Evaluation in Practice

A Methodological Approach

SECOND EDITION

Richard D. Bingham

Cleveland State University

Claire L. Felbinger

American University

CHATHAM HOUSE PUBLISHERS

SEVEN BRIDGES PRESS, LLC

NEW YORK · LONDON

PART III

Quasi-Experimental Designs

It is the absence of random assignments that distinguishes quasi-experiments from true experiments. Quasi-experiments employ comparison groups just as do experiments. In the quasi-experimental design, the researcher strives to create a comparison group that is as close as possible to the experimental group in all relevant respects. And, just as there are a variety of experimental designs, there are a similar variety of basic designs that fall into the quasi-experimental category. These include the pretest-posttest comparison group design, the regression-discontinuity design, the interrupted time-series comparison group design, and the posttest-only comparison group design.

Laura Langbein identifies a number of ways in which quasi-experiments are distinct from randomized experiments in practice. First, quasi-experiments tend to be retrospective—that is, to occur after a program is in place. When the decision to evaluate comes after the program has been implemented, the option of conducting a true experiment is usually foreclosed.

Second, the internal validity of most quasi-experiments is usually more questionable than that of experiments. The researcher must identify and measure all the relevant characteristics of the experimental group and attempt to construct a similar comparison group. Successfully executing this task is frequently problematic (Langbein 1980, 87–89). Although quasi-experiments are not true experiments, they are far from useless. Peter Rossi and Howard Freeman are actually quite positive about them:

> The term "quasi-experiment" does not imply that the procedures described are necessarily inferior to the randomized controlled experiment in terms of reaching plausible estimates of net effects. It is true that, without randomization, equivalence . . . cannot be established with as much certainty. The possibility always remains that the outcome of a program is really due to a variable or process that has not been considered explicitly in the design or analysis. However, quasi-experiments, properly constructed, can provide information on impact that is free of most, if not all, of the confounding processes (threats to validity). . . . Indeed, the findings from a properly executed quasi-experimental design can be more valid than those from a poorly executed randomized experiment. (1985, 267)

References

Langbein, Laura Irwin. 1980. *Discovering Whether Programs Work: A Guide to Statistical Methods for Program Evaluation.* Glenview, Ill.: Scott, Foresman.

Rossi, Peter H. and Howard E. Freeman. 1985. *Evaluation: A Systematic Approach.* 3d ed. Beverly Hills, Calif.: Sage.

CHAPTER 8

Pretest-Posttest Comparison Group Design

THE PRETEST-POSTTEST comparison group design, sometimes referred to as the nonequivalent group design, is similar in every respect to the pretest-posttest control group design except that the individuals (cities, organizations) in the program being evaluated and in the comparison group are not assigned randomly—the comparison group is nonequivalent. The researcher attempts to identify a group for comparison comparable in essential respects to those in the program. The validity of the quasi-experiment depends in large part on how closely the comparison group resembles the experimental group in all essential respects. Table 8.1 illustrates the design.

Obviously a multitude of uses exist for this design, and, in fact, it is one of the designs most frequently found in the literature. If random assignment is not possible, researchers attempt to find a group similar to the experimental group to use as a "pseudo-control." In many cases this is extremely difficult. In fact, it is so difficult that one research scholar, Barker Bausell, recommends against its use, as follows:

> Although this is probably a minority opinion, I recommend against the use of nonequivalent control groups, especially for beginning researchers. In many ways, quasi-experimental research (as these models are sometimes called) is more difficult to perform well than research involving the random assignment of subjects. Such research is also seldom conclusive, since it is subject to so many alternative explanations. (1986, 138)

One way in which evaluators seek to establish a comparison group as similar as possible to the experimental group is through the matched-pair design. This approach assigns subjects to experimental and "control" conditions on the basis of some common characteristics the evaluator wishes to measure. For example, in education research, students in an experimental group might be matched with other students on the basis of grade and reading level to form a comparison group. Only these two factors are controlled. Other factors that could affect the outcome (home, environment, intelligence) are not controlled.

When is this design most often used? It is used in retrospective analysis when it is too

TABLE 8.1
The Pretest-Posttest Comparison Group Design

	Pretest	Program	Posttest
Group E	O	X	O
Group C	O		O

late to assign a control group. It is also frequently used in evaluating programs in which the participants are volunteers and thus a true control group cannot be found.

Finding an adequate comparison group for a volunteer program is precisely the problem that was facing Tim Newcomb in the article to follow: "Conservation Program Evaluations: The Control of Self-Selection Bias."

Supplementary Readings

Judd, Charles M., and David A. Kenny. 1981. *Estimating the Effects of Social Interventions.* Cambridge, England: Cambridge University Press, chap. 6.

Mohr, Lawrence B. 1988. *Impact Analysis for Program Evaluation.* Chicago: Dorsey, chap. 4.

Rossi, Peter H., Howard E. Freeman, and Mark W. Lipsey. 1999. *Evaluation: A Systematic Approach.* 6th ed. Thousand Oaks, Calif.: Sage, chap. 9.

READING

Conservation Program Evaluations
The Control of Self-Selection Bias

Tim M. Newcomb
Seattle City Light

The Seattle City Light Department evaluated its weatherization program for low-income homeowners to determine how much electricity was conserved, and whether the program was cost-effective. The study employed a control group of future program participants in order to avoid the biased estimate that could result because participants are a voluntary, nonrandom group of utility customers. Surveys, energy audits, and electricity meter data confirmed the similarity of the experimental and control groups prior to home weatherization. Following the installation of conservation measures, the consumption of the experimental group declined by an average of 3400 Kwh per year, when compared with the control group. This experimental design is believed to have nullified the potential biases due to selection, testing, regression, history, and instrumentation. The bias due to mortality, defined in this study as customer turnover, was seen to be very minor. This design can be applied to evaluate programs that operate continuously, provided that the guidelines for the program do not change and that clearly defined measurements of program success are available.

PUBLIC AND PRIVATE electric utilities across the United States have developed a variety of conservation programs for their residential customers during the past five years. This effort has been spurred in part by the federal Residential Conservation Service, which began in 1979 and requires the larger utilities to offer a home energy audit service to their customers. From the point of view of the utilities, these programs are beneficial to the extent that they reduce the residential load in a cost effective way.

Precise analyses of the cost-effectiveness of these programs are difficult to make because relatively little is known about the amount of electricity conserved by the programs. Several recent reviews have noted the lack of accurate research on residential electricity conservation and have called for better evaluation strategies to measure conservation (Berry, 1981; Pease, 1982). In the early stages of conservation program planning, estimates of potential electricity savings came from engineering studies. Recent evaluation studies by Seattle City Light

(Weiss and Newcomb, 1982; Newcomb, 1982) and by Oak Ridge National Laboratory (Hirst et al., 1983a) found that conservation estimates developed prior to program implementation were considerably larger than the observed amount of conservation.

Inaccuracies in engineering estimates may be due, in part, to the tremendous variation in electricity consumption patterns among residential customers. Olsen and Cluett (1979) and Hirst et al. (1983a) report that some of this variation can be explained by differences in home size, income level, and number of occupants. More than half of the variation among households remains after these factors are accounted for. Engineering studies, which are conducted in a few homes, cannot take into account the complex relationships among unique home dwellers, unique homes, and varying combinations of energy conservation measures. For these reasons, accurate estimation of the conservation achieved by a residential program must be based upon actual field measurements of electricity use by the participating population.

The purpose of this article is to describe a quasi-experimental design for estimating average household electricity conservation that controls for the major threats to internal validity and does not require complex statistical methods for analysis. The design uses a nonequivalent control group as described by Campbell and Stanley (1966) consisting of residential customers who participated in the same conservation program at a later date. This research strategy is useful in situations where a program operates over a period of several years and maintains constant and restrictive guidelines for accepting participants. Under these conditions, early participants will resemble later ones with respect to important preprogram characteristics.

This approach to estimating electricity conservation has been applied to three residential conservation program evaluations at Seattle City Light. The evaluation of electricity savings for the Low Income Electric Program (LIEP) will be presented here as an example of the approach.

Seattle City Light serves approximately 270,000 residential customers and plans to weatherize 20,000 homes by 1989 through the LIEP. This is expected to yield between 60 and 70 million kilowatt-hours of conserved electricity each year. The predicted cost of the weatherization over 8 years is 25 million in 1980 dollars. The LIEP offers a free home energy audit and free home weatherization to applicants whose total annual household income is less than 90% of the median income for the Seattle-Everett Standard Metropolitan Statistical Area. All applicants must accept ceiling and underfloor insulation as needed, as well as duct insulation, hot-water tank insulation, and repairs that are judged necessary by the utility auditors. Wall insulation, pipe insulation, and weatherstripping and caulking are optional. The participants included in this analysis all had electric heat as their primary source of space heat. The average 1981 LIEP participant received approximately $1400 in weatherization materials and labor at no cost.

The average cost to Seattle City Light for a weatherization job was approximately $2500 in 1981, including administration, home energy audits, and inspection of weatherization. An accurate assessment of the electricity savings is essential to evaluate and justify such large expenditures.

Reasons for Selecting LIEP Participants as a Control Group

A simple pre-post experimental design could not provide accurate program conservation estimates for several important reasons. Weather fluctuations from one year to the next have a strong impact upon electricity use in homes that use electric space heat, and could easily mask or magnify the effects of home weatherization. Seattle City Light increased its residential rates approximately 65% in August of 1980, causing its customers to decrease their consumption of electricity. This effect coincided with the LIEP weatherization and, while the rate hike cannot be separated cleanly from the effects of the program, it probably produced reductions in electricity use beyond the results of weatherization. There were no time series data available of sufficient quality to allow the estimation of weather and rate hike effects using multiple regression.

All of these reasons demonstrate the need for the inclusion of a control group in this study that is as similar as possible to the LIEP participants. No group of residential customers fitting this description had been selected for analysis prior to the program, and current household income data were not available for any group except participants. Further, since the program is voluntary, participants almost certainly differ from the average low-income customer in attitudes and education. Recent studies by Olsen and Cluett (1979) and by Berry (1981) report that participants in voluntary residential conservation programs tend to have more education and use more electricity than the average customer. Without knowing beforehand exactly what the attributes of the program participants were, the best choice for a control group was the group of people who would later sign up for the program. This choice was confirmed through a comparison of survey data and electricity consumption figures for the LIEP weatherized homes and the control group to be presented in a later section.

A description of the experimental and the control groups, and the general approach to calculating electricity conservation estimates from the program are outlined as follows:

- The program participants, or "experimental group," received home conservation measures in 1981. The electricity use for this group prior to the program (September 1980 to May 1981) will be referred to as E_1. The electricity use for this group after the program weatherized the homes (September 1981 to May 1982) is referred to as E_2.
- The "control group" of homeowners received home conservation measures in 1982. The electricity consumption during the period from September 1980 to May 1981 will be called C_1. The electricity in use from September 1981 to May 1982 for this group is C_2. In this case, C_2 represents electricity use for the control group before the homes were weatherized.
- The computation of the estimates of conservation due to the program involves computing the pre-to-post change in consumption for the experimental group, $E_2 - E_1$, and the corresponding change for the control group, $C_2 - C_1$. The latter difference is subtracted from the former to find the estimate of conserved electricity, which is $(E_2 - E_1) - (C_2 - C_1)$.

Estimating the Impact of Weatherization on Electricity Use

Comparison of the Preprogram Attributes of the Weatherized Homes and the Control Homes

Three sources of information about the experimental and control groups of homes were available. Data collected at the time of the home energy audit describe the age and size of the home, the number of occupants, and the type of space heating. A mail survey of a random sample of each of the groups gathered information on the conservation actions taken in addition to the LIEP, and on the reasons for the homeowner's participation. Finally, the utility's customer metering system provided electricity consumption figures for each customer in periods of two months. Each of these sources can be used to test the assumption that the two groups were similar before the weatherization occurred.

Table 1 compares the groups with respect to average heated floor area, percentage of homes having each type of electric space heat, and average number of nighttime occupants. The reports by Hirst et al. (1983a) and by Olsen and Cluett (1979) showed that home size and number of occupants were positively correlated with level of consumption. The fact that there is little difference between the experimental and control groups for these two measures is strong support for the argument that the groups are similar. The close similarity in types of space heating is also significant because different types of space heating respond differently to fluctuations in temperature.

The survey responses to the mailed questionnaire are shown in table 2. The response rates to these surveys were 81% and 76% for the experimental group and the control group, respectively, indicating that the responses are representative of the respective survey populations sampled. There is little difference with regard to the respondents' stated reasons for participating in LIEP, except for a small increase in those who mentioned saving on electricity bills in 1982. The percentage of respondents who took each of the four conservation actions in addition to the LIEP is also very similar, which shows that any measured effects of the program weatherization are not affected by differences in extraprogram weatherization. The two groups appear to resemble each other with respect to their motivation and their reliance on the program for changes in home weatherization.

The single most critical test of the similarity of the two groups must be their respective levels of electricity use during the same time period before the experimental group received weatherization. Any differences in either the amount of electricity used or in the response curves of electricity use to weather would require adjustments before the impact of the program could be measured. Table 3 presents the average electricity consumption figures and the standard errors for these averages, for five two-month billing periods prior to LIEP participation. The differences between the groups are not significant at the 5% level for any of the five two-month periods. Since electricity use is

TABLE 1

LIEP Participants, 1981 and 1982: Selected Demographic and Dwelling Characteristics

	1981 Participants Experimental Group (N = 326)	*1982 Participants Control Group (N = 227)*	*t-Tests of Differences Between Groups Two-Tailed*
Floor area of home			
Mean square feet	1258	1323	t = –1.26
Standard Error	35	38	p > .10
Number of occupants			
Mean number	2.6	2.5	t = 0.71
Standard error	0.1	0.1	p > .10
Type of electric heat			
% Furnaces	20.4	21.6	t = –0.34
% Baseboard	70.6	73.6	p > .50
% Other	3.5	3.0	t = –0.77
			p > .10
			t = 0.33
			p > .50

the criterion variable in this analysis, the similarity offers strong support for the choice of later participants as the control group.

The evidence for similarity was judged strong enough to support the estimation of electricity conservation using the calculations described earlier. A further discussion of the advantages and disadvantages of this design follows the estimation of the annual program conservation.

Calculation of the Electricity Conservation Estimates

The description of the quasi-experimental design that has been developed up to this point is oversimplified. This is so because homes were weatherized under LIEP continuously during 1981 and 1982, rather than in a short period. Figure 1 shows that as homes were weatherized in 1981, they were dropped from the preprogram consumption averages, and added to the postprogram averages. As homes in the control group were weatherized in 1982, they were dropped from the control group consumption averages.

All of the preprogram bimonthly billing records for the experimental group were grouped by two-month periods and an average was computed for each period, although the number of cases in each preprogram period changed (decreased) as homes received weatherization and were removed from the preprogram group. In a similar fashion, all of the electricity billing data for postprogram periods for the experimental group were

TABLE 2

LIEP participants, 1981 and 1982: Survey Responses – Reasons for Participating in LIEP and Conservation Actions Taken Outside LIEP

	1981 Participants Experimental Group (N = 197) %	*1982 Participants Control Group (N = 110)* %	*t-Tests of Differences Between Groups Two-Tailed*
Reasons for participating			
Program is free	70	70	$t = 0$
Save on bills	89	97	$t = -2.85$, $p < .01$
Believe in conservation	75	77	$t = -0.40$, $p > .50$
Increase value of home	55	58	$t = -0.51$, $p > .50$
Increase comfort of home	80	77	$t = 0.61$, $p > .50$
Help solve energy crisis	61	57	$t = 0.68$, $p > .10$
Conservation actions taken outside LIEP within 12 months of weatherization			
Installed thermal windows	10	12	$t = -0.54$, $p > .50$
Installed storm windows	16	21	$t = -1.06$, $p > .10$
Installed energy-efficient water heater	9	7	$t = 0.63$, $p > .50$
Installed solar water heater	1	1	$t = 0$

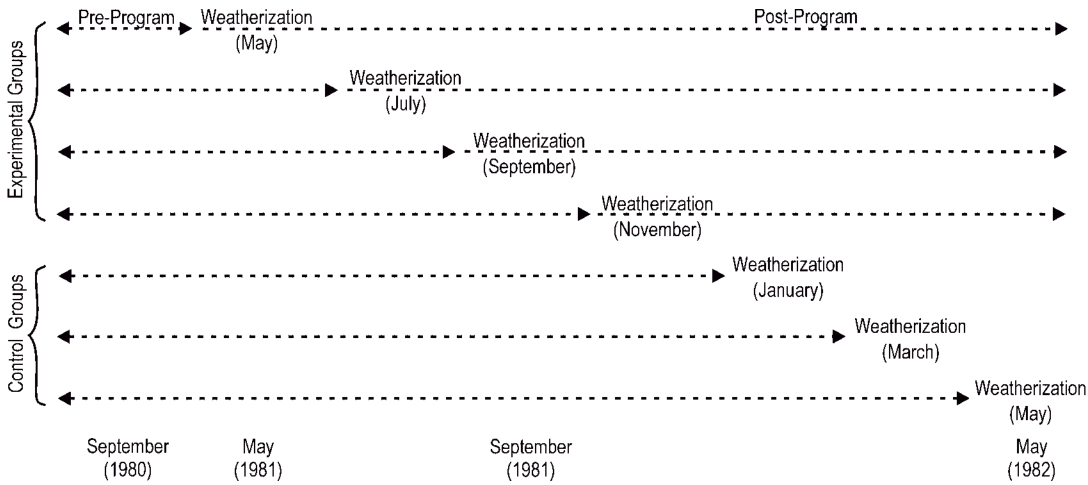

FIGURE 1

grouped by period and averaged. The number in each billing period changed (increased) over time as more homes were weatherized.

The control group remained stable throughout the preprogram period. However, as these homes entered the program in early 1982, they were dropped from the control group, and the number of cases available for control purposes declined during successive billing periods in 1982 through May of 1982, which was the last period used in the analysis.

The varying numbers of homes in the two groups is a problem that will occur in a design of this type unless the treatment is applied to the entire group of participants in a brief period. The evaluation of programs operating in a natural setting is more likely to face the continuous mode of program operation described here. An accurate system for tracking and assigning participants to one or the other of the two groups as the treatment occurs is helpful for a design of this kind.

Figure 2 is a graph of each two-month electricity consumption average for the experimental group and the control group using the method described above. While the two groups are similar during the period up to September 1980, they diverge after that time. The small difference during the September 1980 period is due to the relatively minor impact of weatherization during summer months when consumption for space heating is very low. The number of heating degree days during the period from November 1980 to April 1981, was only 87% of the number of heating degree days from November 1981 to April 1982. The colder winter of 1981–1982 had an obvious impact upon the consumption of the control group homes in Figure 2. This increased consumption was not shared by the experimental group; however, a simple pre-post design would have reduced the estimate of electricity conservation over this period for the experimental group because the colder winter raised the consumption of electricity by comparison with the previous year.

Table 3 displays the mean values for the experimental and control groups, before and after the experimental group received weatherization. The t-test values for each of the ten differences between mean values are listed in the third column with the probability that such a value would occur by chance. One-tailed t-tests were used for both the before and after periods, because the hypothesis to be tested is that the experimental group does not have a lower bi-

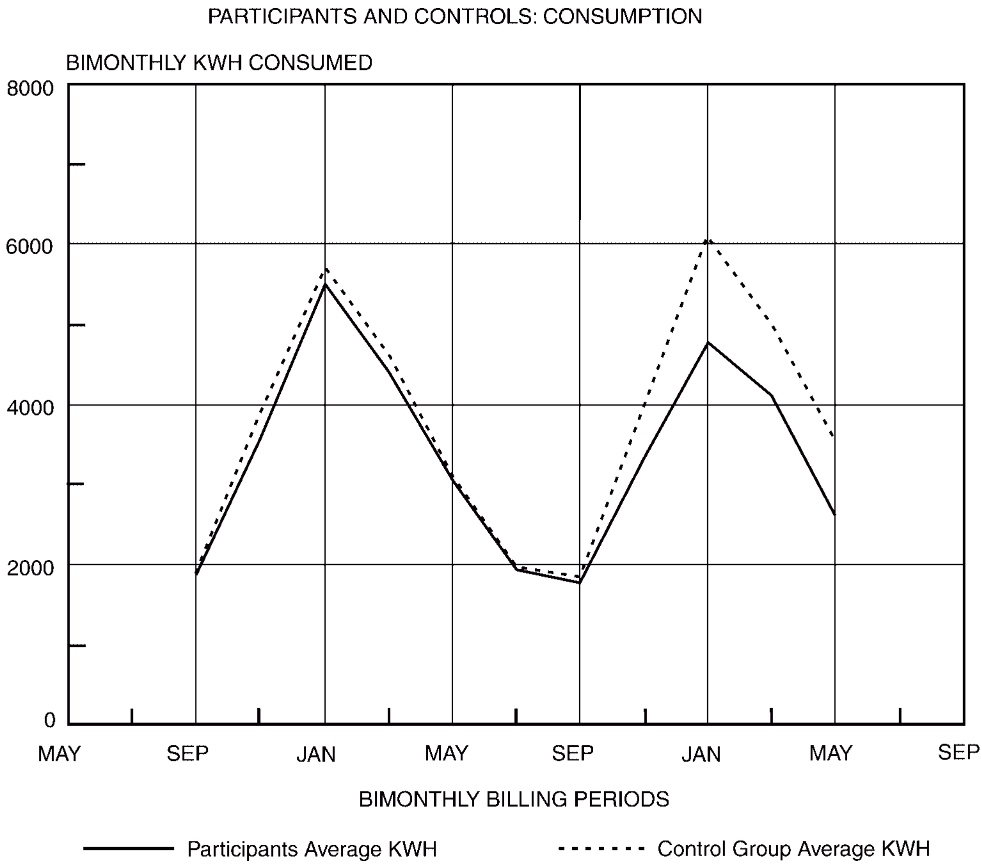

FIGURE 2
LIEP Electricity Consumption (KWH) 1980–1982

monthly electricity level at each period than the control group. While all of the preperiod differences between means are insignificantly different from zero, four of the five postperiod differences are highly significant.

The actual computations of electricity conservation are presented in Table 4. The first column contains the pre-to-post changes by billing period for the experimental group, referred to earlier as E_2 - E_1. The corresponding differences for the control group are given in column 2, and referred to as C_2 - C_1. The NET CHANGE, $(E_2 - E_1) - (C_2 - C_1)$, is computed in column three. The total net change for the ten-month period under study is -3083Kwh. To compute an estimated annual amount of electricity conservation, the ratio of electricity use by the control group for one full year (July 1980 to May 1981) to that group's use for ten months (September 1980 to May 1981), equal to 1.11, was multiplied by -3083 Kwh/10 months. The result is -3422 Kwh/year of electricity conservation. The value of the conserved electricity to the average LIEP participant is $485 (1981 dollars) over the 30-year lifetime of the weatherization measures (Weiss, 1983; this is also the net present value to the customer since the weatherization was free).

TABLE 3

LIEP Participants, 1981 and 1982: Electricity Consumption Before and After Weatherization of the Experimental Group

Bimonthly Billing Period	*1981 Participants Experimental Group*	*1982 Participants Control Group*	*t-Tests of Differences Between Groups One-Tailed*
Preperiod			
September 1980			
Mean Kwh/2 months	1893	1902	t = –0.08
Standard Error	98	55	p > .05
N	68	201	
November 1980			
Mean Kwh/2 months	3686	3829	t = –0.70
Standard Error	160	128	p > .05
N	113	208	
January 1981			
Mean Kwh/2 months	5195	5434	t = –1.13
Standard Error	155	144	p > .05
N	185	210	
March 1981			
Mean Kwh/2 months	4389	4567	t = –1.11
Standard Error	105	121	p > .05
N	293	209	
May 1981			
Mean Kwh/2 months	3010	3114	t = –0.91
Standard Error	76	86	p > .05
N	293	209	
Postperiod			
September 1981			
Mean Kwh/2 months	1780	1871	t = –0.85
Standard Error	91	57	p > .05
N	68	209	
November 1981			
Mean Kwh/2 months	3232	3960	t = –3.95
Standard Error	137	123	p < .05
N	113	209	
January 1982			
Mean Kwh/2 months	5031	6169	t = –5.03
Standard Error	156	164	p < .01
N	185	208	
March 1982			
Mean Kwh/2 months	4116	5027	t = –4.92
Standard Error	98	157	p < .01
N	293	160	
May 1982			
Mean Kwh/2 months	2631	3519	t = –4.15
Standard Error	70	202	p < .01
N	293	39	

TABLE 4

LIEP Participants, 1981 and 1983: Computation of Electricity Conservation from Bimonthly Changes in Consumption of the Experimental and Control Groups

Bimonthly Comparison Periods	*Changes for the 1981 Participants Experimental Group* $E_2 - E_1$	*Changes for the 1982 Participants Control Group* $C_2 - C_1$	*Net Change Due to LIEP Program Weatherization* $(E_2 - E_1) - (C_2 - C_1)$
September 1981 minus September 1980	–113 Kwh/2 mo.	–31 Kwh/2 mo.	–82 Kwh/2 mo.
November 1981 minus November 1980	–454	131	–585
January 1982 minus January 1981	–164	735	–899
March 1982 minus March 1981	–273	460	–733
May 1982 minus May 1981	–379	405	–784

Note: Net change over ten-month period: –3083 Kwh.

Discussion

From the point of view of quasi-experimental design theory, the design described here has some significant strengths. The assumption that the control and experimental groups are alike can be examined carefully. The somewhat unique group of customers that participated in the LIEP can be compared to another very similar group without performing a multiple-regression analysis of the relationship between income levels, electricity consumption, and weatherization levels.

Most of the threats to internal validity described by Campbell and Stanley (1966) can be discounted in this study. The problem of selection in the LIEP study is a recurrent one in residential energy conservation research (Hirst et al., 1983b). Since voluntary participants may be unlike a random sample of customers, this difference must either be controlled through the proper design, or counteracted through statistical techniques. The nonequivalent control group design used here effectively dispenses with this threat through its comparison of two groups of customers who have volunteered for the same program during adjacent time periods.

The problem of mortality in the experimental group has its counterpart here in those LIEP participants who had to be excluded from analysis either because they moved into homes shortly before the weatherization, or moved out shortly after the weatherization. For these cases there were not enough two-month billing periods under one owner to permit pre-post comparison. Approximately 10% of the controls and 13% of the experimentals were discarded for this reason. An examination of the impact of change of ownership on consumption shows that in 50% of the cases the level of electricity consumption changed by more than 25% with the ownership change. Given this large impact, a comparison of preprogram and postprogram levels regardless of ownership is not possible. These large changes underscore the high level of variability in electricity use among the single-family customers of Seattle City Light.

Table 5 presents preprogram consumption data for the experimental group and for a sample of homes excluded due to ownership changes. The comparison suggests that the excluded cases may have used slightly less electricity than the cases that were included in the analysis. The result of this exclusion may have been to overestimate the average conservation due to the program.

The use of future program participants has limitations for program evaluation that stem from the method's dependence upon constant program guidelines for participation. As the requirements for entry into a program change, the possibility of finding a similar group of future participants disappears.

The design employed in this analysis can only be used when a program operates continuously under relatively unchanging guidelines for a sufficient period to permit the accumulation of large groups for statistical analysis. Even under these conditions, such a design will be difficult to defend if the dependent variables are qualitative, or less well-defined than bimonthly electricity consumption. In such cases, it must be defended on theoretical grounds rather than by comparing preprogram characteristics of the experimental and control groups.

TABLE 5

The LIEP Experimental Group and Excluded LIEP Cases Average Preprogram Electricity Use for Four Billing Periods

Bimonthly Billing Period	*1981 Participants Experimental Group*	*1981 Participants Excluded*	*t-Tests of Differences Between Groups Two-Tailed*
September 1980			
Mean Kwh/2 months	1893	1840	t = 0.29
Standard Error	98	155	p > .50
N	68	20	
November 1980			
Mean Kwh/2 months	3686	3837	t = –0.41
Standard Error	160	336	p > .50
N	113	24	
January 1981			
Mean Kwh/2 months	5195	4870	t = 0.90
Standard Error	155	328	p > .10
N	185	28	
March 1981			
Mean Kwh/2 months	4389	3928	t = 1.55
Standard Error	105	278	p > .10
N	293	25	

References

Berry, L. (1981) *Review of Evaluations of Utility Home Energy Audit Programs*, ORNL/CON-58. Tennessee: Oak Ridge National Laboratory.

Campbell, D. T. and J. C. Stanley (1966) *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally.

Hirst, E., E. Broneman, R. Goeltz, J. Trimble, and D. Lerman (1983a) *Evaluation of the BPA Residential Weatherization Pilot Program*, ORNL/CON-124. Tennessee: Oak Ridge National Laboratory.

____J. Trimble, R. Goeltz, and S. Cardell (1983b) *Methods to Deal with Self-Selection in Estimating Conservation Program Energy Savings: Use of Synthetic Data to Test Alternative Approaches*, ORNL/CON-120. Tennessee: Oak Ridge National Laboratory.

Newcomb, T. M. (1982) *Electricity Conservation Estimates for the Low Income Electric Program.* Washington: Seattle City Light Department.

Olsen, M. and C. Cluett (1979) *Evaluation of the Seattle City Light Neighborhood Energy Conservation Program.* Seattle, WA: Battelle Human Affairs Research Center.

Pease, S. (1982) *An Appraisal of Evaluations of Utility-Sponsored Programs for Residential Energy Conservation*, N-1925-DOE. Washington, DC: U.S. Department of Energy.

Weiss, C. (1983) *Low Income Electric Program Cost-Effectiveness Analysis.* Washington: Seattle City Light.

____and T. M. Newcomb (1982) *Home Energy Loan Program Energy Savings Analysis.* Washington: Seattle City Light.

Tim M. Newcomb is a Program Evaluator in the Conservation Division of Seattle City Light, Seattle, Washington. His current interests are in using Arima Time Series models to estimate electricity conservation, and in the use of surveys to market conservation programs.

Explanation and Critique

Developing an evaluation is an exercise of the dramatic imagination. The art of evaluation involves using the science of evaluation to create a design and gather information that is appropriate for a specific situation and a particular policymaking context. The quasi-experiment with constructed controls (comparison group) provides the researcher with the opportunity to exercise artistic creativity. Tim Newcomb rose to this challenge. "Conservation Program Evaluations: The Control of Self-Selection Bias" is absolutely elegant.

Seattle City Light operated a Low Income Electric Program (LIEP) that offered a free energy audit and free home weatherization to applicants whose total annual household income was less than 90 percent of the median income for the Seattle-Everett Standard Metropolitan Statistical Area. The average 1981 LIEP participant received approximately $1,400 in weatherization materials at no cost. The average cost to Seattle Light for each job was $2,500, including administration and energy audits. The company planned to weatherize 20,000 homes by 1989 through LIEP at a cost of $25 million in 1980 dollars. Given such cost, it was imperative to develop an approach to accurately estimate electricity conservation that could be attributed to the program. How could this be done?

Because the program was voluntary, Newcomb could not use an experimental design and randomly assign low-income utility customers into program and control groups. Furthermore, a comparison of low-income utility participants with nonparticipants would be inappropriate because simply by volunteering for the program a select group of low-income utility customers showed their concern with utility costs. They thus might be assumed to conserve more in response to the 1980 rate increase than nonvolunteers. (This rate increase of some 65 percent was an effect of history which could have had an impact on the evaluation without the comparison group.) The simple pretest-posttest design could not be used because of year-to-year fluctuations in the weather. Nor were time-series data available to allow the estimation of weather and rate hike effects using multiple regression.

Thus, Newcomb was faced with the need to find a comparison group for the study as similar as possible to LIEP participants. Obviously, no true control group was available. He described his predicament:

> No group of residential customers fitting this description had been selected for analysis prior to the program, and current household income data were not available for any group except participants. Further, since the program is voluntary, participants almost certainly differ from the average low-income customer in attitudes and educations.

Newcomb found his comparison group and overcame the self-selection problem by using as "constructed controls" the group of people who later signed up for the program. The characteristics of the comparison group in terms of housing, income, preweatherization electricity use, and so on were proven to be comparable. Most importantly, the self-selection problem was overcome through Newcomb's unique design. The pretest-posttest comparison group design used in this study effectively dispensed with the selection threat to validity through its comparison of two groups of customers who volunteered for the same program during adjacent time periods.

Substantively, what did this evaluation show? Newcomb compared the energy consumption of both groups for the before-and-after periods by using the one-tailed *t*-test. The *t*-test was used to determine whether the average energy consumption of weatherized residents was different from the average consumption by those not weatherized. Thus Newcomb used a difference of means test with the *t*-statistic to determine whether there was a statistically significant difference between the means. In general, the *t*-test is an inferential statistic used to determine whether the null hypothesis, the test or alternate hypothesis (that there is a difference) is accepted. The one-tailed *t*-test was appropriate in this case because Newcomb's hypothesis was that the experimental group would have lower bimonthly electricity consumption rates than the control group. When a direction (lower) is implied in a hypothesis (rather than just "different"), a one-tailed test is dictated.

Newcomb found that there was no significant difference between the energy consumption of the two groups in all the pretest periods. After the homes of the experimental group had been weatherized, however, there was a significant difference in energy consumption between the two groups for four of the five periods. It was only in the warmest months that the differences between the two groups was not significant.

By using this carefully selected comparison group, Newcomb was able to show that the program resulted in a savings of 3,422 kilowatt hours per year for the average participant. The value of the conserved electricity (1981 dollars) amounted to $485 per participant over the 30-year lifetime of the weatherization measures. Given variations in the temperature from winter to winter, Newcomb never could have made these calculations without having a credible comparison group. More simply, because winters are different, the energy consumption of the families participating in the program will vary from year to year; therefore, it is not possible to simply compare this year's energy use with last year's.

Reference

Bausell, R. Barker. 1986. *A Practical Guide to Conducting Empirical Research.* New York: Harper and Row.