signed up than the program could accommodate and thus some additional criteria to select participants had to be used. The program could have allowed "high-potential women" to participate (based on objective or subjective criteria) and then follow the career performances of only the participants (i.e., how many publications, their salaries, and their success in receiving tenure). As indicated by the discussion above, however, this kind of evaluation would not give the program designers and funders an accurate picture of the effectiveness of mentoring. Events that affect the career performances of female faculty may have occurred at the same time that the mentoring program was implemented and looking only at the participants would give a biased estimate of the program's effect. Comparing the "high-potential" women who were selected to those who were not selected (or did not volunteer to participate) would also give a biased estimate because these women may have been fundamentally different from the other female faculty (for example, they may have succeeded regardless of whether they were in a mentoring program).

Rather than develop some ad hoc criteria for selecting participants, someone came up with the idea of selecting participants randomly from the pool of eligible and interested women and then following the career performances of women from this pool who received mentoring and those who did not. The only additional cost of the program was the need to gather outcome measures on the women who did not receive mentoring so that one could determine if mentoring

had an effect (this cost was small, particularly since they reduced the participant pool below the original target and used the saved funds to monitor the control group). In addition to providing an objective measure of the mentoring program's effectiveness (the evaluation is not yet complete), this kind of experiment has another desirable feature: political or social power does not determine who gets to participate and who does not (it is random) and thus many perceive the program as being "fairer."

III. An Example: The *Water Use It Wisely* Program

The *Water Use It Wisely* (WUIW) program is sponsored by Georgia Natural Gas in partnership with Georgia's Department of Natural Resources and the Pollution Prevention Assistance Division. One of its main initiatives is to educate residential consumers on the need to conserve water and the different ways through which consumers can do so. For example, the WUIW program has educational inserts that accompany Georgia Natural Gas bills. These inserts give water-saving tips to households (e.g., "only wash full loads of laundry," "set a timer when watering your lawn or garden with a hose"). A valid research question is: ***do the educational inserts reduce water use among recipient households compared to their use in the absence of the program?*** Answering questions like these can help Georgia policymakers and collaborating businesses prioritize among program alternatives and improve water efficiency efforts.

To answer the question, one can simply randomly assign some households to receive inserts over the course of a year and others to not receive the inserts. An analysis of the differences in water use between the receivers and the non-receivers provides a statistical measure of the changes in water use that can be attributed to the education program. The nature of these inserts can also be varied randomly: some receivers simply observe water-saving tips, others receive the tips and appeals to their social consciences, and a third group receives water-saving tips and coupons for water-saving devices. More complicated experiments can assign other treatments to households in the state (e.g., water audits, post-audit technical assistance, and public recognition or prizes for demonstrable water reductions).

Note that little change to the existing program is necessary. In fact, costs are reduced because not everyone receives the information. The cost savings can be used to pay for the monitoring of water use among recipients and non-recipients, in collaboration with water suppliers. If household water use is too costly to monitor, then the experimental treatments can be assigned at the level of communities or towns (i.e., the experimental design can randomly assign communities to two groups – a group that receives water conservation information and a group that does not receive the information). Note that although I focused on a single program as an example, most other water efficiency programs (e.g., technical assistance or water audits to private firms or farmers) are also best evaluated through a randomized experimental design

IV. Conclusions

Randomized policy experiments are popular in fields such as education, social welfare reform and public health. They provide accurate information about policy effectiveness and help to improve future decision-making. The use of such experiments in the design of environmental policy, however, is nonexistent. Their absence is unfortunate given that they are so sorely needed, while requiring few additional funds to implement. In many cases, these experiments can take place in the context of existing initiatives.

Government agencies, firms and nonprofit organizations can serve as active partners in the design and implementation of policy experiments, particularly by bringing in their local knowledge and grassroots mobilization capacity (see, for example, efforts underway in the African education sector[2]). With widespread cynicism about the effectiveness of conservation investments, providing clear evidence on the effects of different interventions can also help galvanize support for more conservation financing.

In the specific case of water efficiency programs in Georgia, the use of randomized policy experiments can improve water resource management in the state in at least the five ways: (1) allow policymakers to compare program alternatives at the pilot project stage prior to full-scale implementation; (2) indicate where modifications are needed; (3) allow easier comparisons of program costs and benefits; (4) promote public and institutional support for water efficiency programs that yield proven results; and (5) assist agencies such as the Department of Natural Resources in justifying current funding levels and acquiring new funding for effective programs.

[2] Kremer, Michael (2003) "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons," *American Economic Review* 93(2): 102-115.