(Comparable) Non equivalent control group design

Neighborhood Stratification in Studies with Volunteers

Loren Kirkeide, Principal Load Research Analyst, SRP November 12, 2009

I oversee the load research group at SRP, an electric utility in the Phoenix metropolitan area. A manager in the Corporate Pricing department asked for my opinion on two studies emailed to him by a consultant. The studies revealed a statistically significant decrease in total energy usage when customers switched to a time-of-use (TOU) rate from a non-TOU rate. Although the studies were sent to SRP by a consultant involved in conducting experiments, I saw problems with the experimental designs in the studies. I'm not disputing the results for their specific TOU rate designs, and the consultant did not conduct the studies, but I wanted to get his thoughts on these experimental designs. I have read other studies using these or similar methodologies to recruit volunteers for treatment groups and randomly select a sample of residential customers to serve as a control group, but I finally decided to see if another researcher had similar concerns. I asked the consultant for his opinions on the recruitment of volunteers and the random selection of a control group. He offered a few comments, but did not express my concern over the control group sampling method. His response after I expressed this concern was essentially neutral, but I sensed he wanted to think about it.

This has prompted me to put something in writing to hopefully generate discussion about this specific issue and share ideas on how to address it. This issue concerns the recruitment of volunteers to test a new program, such as a new TOU rate. The management of many utilities has generally not supported mandatory customer participation in "experimental" programs, so some form of volunteer recruitment is required. Of course, mandatory participation is only an issue if the intent is to understand behavior in a program likely to become mandatory. And management is generally unsupportive of recruiting volunteers and then telling half of them they cannot participate in the program because they have been randomly assigned to a control group. This approach would apply if the intent is to have a voluntary program. One is often left with recruiting volunteers for the treatment group and then selecting a control group for comparison. As mentioned above, a common approach is to randomly select customers for the control group. The treatment and control groups also may be stratified based on amount of energy usage.

We know recruiting volunteers is not random selection, but that will likely remain very difficult to overcome in real-world research. I also suspect we'll agree that "volunteerism" for electric utility studies is not randomly distributed across all customers in a utility's entire service area. Some customers are more willing to volunteer than others. It's not unreasonable to assume that some neighborhoods likely possess a much greater propensity to volunteer for a TOU study, for example, than other neighborhoods. This creates a problem beyond the "volunteer" effect in these studies. Neighborhoods differ on many socioeconomic and demographic factors, and they differ on the type, size and age of the homes. If we agree that people in the same neighborhood are more similar than people in different neighborhoods and the housing in the same neighborhood is more similar than housing across neighborhoods, we need to take a closer look at how we select control groups for comparisons with treatment groups.

We can begin with acknowledging that a randomly selected control group does not make a study using volunteers for the treatment group at least "half right" for statistical analysis. Randomly selected control groups do not offer any partial fix for the treatment groups formed from the recruitment of volunteers. Neighborhood differences caused by the very different selection methods for treatment and control groups

may well be so significant that results based on comparisons between the control and treatment groups could be quite misleading. The treatment effect may be so significantly confounded with a neighborhood effect that we cannot avoid questioning the reliability of the results. Neighborhood stratification at least eliminates the potential for a neighborhood effect ending up embedded in the treatment effect. The goal is to eliminate treatment and control group differences in "between-neighborhoods" variance due to all the factors mentioned above, so that we only have differences in "within-neighborhoods" variance between the groups. This allows us to say that our control group exactly reflects the neighborhoods from which we obtained our treatment group.

Some may view this as a matched-pairs design, while others may define this as a form of stratification. Within-neighborhoods variance is still substantial. Even with neighborhood stratification, there may be cases where the treatment and control customers for a neighborhood will be comparing a single person to a family with young children, an 1,800 square-foot house to a 2,600 square-foot house, a $50,000 household income to a $90,000 household income, etc., so I personally lean toward defining it as a special form of stratification. Information is usually not readily available to match treatment and control groups on important lifestyle/household/demographic/dwelling type factors, but creating a matched-pairs design is not the goal. The goal is to select a reasonably comparable control group matching the "neighborhood mix" of our volunteers in the treatment group. In addition, stratification based on neighborhood and energy usage helps further improve similarity on such other crucial factors as air conditioning, appliances, home electronics (e.g., big TVs), swimming pools, etc. We definitely know where they live and how much electricity they use. Stratifying on neighborhood and energy usage reduces confounding of our treatment effects with differences between treatment and control groups due to these other uncontrolled neighborhood effects. I also suspect that other large service areas share a similarity with SRP in temperature variations across the service area. The daily highs and lows significantly differ across the service area to a point where it would be inappropriate not to control for this with the control group. Neighborhood stratification ensures that the temperatures experienced by the treatment and control groups as a whole are identical.

The neighborhood stratification design is simply determined by the volunteers in the treatment group(s). When a volunteer is recruited for the treatment group, a customer within a predefined neighborhood grouping (within a few blocks or few houses of the volunteer) who is closest in energy usage to the volunteer/treatment customer is selected to be the control customer for the neighborhood. In cases where the study involves the use of existing customers for a treatment group, a control customer is selected for comparison to a treatment customer as described above. For example, you may want to conduct a study on customers who are already participating in a program, so no recruitment is needed.

Neighborhood stratification is much more difficult than energy usage stratification, but I believe it's worth the effort to help improve the pretest comparability of treatment and control groups. At the very least, it's time to start discussing experimental design issues now to help improve future studies. Smart meter technology is setting the stage for an enormous increase in experiments testing the effects of programs on energy usage behavior. This is an ideal time for us to address experimental design issues and learn from each other, rather than further delay such a discussion until after many more millions of dollars have been spent on future studies. Aligning treatment and control groups at a neighborhood level fails to specifically control for socioeconomic and dwelling-type characteristics, but it does help reduce such group differences, and it's a much easier and cost-effective step in that direction.