



Chapter 17: Residential Behavior Protocol

The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures

Created as part of subcontract with period of performance
September 2011 – December 2014

James Stewart
The Cadmus Group
Waltham, Massachusetts

Annika Todd
Lawrence Berkeley National Laboratory
Berkeley, California

NREL Technical Monitor: Charles Kurnik

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Subcontract Report
NREL/SR-7A40-62497
January 2015

Contract No. DE-AC36-08GO28308

Chapter 17: Residential Behavior Protocol

The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures

Created as part of subcontract with period of performance
September 2011 – December 2014

James Stewart
The Cadmus Group
Waltham, Massachusetts

Annika Todd
Lawrence Berkeley National Laboratory
Berkeley, California

NREL Technical Monitor: Charles Kurnik

Prepared under Subcontract No. LGJ-1-11965-01

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at <http://www.osti.gov/scitech>

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: <mailto:reports@adonis.osti.gov>

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/help/ordermethods.aspx>

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

NREL prints on paper that contains recycled content.

Acknowledgments

The chapter author wishes to thank and acknowledge Bill Provencher and Kevin Cooney of Navigant, as well as Northeast Energy Efficiency Partnerships and OPower, for their thoughtful contributions.

Acronyms

BB	behavior-based
RCT	randomized control trial
RED	randomized encouragement design
ITT	intent-to-treat
D-in-D	difference-in-differences
TOT	treatment effect on the treated
OLS	Ordinary Least Squares
LATE	local average treatment

Table of Contents

1	Measure Description	1
2	Applicability Conditions of Protocol	2
2.1	Examples of Protocol Applicability	3
3	Savings Concepts	5
3.1	Definitions	5
3.2	Experimental Research Designs	6
3.3	Basic Features	7
3.3.1	Common Features of Randomized Control Trial Designs	7
3.4	Common Designs	9
3.4.1	Randomized Control Trial With Opt-Out Program Design	9
3.4.2	Randomized Control Trial With Opt-In Program Design	10
3.4.3	Randomized Encouragement Design	12
3.4.4	Persistence Design	14
3.5	Evaluation Benefits and Implementation Requirements of Randomized Experiments	15
4	Savings Estimation	17
4.1	Sample Design	17
4.1.1	Sample Size	18
4.1.2	Random Assignment to Treatment and Control Groups by Independent Third Party	18
4.1.3	Equivalency Check	19
4.2	Data Requirements and Collection	19
4.2.1	Energy Use Data	19
4.2.2	Makeup of Analysis Sample	20
4.2.3	Other Data Requirements	20
4.2.4	Data Collection Method	21
4.3	Analysis Methods	21
4.3.1	Panel Regression Analysis	21
4.3.2	Panel Regression Model Specifications	22
4.3.3	Simple Differences Regression Model of Energy Use	22
4.3.4	Simple Differences Regression Estimate of Heterogeneous Savings Impacts	23
4.3.5	Simple Differences Regression Estimate of Savings During Each Time Period	24
4.3.6	Difference-in-Differences Regression Model of Energy Use	25
4.3.7	D-in-D Estimate of Savings for Each Time Period	26
4.3.8	Randomized Encouragement Design	27
4.3.9	Models for Estimating Savings Persistence	28
4.3.10	Standard Errors	29
4.3.11	Opt-Out Subjects and Account Closures	30
4.4	Energy Efficiency Program Uplift and Double Counting of Savings	31
5	Reporting	34
	References	35

List of Figures

Figure 1. Illustration of RCT with opt-out program design.....	9
Figure 2. Illustration of RCT with opt-in program design	11
Figure 3. Illustration of RED program design	12
Figure 4. Example of D-in-D regression savings estimates	26
Figure 5. Calculation of double-counted savings.....	32

List of Tables

Table 1. Benefits and Implementation Requirements of Randomized Experiments.....	15
Table 2. Considerations in Selecting an Experimental Design	16

1 Measure Description

Residential behavior-based (BB) programs use strategies grounded in the behavioral and social sciences to influence household energy use. These may include providing households with real-time or delayed feedback about their energy use; supplying energy efficiency education and tips; rewarding households for reducing their energy use; comparing households to their peers; and establishing games, tournaments, and competitions.¹ BB programs often target multiple energy end uses and encourage energy savings, demand savings, or both. Savings from BB programs are usually a small percentage of energy use, typically less than 5%.²

Utilities introduced the first large-scale residential BB programs in 2008. Since then, dozens of utilities have offered these programs to their customers.³ Although program designs differ, many share these features:

- They are implemented using a randomized experimental design, where eligible homes are randomly assigned to treatment or control groups.
- They are large scale by energy efficiency program standards, targeting thousands of utility customers.
- They provide customers with analyses of their historical consumption, energy savings tips, and energy efficiency comparisons to neighboring homes, either in personalized home reports or through a Web portal, or offer incentives for savings energy.
- They are typically implemented by outside vendors.⁴

Utilities will continue to implement residential BB programs as large-scale, randomized control trials (RCTs); however, some are now experimenting with alternative program designs that are smaller scale; involve new communication channels such as the Web, social media, and text messaging; or that employ novel strategies for encouraging behavior change (e.g., Facebook competitions).⁵ These programs will create new evaluation challenges and may require different evaluation methods than those currently employed to verify any savings they generate. Quasi-experimental methods, however, require stronger assumptions to yield valid savings estimates and may not measure savings with the same degree of validity and accuracy as experimental methods.

¹ See Ignelzi et al. (2013) for a classification and descriptions of different BB intervention strategies and Mazur-Stommen and Farley (2013) for a survey and classification of current BB programs.

² See Alcott (2011), Davis (2011), and Rosenberg et al. (2013) for savings estimates from residential BB programs.

³ See the 2013 Consortium for Energy Efficiency (CEE) database for a list of utility behavior programs; it is available for download: <http://library.cee1.org/content/2013-behavior-program-summary-public-version>.

⁴ Vendors that offer residential BB programs include Aclara, C3 Energy, Opower, and Simple Energy.

⁵ The 2013 CEE database includes descriptions of many residential BB programs with alternative designs such as community-focused programs, college dormitory programs, K-12 school programs, and programs relying on social media.

2 Applicability Conditions of Protocol

This protocol recommends the use of RCTs or randomized encouragement designs (REDs) for estimating savings from BB programs that satisfy the following conditions:⁶

- Residential utility customers are the target.
- Energy or demand savings are the objective.
- An appropriately sized analysis sample can be constructed.
- Accurate energy use measurements for sampled units are available.

The next section of this protocol carefully defines and explains these evaluation methods. A significant body of evidence indicates that randomized experimental approaches work; that is, they result in unbiased and robust estimates of program energy and demand savings.

This protocol applies only to residential BB programs. In theory, evaluators can apply the experimental methods recommended in this protocol to nonresidential BB programs, and there are examples of evaluators applying such methods.⁷ However, utilities have offered relatively few BB programs to nonresidential customers thus far. Thus, knowledge about the efficacy of evaluation methods in the nonresidential sector is lacking. As more evidence accumulates, the National Renewable Energy Laboratory could expand this protocol to include nonresidential programs.

This protocol also addresses best practices for estimating energy and demand savings. There are no significant conceptual differences between measuring energy savings and measuring demand savings; thus, evaluators can apply the algorithms in this protocol for calculating BB program savings to either. The protocol does not directly address the evaluation of other BB program objectives, such as increasing utility customer satisfaction, educating customers about their energy use, or increasing awareness of energy efficiency.⁸

This protocol also requires that the analysis sample be large enough to detect the expected savings with a high degree of confidence. Because most BB programs result in small percentage savings, a large number of sampled units are required to detect savings. This protocol does not address program evaluations conducted with insufficiently sized samples.

Finally, this protocol requires that the energy use of participants or households affected by the program (for the treatment and control groups) can be clearly identified and measured. Typically, the analysis unit is the household; in this case, treatment group households must be identifiable and individual household energy use must be measurable. However, depending on the BB

⁶ As discussed in Considering Resource Constraints in the Introduction of this UMP report, small utilities (as defined under the U.S. Small Business Administration regulations) may face additional constraints in undertaking this protocol. Therefore, alternative methodologies should be considered for such utilities.

⁷ For example, PG&E offers a Business Energy Reports Program, which it implemented as a field experiment. See http://beccconference.org/wp-content/uploads/2013/12/BECC_PGE_BER_11-19-13_seelig.pdf

⁸ Process evaluation objectives may be important, and omission of them from this protocol should not be interpreted as a statement that these objectives should not be considered by program administrators.

program, the analysis units may not be households. For example, for a BB program that generates an energy competition between hundreds of housing floors at a university, the analysis unit may be floors; in this case, the energy use measurement of individual floors must be available.

The characteristics of BB programs that *do not* determine the applicability of the evaluation protocol include:

- Whether the program is opt in or opt out⁹
- The specific behavior-modification theory or strategy
- The channel(s) through which program information is communicated.

This protocol does not recommend quasi-experimental methods to evaluate BB programs covered by this protocol. Evaluators of BB programs have employed quasi-experimental methods,¹⁰ but more knowledge about the efficacy of these methods is needed before they can be recommended.¹¹ As more evidence accumulates, the National Renewable Energy Laboratory may update this protocol as necessary.

Although this protocol strongly recommends RCTs or REDs for estimating savings from residential BB programs when the applicability criteria are met, it also recognizes that other considerations such as regulatory requirements or program objectives may take precedence, and evaluators may not always be able to apply these methods. In these cases, evaluators will have to employ quasi-experimental methods, which require stronger assumptions than do experimental methods to yield valid savings estimates. If these assumptions are violated, quasi-experimental methods may produce biased results. The extent of the biases in the estimates is not knowable *ex ante*, so results will be less reliable. Because there is currently not enough evidence of quasi-experimental methods that perform well, this protocol refrains from recommending non-RCT evaluation methods. A good reference for the application of quasi-experimental methods to behavior-based program evaluation is See Action (2012) or Cappers et al. (2013).

2.1 Examples of Protocol Applicability

Examples of residential BB programs for which the evaluation protocol applies follow:

- **Example 1:** A utility sends energy reports encouraging conservation steps to thousands of randomly selected residential customers.

⁹ In opt-in programs, customers enroll or select to participate. In opt-out programs, the utility enrolls the customers, and the customers remain in the program until they opt out. An example opt-in program is having a utility Web portal with home energy use information and energy efficiency tips that residential customers can use if they choose. An example opt-out program is sending energy reports to utility selected customers.

¹⁰ For example, see Harding and Hsiaw (2012), who use variation in timing of adoption of an online goal-setting tool to estimate savings from the tool.

¹¹ Allcott (2011) shows that a within-subject design using a pre-post comparison of monthly energy use of households receiving energy reports overestimates savings compared to difference-in-differences (D-in-D) estimation using treatment and control group subjects.

- **Example 2:** Several hundred residential customers enroll in a Wi-Fi-enabled thermostat pilot program offered by the utility.
- **Example 3:** A utility invites thousands of residential customers to use its Web portal to track their energy use in real time, set goals for energy saving, find ideas about how to reduce their energy use, and receive points or rewards for saving energy.
- **Example 4:** A utility sends voice, text, and email messages to thousands of residential utility customers encouraging—and providing tips for—reducing energy use during an impending peak demand event.

Examples of programs for which the protocol does not apply follow:

- **Example 5:** A utility uses a mass-media advertising campaign that relies on radio and other broadcast media to encourage residential customers to conserve energy.
- **Example 6:** A utility initiates a social media campaign (e.g., using Facebook or Twitter) to encourage energy conservation.
- **Example 7:** A utility runs a pilot program to test the savings from in-home energy-use displays, and enrolls too few customers to detect the expected savings.
- **Example 8:** A utility runs a BB program in a large college dormitory to change student attitudes about energy use. The utility randomly assigns some rooms to the treatment group. The dorm is master-metered.

The protocol does not apply to Example 5 or Example 6 because the evaluator cannot identify who received the messages. The protocol does not apply to Example 7 because too few customers are in the pilot to accurately detect energy savings. The protocol does not apply to Example 8 because energy-use data are not available for the specific rooms in the treatment and control groups.

3 Savings Concepts

This protocol applies to residential BB programs that satisfy the applicability conditions described in Section 2. RCTs or REDs and regression analysis of energy use for periods before and during the treatment for treatment and control group subjects are recommended for estimating energy or demand savings from BB programs. The protocol recommends RCTs and REDs because they yield unbiased and robust estimates of savings caused by the program; that is, net savings. Unless otherwise noted, all references in this protocol to savings are to net savings.

Section 3.1 defines some key concepts; Section 3.2 describes specific evaluation methods.

3.1 Definitions

Control group. In an experiment, the control group comprises subjects (e.g., utility customers) who do not receive the program intervention or treatment.

Experimental design.¹² Randomized experimental designs rely on observing the energy use of subjects who were randomly assigned to program treatments or interventions in a controlled process.

External validity. Savings estimates are externally valid if evaluators can apply them to different populations or different time periods from those studied.

Internal validity. Savings estimates are internally valid if the savings estimator is expected to equal the causal effect of the program on consumption.

Opt-in program. Utilities use opt-in BB programs if the customers must agree to participate, and the utility cannot administer treatment without consent.

Opt-out program. Utilities use opt-out BB programs if customers need not agree to participate. The utility can administer treatment without consent, and customers remain enrolled until they ask the utility to stop the treatment.

Quasi-experimental design. Quasi-experimental designs rely on a comparison group who is not obtained via random assignment. Such designs observe energy use and determine program treatments or interventions based on factors that may be partly random but not controlled.

Randomized Control Trial (RCT). An RCT yields an unbiased estimate of savings. Evaluators randomly assign subjects from a study population to a treatment group or a control group. Subjects in a treatment group receive one program treatment (there may be multiple treatments), while subjects in the control group receive no treatment. The RCT ensures that receiving the treatment is uncorrelated with the subjects' pre-treatment energy use, and that evaluators can attribute any difference in energy use between the groups to the treatment.

¹² When this protocol uses the term *experimental methods*, it refers to randomized experiments such as RCTs or REDs, not other experimental evaluation approaches such as natural experiments or quasi-experiments.

Randomized Encouragement Design (RED). In an RED, evaluators randomly assign subjects to a treatment group that receives *encouragement* to participate in a program or to a control group who does not receive encouragement. The RED yields an unbiased estimate of the effect on energy use of encouraging energy-efficient behaviors and the effect on customers who participate because of the encouragement.

Treatment. A treatment is an intervention administered through the BB program to subjects in the treatment group. Depending on the research design, the treatment may be a program intervention or encouragement to accept an intervention.

Treatment effect. This is the effect of the BB program intervention(s) on energy use for a specific population and time period.

Treatment group. The treatment group includes subjects who receive the treatment.

3.2 Experimental Research Designs

This section outlines experimental methods for evaluating BB programs. The most important benefit of an RCT or RED is that, if carried out correctly, the experiment results in an unbiased estimate of the program's causal impact.¹³ Unbiased savings estimates have internal validity. A result is internally valid if the evaluator can expect the value of the estimator to equal the savings caused by the program intervention. The principal threat to internal validity in BB program evaluation derives from potential selection bias about who receives a program intervention. RCTs and REDs yield unbiased savings estimates because they ensure that receiving the program intervention is uncorrelated with the subjects' energy use.

Experimental research designs may yield savings estimates that are applicable to other populations or time periods, making them externally valid. Whether savings have external validity will depend on the specific research design, the study population, and other program features.

A benefit of field experiments is their versatility: evaluators can apply them to a wide range of BB programs regardless of whether they are opt-in or opt-out programs. Evaluators can apply experimental methods to any program where the objective is to achieve energy or demand savings; evaluators can construct an appropriately sized analysis sample; and accurate measurements of the energy use of sampled units are available.

Experimental methods generally yield highly robust savings estimates that are not model dependent; that is, they do not depend on the specification of the model used for estimation.

The choice of whether to use an RCT or RED to evaluate program savings should depend on several factors, including whether it is an opt-in or opt-out program and the utility's tolerance for subjecting customers to the requirements of an experiment. For example, using an RCT for an opt-in program might require delaying or denying participation for some customers. A utility may prefer to use an RED to accommodate all the customers who want to participate.

¹³ List (2011) describes many of the benefits of employing randomized field experiments.

Implementing an RCT or RED design requires upfront planning. Program evaluation must be an integral part of the program planning process; this need will be evident in the experimental research design descriptions described in Section 3.3.

3.3 Basic Features

This section outlines several types of RCT research designs, which are simple but extremely powerful research tools. The core feature of RCT is the random assignment of study subjects (e.g., utility customers, floors of a college dormitory) to a treatment group that receives or experiences an intervention or to a control group that does not receive the intervention.

Section 3.3.1 outlines some common features of RCTs and discusses specific cases.

3.3.1 Common Features of Randomized Control Trial Designs

The key requirements of an RCT are incorporated into the following steps:

1. **Identify the study population:** The program administrator screens the utility population if the program intervention is offered to certain customer segments only, such as single-family homes. Program designers can base eligibility on dwelling type (e.g., single family, multifamily), geographic location, completeness of recent billing history, heating fuel type, utility rate class, or other energy use characteristics.
2. **Determine sample sizes:** The numbers of subjects to assign to the treatment and control groups depend on the type of randomized experiment (e.g., REDs and opt-out RCTs generally require more customers) and hypothesized savings. The number of subjects assigned to the treatment versus control groups should be large enough to detect the hypothesized program effect with sufficient probability.¹⁴

Evaluators can use a statistical power analysis to determine the number of subjects required. This results in minimum sample sizes for the treatment and control groups as a function of the hypothesized program effect, the coefficient of variation of energy use, the specific analysis approach that will be used (e.g., simple differences of means, a repeated measure analysis), and tolerances for Type I and Type II statistical errors.¹⁵ Most statistical software (including SAS, STATA, and R) now include packages for performing statistical power analyses. It is not uncommon for BB programs with expected savings of less than 5% to require thousands of subjects in the treatment and control groups.¹⁶

An important component of the random assignment process is to verify that the treatment and control groups are statistically equivalent or balanced in their observed covariates. At

¹⁴ The number of subjects in the treatment group may also depend on the size of the program savings goal.

¹⁵ A Type I error occurs when a researcher rejects a null hypothesis that is true. Statistical confidence equals 1 minus the probability of a Type I error. A Type II error occurs when a researcher accepts a null hypothesis that is false. Many researchers agree that the probability of a 5% Type I error and a 20% Type II error is acceptable. See List et al. (2010).

¹⁶ EPRI (2010) illustrates that, all else equal, repeated measure designs, which exploit multiple observations of energy use per subject both before and after program intervention, require smaller analysis sample sizes than other types of designs.

a minimum, evaluators should check before the intervention for statistically significant differences in average pre-treatment energy use and in the distribution of pre-treatment energy use between treatment and control homes.

3. **Randomly assign subjects to treatments and control:** Study subjects should be randomly assigned to treatment and control groups. To avoid the appearance of a conflict of interest and to ensure the integrity of the RCT, this protocol highly recommends that a qualified independent third party perform the random assignment. Also, to preserve the integrity of the experiment, customers must not choose their assignments. The procedure for randomly assigning subjects to treatment and control groups should be transparent and well documented.
4. **Administer the treatment:** The intervention must be administered to the treatment group and withheld from the control group. To avoid a Hawthorne effect, in which subjects change their energy use in response to observation, control group subjects should receive minimal information about the study. Depending on the research subject and intervention type, the utility may administer treatment once or repeatedly and for different durations. However, the treatment period should be long enough for evaluators to observe any effects of the intervention.
5. **Collecting data:** Data must be collected from all study subjects, not only from those who chose to participate or only from those who did not drop out of the study or experiment. Preferably, evaluators collect multiple pre- and post-treatment energy use measurements. Such data enable the evaluator to control for time-invariant differences in average energy use between the treatment and control groups to obtain more precise savings estimates. Step 6 discusses this in further detail.
6. **Estimate savings:**¹⁷ Evaluators should calculate savings as the difference in energy use or difference-in-differences (D-in-D) of energy use between the subjects who were initially assigned to the treatment versus the control group. To be able to calculate an unbiased savings estimate, evaluators must compare the energy use from the entire group of subjects who were originally randomly assigned to the treatment group to the entire group of subjects who were originally randomly assigned to the control group. For example, the savings estimate would be biased if evaluators used only data from utility customers in the treatment group who chose to participate in the study.

The difference in energy use between the treatment and control groups, usually called an intent-to-treat (ITT) effect, is an unbiased estimate of savings because subjects were randomly assigned to the treatment and control groups. The effect is an ITT because, in contrast to many randomized clinical medical trials, ensuring that treatment group subjects in most BB programs comply with the treatment is impossible. For example, some households may opt out of an energy reports program, or they may fail to notice or simply ignore the energy reports. Thus, the effect is ITT, and the evaluator should base

¹⁷ This protocol focuses on estimating average treatment effects; however, treatment effects of behavior programs may be heterogeneous. Costa and Kahn (2010) discuss how treatment effects can depend on political ideology and Allcott (2011) discusses how treatment effects can depend on pretreatment energy use.

the results on the initial assignment of subjects to the treatment group, whether or not subjects actually complied with the treatment.

The savings estimation approach should be well documented, transparent, and performed by an independent third party.

3.4 Common Designs

Section 3.1 describes some of the RCT designs commonly used in BB programs.

3.4.1 Randomized Control Trial With Opt-Out Program Design

One common type of RCT includes the option for treated subjects to opt out of receiving the program treatment. This design reflects the most realistic description of how most BB programs work. For example, in energy reports programs, some treated customers may ask the utility to stop sending them reports.

Figure 1 depicts the process flow of an RCT in which treated customers can opt out of the program. In this illustration, the utility initially screened utility customers to refine the study population.

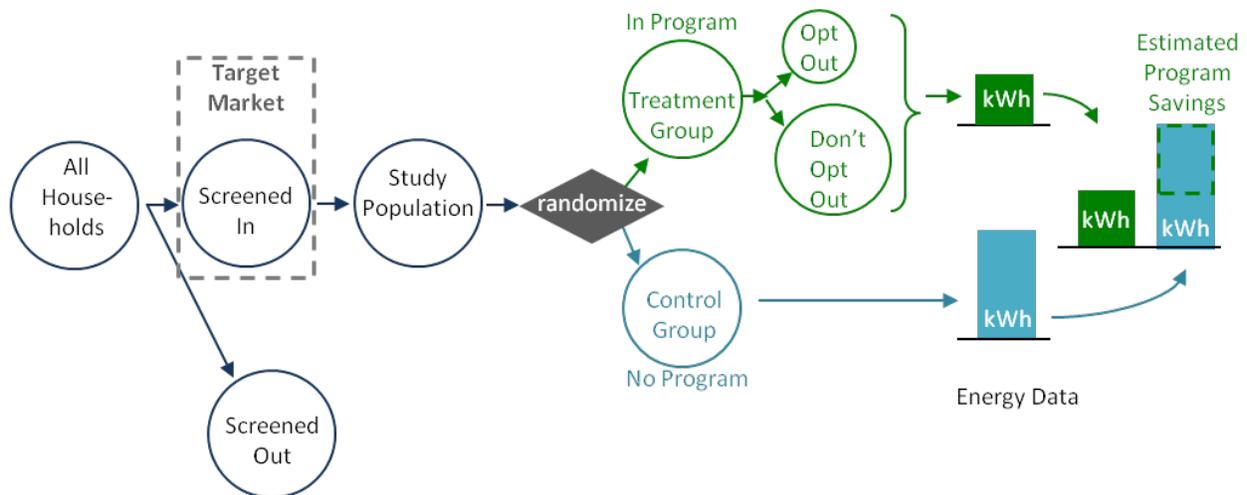


Figure 1. Illustration of RCT with opt-out program design¹⁸

Customers who pass the screening constitute the study population or sample frame. The savings estimate will apply to this population. Alternatively, the utility may want to study only a sample of the screened population, in which case a third party should sample randomly from the study population. The analysis sample must be large enough to meet the minimum size requirement for the treatment and control groups. The program savings goals and desired statistical power will determine the size of the treatment group.

¹⁸ This graphic and the following ones are variations of those that appeared in SEE Action (2012). A coauthor of the SEE Action report and the creator of that reports' figures is one of the authors of this protocol.

The next steps in an RCT with opt-out program design are to (1) randomly assign subjects in the study population to the program treatment and control groups; (2) administer the program treatments; and (3) collect energy use data.

The distinguishing feature of this experimental design is that customers can opt out of the program. As Figure 1 shows, evaluators should include opt-out subjects in the energy savings analysis to ensure unbiased savings estimates. Evaluators can then calculate savings as the difference in average energy use between treatment group customers, including opt-out subjects and control group customers. Removing opt-out subjects from the analysis would bias the savings estimate because identifying subjects in the control group who would have also opted out had they received the treatment is impossible. The resulting savings estimate is therefore an average of the savings of treated customers who remain in the program and of customers who opted out.

Depending on the type of BB program, the percentage of customers who opt out may be small, and may not affect the savings estimates significantly (e.g., few customers generally opt out of energy reports programs).

3.4.2 Randomized Control Trial With Opt-In Program Design

RCT with opt-out subjects assumes that the BB program treatments can be administered to subjects without their agreement. This is the case for programs in which, for example, a utility mails energy reports to customer homes or leaves door hangers with energy savings tips on customer homes. However, the utilities must have consent to administer some interventions. Examples include offering Web-based home audit or energy consumption tools; programmable, communicating thermostats with wireless capability; an online class about energy rates and efficiency; or in-home displays. All these interventions require that customers opt in to the program.

An opt-in RCT, (Figure 2) can accommodate the necessity for customers to opt in to some BB programs. This design results in an unbiased estimate of the ITT effect for customers who opt in to the program. The estimate of savings will have internal validity; however, it will not have external validity, because it will not apply to subjects who do not opt in.

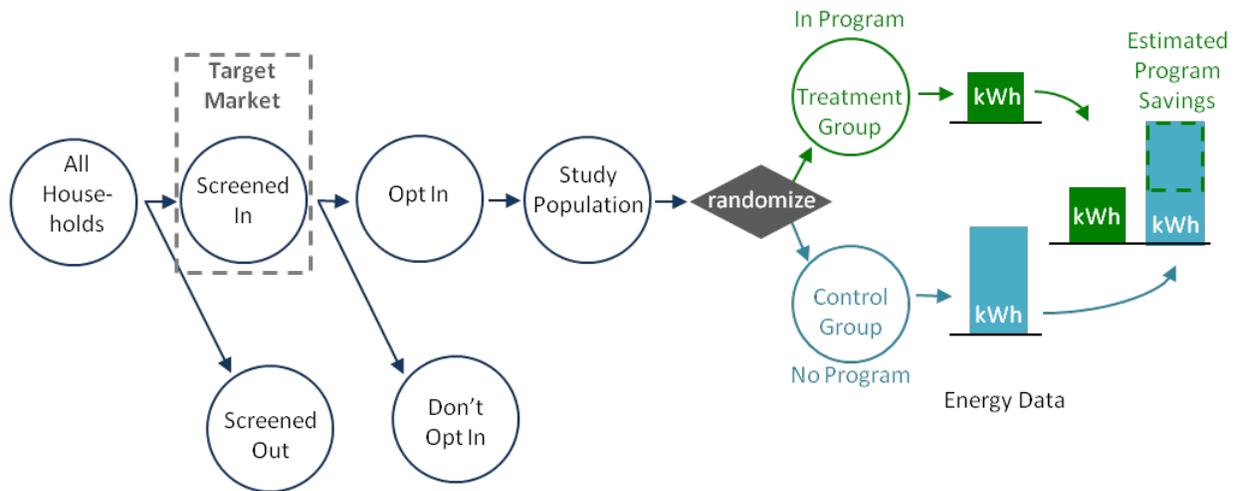


Figure 2. Illustration of RCT with opt-in program design

Implementing opt-in RCTs is very similar to implementing opt-out RCTs. The first step, screening utility customers for eligibility to determine the study population, is the same. The next step is to market the program to eligible customers. Some eligible customers may then agree to participate. Then, an independent third party randomly assigns these customers to either a treatment group that receives the intervention or a control group that does not. The utility delays or denies participation in the program to customers assigned to the control group. Thus, only customers who opted in and were assigned to the treatment group will receive the treatment.

Randomizing only opt-in customers ensures that the treatment and control groups are equivalent in their energy use characteristics. In contrast, other quasi-experimental approaches, such as matching participants to nonparticipants, cannot guarantee either this equivalence or the internal validity of the savings estimates.

After the random assignment, the opt-in RCT proceeds the same as an RCT with opt-out subjects: the utility administers the intervention to the treatment group. The evaluator collects energy use data from the treatment and control groups, then estimates energy savings as the difference in energy use between the groups. The evaluator does not collect energy use data for customers who do not opt in to the program.

An important difference between the opt-in RCTs and RCTs with opt-out subjects is how to interpret the savings estimates. In the RCT with opt-out subjects, the evaluator bases the savings estimate on a comparison of the energy use between treatment and control groups, which pertains to the entire study population. In contrast, in the opt-in RCT, the savings estimate pertains to the subset of customers who opted into the program, and the difference in energy use represents the treatment effect on customers who opted in to the program. Opt-in RCT savings

estimates have internal validity; however, they do not apply to customers who did not opt in to the program.

3.4.3 Randomized Encouragement Design

For some opt-in BB programs, delaying or denying participation to some customers may be undesirable. In this case, neither the opt-out nor the opt-in RCT design would be appropriate, and this protocol recommends an RED. Instead of randomly assigning subjects to receive or not receive an intervention, a third party randomly assigns them to a treatment group that is *encouraged* to accept the intervention (i.e., to participate in a program or adopt a measure), or to a control group that does not receive encouragement. Customers who receive the encouragement can refuse the intervention, and, depending on the program design, control group customers who learn about the intervention may be able to participate.

The RED yields an unbiased estimate of the effect of encouragement on energy use and, depending on the program design, can also provide an unbiased estimate of either the effect of the intervention on customers who accept it because of the encouragement or the effect of the intervention on all customer who accept it.

Figure 3 illustrates the process flow for a program using an RED. As with the RCT with opt-out and opt-in RCT, the first two steps are to identify the sample frame and select a study population. Next, like the RCT with opt out, a third party randomly assigns subjects to a treatment group, which receives encouragement, or to a control group, which does not. For example, a utility might employ a direct mail campaign that encourages treatment group customers to use an online audit tool. The utility would administer the intervention to treatment group customers who opt-in. Although customers in the control group did not receive encouragement, some may learn about the program and decide to sign up. The program design shown in Figure 3 allows for control group customers to receive the behavioral intervention.

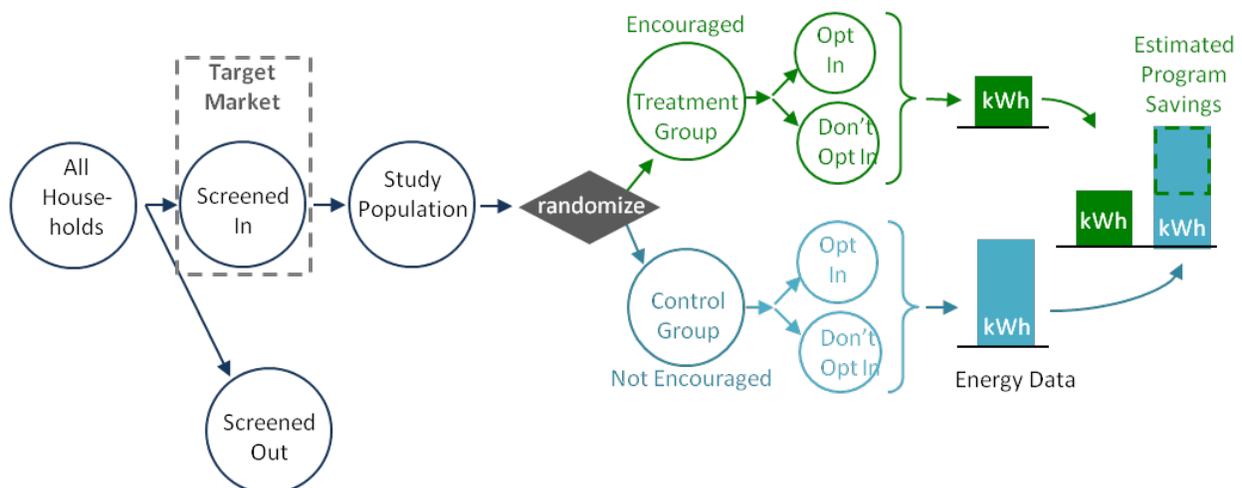


Figure 3. Illustration of RED program design

In Figure 3, the difference in energy use between homes in the treatment and control groups is an estimate of savings from the encouragement, not from the intervention. However, evaluators can

also use the difference in energy use to estimate savings for customers who accept the intervention because of the encouragement. To see this, consider that the study population comprises three types of subjects: (1) always takers, or those who would accept the intervention whether encouraged or not; (2) never takers, or those who would never accept the intervention even if encouraged; and (3) compliers, or those who would accept the intervention only if encouraged. Compliers participate only after receiving the encouragement.

Because eligible subjects are randomly assigned to groups depending on whether they receive encouragement, the treatment and control groups should have equal frequencies of always takers, never takers, and compliers in expectation. After treatment, the only difference between the treatment and control groups is that compliers in the treatment group accept the treatment and compliers in the control group do not. In both groups, always takers accept the treatment and never takers always refuse the treatment. Therefore, the difference in energy use between the groups reflects the treatment effect of encouragement on compliers.

To estimate the effect of the intervention on compliers (known as the local average treatment effect [LATE]), it is necessary to scale the treatment effect of the encouragement by the difference between treatment and control groups in the percentage of customers who receive the intervention:¹⁹

$$\frac{1}{(\% \text{ of encouraged customers who accepted} - \% \text{ of not encouraged customers who did not accept})}$$

The LATE does not capture the program effect on always takers because always takers in the control group are permitted to participate.

For BB programs with REDs that do not permit control group customers to participate, evaluators can estimate the treatment effect on the treated (TOT). The TOT is the effect of the program intervention on all customers who accept the intervention. In this case, the difference in energy use between the treatment and control groups reflects the impact of the encouragement on the always takers and compliers in the treatment group. Scaling the difference by the inverse of the percentage of customers who accepted the intervention yields an estimate of the TOT impact.²⁰

Successful application of an RED requires that compliers constitute a sufficiently large percentage of the encouraged population.²¹ If the RED generates too few compliers, the effects of the encouragement and receiving the intervention cannot be precisely estimated. Therefore, before employing an RED, evaluators should ensure that the sample size is sufficiently large and

¹⁹ This approach of estimating savings from the intervention because of encouragement assumes zero savings for customers who received encouragement but did not accept the intervention. If encouraged customers who did not accept the intervention reduced their energy use in response to the encouragement, the savings estimate for compliers will be biased upward.

²⁰ If the effect of program participation is the same for compliers as for others, those who would have participated without encouragement (always-takers) and those who do not participate (never takers), the RED will yield an unbiased estimate of the population average treatment effect.

²¹ For an example of the successful application of an RED, see SMUD (2013).

that the encouragement will result in the required number of compliers. If the risk of an RED generating too few compliers is significant, evaluators may want to consider alternative approaches, including quasi-experimental methods.

3.4.4 Persistence Design

Studies of home energy reports programs show that program savings persist while homes continue to receive reports.²² However, utilities and regulators may want to know what happens to BB program savings after the behavioral intervention ends. They may wish to measure whether their savings persist after the utility stops sending reports and for how long, as well as the rate of the savings “decay.” As Allcott and Rodgers (forthcoming) demonstrate, the rate of savings decay after treatment ends has significant implications for the performance of efficiency program portfolios and measuring cost effectiveness of BB programs. Initial studies of home energy reports programs indicate that some portion of savings may persist after the treatment stops, although further research is needed.²³

This protocol recommends that evaluators employ RCTs to estimate the persistence of BB program savings after participants stop receiving the intervention. The application of an RCT to a savings persistence study proceeds similarly to the application of RCTs previously discussed.

The utility is assumed to implement the BB program as an RCT with opt-out design; that is, customers from the study population were randomly assigned to a treatment group that received an intervention or to a control group that did not. Customers are able to opt out of the program (see Figure 1).

The persistence study starts with identifying the study population, in this case, the population of treated customers who received the intervention. The utility may choose to screen this population and study persistence by energy use or by socio-demographic characteristics. The persistence study population must include customers who opted out, because evaluators will need to make energy use comparisons between the persistence study population and the original control group, which includes customers who would have opted out.

The next step is to randomly assign customers in the persistence study population to one of two groups. Customers in the “discontinued treatment” group will stop receiving the intervention; customers in the “continued treatment” group will continue receiving intervention. The utility then administers the study and collects energy use data after sufficient time has passed to observe the persistence effects.

To estimate savings persistence after the end of treatment, the evaluator compares the energy use of customers in the discontinued treatment group with the energy use of customers in the original control group. This represents the post-treatment savings for customers who no longer received the intervention.

²³ Allcott and Rodgers (forthcoming), Brattle (2012), SMUD (2012), PSE (2012) estimate BB program savings after the treatment is discontinued. Allcott and Rodgers estimate a savings decay rate of about 19%/year.

To estimate the decay of savings after the end of treatment, the evaluator compares the energy use of the discontinued treatment group before and after treatment was discontinued. This represents the change in savings after the end of treatment for customers in the discontinued treatment group.

To estimate savings that were foregone because treatment was discontinued, the evaluator compares the savings of the continued and discontinued treatment groups after the end of treatment.

3.5 Evaluation Benefits and Implementation Requirements of Randomized Experiments

This protocol strongly recommends the use of randomized field experiments (RCTs or REDs) for evaluating residential BB programs. Sections 3.1–3.4 described the benefits—and some requirements—of evaluating BB programs using randomized experiments; this section summarizes these benefits and requirements.

Table 1 summarizes the benefits and implementation requirements.

Table 1. Benefits and Implementation Requirements of Randomized Experiments

Evaluation Benefits	Implementation Requirements
<ul style="list-style-type: none"> • Yield unbiased, valid estimates of causal program impacts, resulting in a high degree of confidence in the savings • Yield savings estimates that are robust to changes in model specification • Are versatile, and can be applied to opt-out and opt-in BB programs • Are widely accepted as the “gold standard” of good program evaluations • Result in transparent and straightforward analysis and evaluation • Can be designed to test specific research questions such as persistence of savings after treatment ends 	<ul style="list-style-type: none"> • An appropriately sized analysis sample • Accurate energy use measurements for sampled units • Advance planning and early evaluator involvement in program design • Restricted participation or program marketing to randomly selected customers

The principal benefit of randomized experiments is that they yield unbiased and robust estimates of program savings. They are also versatile, widely accepted, and straightforward to analyze. The principal requirements for implementing randomized experiments include the availability of accurate energy use measurements and a sufficiently large analysis study population.²⁴

²⁴ A frequent objection to the use of randomized experiments is that some utility customers may not have the opportunity to participate in a program. However, programs are often limited to a certain subset of customers; for example, a program may start out as limited to customers in a certain county or other geographic location. REDs allow any customers who would like to participate the opportunity to do so, even if they are in the control group. In our view, limiting the availability of the program to certain customers in RCTs is done with the worthy objective of advancing the utility’s knowledge of program savings effects and making future allocation of scarce efficiency resources more optimal.

Also, this protocol specifically recommends REDs or RCTs for estimating BB program savings. Both designs yield unbiased savings estimates. The choice of RED or RCT will depend primarily on program design and implementation considerations, in particular, whether the program has an opt-in or opt-out design. RCTs work well with opt-out programs such as residential energy reports programs. Customers who do not want to receive reports can opt out at any time without adversely affecting the evaluation. RCTs also work well with opt-in programs for which customer participation can be delayed (e.g., customers are put on a “waiting list”) or denied. For situations in which delaying or denying a certain subset of customers is impossible, REDs may be more appropriate. REDs can accommodate all interested customers, but have the disadvantages of requiring larger analysis samples and requiring two analysis steps to yield a direct estimate of the behavioral intervention’s effect on energy use.

Table 2 lists some issues to consider in choosing an RCT or RED.

Table 2. Considerations in Selecting an Experimental Design

Experimental Design	Evaluation Benefits	Implementation and Evaluation Requirements
RCT	<ul style="list-style-type: none"> • Yields unbiased, robust, and valid estimates of causal program impacts, resulting in a high degree of confidence in the savings • Simple to understand • Works well with opt-out programs • Works well with opt-in programs if customers can be delayed or denied 	<ul style="list-style-type: none"> • May require delaying or denying participation of some customers if program requires customers to opt in
RED	<ul style="list-style-type: none"> • Yields unbiased, robust, and valid estimates of causal program impacts, resulting in a high degree of confidence in the savings • Can accommodate all customers interested in participating • Works well with opt-in and opt-out programs 	<ul style="list-style-type: none"> • More complex design and harder to understand • Requires a more complex analysis • Requires larger analysis sample

4 Savings Estimation

Evaluators should estimate BB program savings as the difference in energy use between treatment and control group subjects in the analysis sample. Energy savings for a household in the BB program is the difference between the energy the household used and the energy the household would have used if it had not participated. However, the energy use of a household cannot be observed under two different states. Instead, to estimate savings, evaluators should compare the energy use of households in the treatment group to that of a group of households that are statistically the same but did not receive the treatment (the homes randomly assigned to the control group). In a randomized experiment, assignment to the treatment is random; thus, evaluators can expect control group subjects to use the same amount of energy that the treatment group would have used without the treatment. The difference in their energy use will therefore be an unbiased estimate of energy savings.

Savings can be estimated using energy use data from the treatment period only or from before and during the treatment. If energy use data from only the treatment period are used, evaluators estimate the savings as a simple difference (D). If the analysis also controls for energy use before the treatment, evaluators will estimate the savings as a D-in-D. The availability of energy use data for the period before the treatment will determine the approach, but D-in-D estimation is strongly advised when pretreatment energy use data are available.

Both approaches result in unbiased estimates of savings (i.e., in expectation, the two methods are expected to yield an estimate equal to the true savings), but D-in-D estimation generally results in more precise savings estimates (i.e., the D-in-D estimate will have a smaller standard error) because it accounts for time-invariant energy use that contribute significantly to the variance of energy use between subjects.²⁵

When conducting D-in-D estimation, evaluators should collect at least 1 full year of historical energy use data (the 12 months immediately before the program start date) to ensure that baseline data fully reflect seasonal energy use effects.

How frequently should BB programs be evaluated? Regulators usually determine the frequency of program evaluation. Although requirements vary between jurisdictions, most BB programs are evaluated once per year. Annual evaluation seems appropriate for many BB programs such as home energy reports programs because savings tend to increase over time, or at least for the first several years of the program.

4.1 Sample Design

Utilities should integrate the design of the analysis sample with program planning, because numerous considerations, including the size of the analysis sample, the method of recruiting

²⁵ D-in-D estimation also accounts for differences in the mean energy use between treatment and control group subjects that are introduced when subjects are randomly assigned to the treatment or control group. Evaluators may not expect such differences with random assignment; however, these differences may nevertheless arise.

customers to the program, and the type of randomized experiment, must be addressed before the program begins.

4.1.1 Sample Size

The analysis sample should be large enough to detect the minimum hypothesized program effect with desired probability.²⁶ To determine the minimum number of subjects required, researchers should employ a statistical power analysis. The inputs for this calculation are:

- The hypothesized program effect
- The coefficient of variation of energy use
- The specific analysis approach to be used (e.g., simple differences of means or a repeated measure analysis), tolerances for Type I and Type II statistical errors (as discussed in Section 3.2)
- The correlation of a subject's energy use observations. Most statistical software, including SAS, STATA, and R, include packages for performing statistical power analyses.²⁷

If the BB program will operate for more than several months or allow subjects to opt out, program planners should account for attrition, the loss of some subjects from the analysis sample because of account closures.

4.1.2 Random Assignment to Treatment and Control Groups by Independent Third Party

After determining the appropriate sizes of the treatment and control group samples, an independent and experienced third-party evaluator should randomly assign subjects to the treatment and control groups. If there is a significant risk that the random assignment will result in unbalanced treatment and control groups, this protocol recommends that evaluators first stratify the study population by pretreatment energy use and then randomly assign subjects in each stratum to treatment and control groups. Stratifying the sample will increase the likelihood that treatment and control group subjects have similar pretreatment means and variances.²⁸

Although this protocol strongly recommends that independent and experienced third-party perform the random assignment, circumstances sometimes make this impossible. In such cases, a third-party evaluator should certify that the assignment of treatment and control group subjects was done correctly and did not introduce bias into the selection process.

²⁶ The utility may also base the number of subjects in the treatment group on the amount of savings it desires to achieve.

²⁷ If statistical software is not available and one wishes to manually calculate the sample sizes, Brattle (2011) provides formulas for calculating sample sizes using statistical power calculations. The formulas also appear in Hilbe (1993) and Seed (1997).

²⁸ Shadish et al. (2002) discuss the benefits of stratified random assignment.

4.1.3 Equivalency Check

The third party performing the random assignment must verify that the characteristics of subjects in the treatment group, including pretreatment energy use, are balanced with those in the control group. If subjects in the groups are not equivalent overall, the energy savings estimates may be biased.

To verify the equivalence of energy use, this protocol recommends that the third party test for differences between treatment and control group subjects in both the mean pretreatment period energy use and in the distribution of pretreatment energy use. Evaluators should also test for differences in other available covariates, such as home floor area and heating fuel type.

If significant differences are found, the third party should consider performing the random assignment again. Ideally, random assignment should not result in any differences; however, differences occasionally appear, and it is better to redo the random assignment than to proceed with unbalanced treatment and control groups, which may lead to biased savings estimates. As noted in Section 4.1.2, stratifying the study population by pretreatment energy use will increase the probability that the groups are balanced.

If the evaluator is not the third party who performed the random assignment, the evaluator should also perform an equivalency check. The evaluator may be able to use statistical methods to control for differences in pretreatment energy use that are found after the program is underway.²⁹

4.2 Data Requirements and Collection

4.2.1 Energy Use Data

Estimating BB program impacts using a field experiment requires collecting energy use data from subjects in the analysis sample. This protocol recommends that evaluators collect multiple energy use measurements for each sampled unit for the periods before and during the treatment.³⁰

These data are known as a panel. Panels can consist of multiple hourly, daily, or monthly energy use observations for each sampled unit. In this protocol, a panel refers to a dataset that includes energy measurements for each sampled unit either for the pretreatment and treatment periods or for the treatment period only. The time period for panel data collection will depend on the program timeline, the frequency of the energy use data, and the amount of data collected.

Panel data have several advantages for use in measuring BB program savings:

- **Relative ease of collection.** Collecting multiple energy use measurements for each sampled unit from utility billing systems is usually easy and inexpensive.

²⁹ If energy use data are available for the periods before and during the treatment, it is possible to control for time-invariant differences between sampled treatment and control group subjects using subject fixed effects.

³⁰ A single measurement of energy use for each sampled unit during the treatment period also results in an unbiased estimate of program savings. The statistical significance of the savings estimate depends on the variation of the true but unknown savings and the number of sampled units.

- **Can estimate savings during specific times.** If the panel collects enough energy use observations per sampled unit, estimating savings at specific times during the treatment period may be possible. For example, hourly energy use data may enable the estimation of precise savings during utility system peak hours. Monthly energy use data may enable the development of precise savings estimates for each month of the year.
- **Savings estimates are more precise.** Evaluators can more precisely estimate energy savings with a panel, because they may be able to control for the time-invariant differences in energy use between subjects that contribute to the variance of energy use.
- **Allows for smaller analysis samples.** All else being equal, fewer units are required to detect a minimum level of savings in a panel study than in a cross-section analysis. Thus, collecting panel data may enable studies with smaller analysis samples and data collection costs.

Using panel data has some disadvantages relative to a single measurement per household: (1) evaluators must correctly cluster the standard errors within each household or unit (as described in the following section); and (2) panel data require statistical software to analyze, whereas estimating savings using single measurements in a basic spreadsheet software program may be possible.

This protocol also recommends that evaluators collect energy use data for the duration of the treatment to ensure they can observe the treatment effect for the entire study period. Ideally, an energy efficiency BB program lasts for a year or more because the energy end uses affected by BB programs vary seasonally. For example, these programs may influence weather-sensitive energy uses, such as space heating or cooling, so collecting less than 1 year of data to reflect every season may yield incomplete results.

Collecting data for an entire year may be impossible because some BB programs do not last that long. For these programs, only an unbiased estimate of savings for the time period of analysis may be obtained. Evaluators should exercise caution in extrapolating those estimates to seasons or months outside the analysis period, especially if the BB program affected weather-sensitive or seasonally varying end uses of energy.

4.2.2 Makeup of Analysis Sample

Evaluators must collect energy use measurements for every household or unit that is initially assigned to a control or treatment group, whether or not the household or unit later opts out. Not collecting energy use data for households initially placed in a treatment group but that then opts out results in imbalanced treatment and control groups and a biased savings estimate.

4.2.3 Other Data Requirements

Program information about each program participant must also be collected. These data must include whether the subject is assigned to the treatment or control group, when the treatments are administered, and if and when the subject opt out.

Temperature and other weather data may also be useful but are not necessary. Weather data should be collected for each household from the nearest weather station.

4.2.4 Data Collection Method

Energy use measurements used in the savings estimation should be collected directly from the utility, not from the program implementer, at the end of the program evaluation period. Depending on the program type, utility billing system, and evaluation objectives, the data frequency can be at 15-minute, 1-hour, daily, or monthly intervals.

4.3 Analysis Methods

This protocol recommends using panel regression analysis to estimate savings from BB field experiments where subjects were randomly assigned to either treatment or control groups. Evaluators typically prefer regression analysis to simply calculating differences in unconditional mean energy use, because it generally results in more precise savings estimates. A significant benefit of randomized field experiments is that regression-based savings estimates are usually quite insensitive to the type of model specification.

Section 4.3.1 addresses issues in panel regression estimation of BB program savings, including model specification and estimation, standard errors estimation, robustness checks, and savings estimation. It illustrates some specifications as well as the application of energy-savings estimation.

4.3.1 Panel Regression Analysis

In panel regressions, the dependent variable is usually the energy use of a subject (a home, apartment, or dormitory) per unit of time such a month, day, or hour. The right side of the equation includes an independent variable to indicate whether the subject was assigned to the treatment or control group. This variable can enter the model singularly or be interacted with another independent variable, depending on the analysis goals and the availability of energy use data from before treatment. The coefficient on the term with the treatment indicator is the energy savings per subject per unit of time. D-in-D models of energy savings must also include an indicator for whether the period occurred before or during the treatment period.

Many panel regressions also include fixed effects. Subject fixed effects capture unobservable energy use specific to a subject that does not vary over time. For example, home fixed effects may capture variation in energy use that is due to differences such as home sizes or makeup of a home's appliance stock. Time-period fixed effects capture unobservable energy use specific to a time period that does not vary between subjects. Including time or subject fixed effects in a regression of energy use of subjects randomly assigned to the treatment or control group will increase the precision but not the unbiasedness of the savings estimates.

Fixed effects can be incorporated into panel regression in several ways.

- Include a separate dummy variable or intercept for each subject in the model. The estimated coefficient on a subject's dummy variable represents the subject's time-invariant energy use. This approach, known as Least Squares Dummy Variables, may, however, not be practical for evaluations with a large number of subjects, because the model requires thousands of dummy variables that may overwhelm available computing resources.

- Apply the fixed-effect estimator, which requires transforming the dependent variable and all the independent variables by subtracting subject-specific means and then running Ordinary Least Squares (OLS) on the transformed data.³¹ This approach is equivalent to Least Squares Dummy Variables.
- Estimate a first difference or annual difference of the model. Differencing removes the subject fixed effect and is equivalent to the dummy variable approach if the fixed-effects model is correctly specified.³²

4.3.2 Panel Regression Model Specifications

This section outlines common regression approaches for estimating treatment effects from residential BB programs. Unless otherwise stated, assume that the BB program was implemented as a field experiment with a randomized control trial or randomized encouragement design.

4.3.3 Simple Differences Regression Model of Energy Use

Consider a BB program in which the evaluator has energy use data for the treatment period only, and wishes to estimate the average energy savings per period from the treatment. Let $t = 1, 2, \dots, T$, where t denotes the time periods during the treatment for which data are available³³, and let $I = 1, 2, \dots, N$, where i denotes the treatment and control group subjects. For simplicity, assume that all treated subjects started the treatment at the same time.

A basic specification to estimate the average energy savings per period from the treatment is:

Equation 1

$$y_{it} = \beta_0 + \beta_1 * Tr_i + \epsilon_{it}$$

Where,

y_{it} = The metered energy use of subject i in period t .

³¹ Greene (2011) Chapter 11 provides more details.

³² Standard econometric formulations assume that fixed effects account for unobservable factors that are correlated with one or more independent variables in the model. This correlation assumption distinguishes fixed-effects panel model estimation from other types of panel models. Fixed effects eliminate bias that would result from omitting unobserved time-invariant characteristics from the model. In general, fixed effects must be included to avoid omitted variable bias. In an RCT, however, fixed effects are unnecessary to the claim that the estimate of the treatment effect is unbiased because fixed effects are uncorrelated with the treatment by design. Although fixed effects regression is unnecessary, it will increase precision by reducing model variance.

Some evaluators may be tempted to choose to use random-effects estimation, which assumes time- or subject-invariant factors are uncorrelated with other variables in the model. However, fixed-effects estimation has important advantages over random-effects estimation: (1) it is robust to the omission of any time-invariant regressors. If the evaluator has doubts about whether the assumptions of the random-effects model are satisfied, the fixed-effects estimator is better; and (2) it yields consistent savings estimates when the assumptions of the random-effects model holds. The converse is not true, making the fixed-effects approach more robust.

Because weaker assumptions are required for the fixed-effects model to yield unbiased estimates, this protocol generally recommends the fixed-effects estimation approach. The remainder of this protocol presents panel regression models that satisfy the fixed-effects assumptions.

³³ For a treatment that is continuous, an example might be $t = 1$ on the first day that the treatment starts, $t = 2$ on the second day, etc.; for a treatment that occurs during certain days only (e.g., a day when the utility's system peaks), an example might be $t = 1$ during the first critical event day, $t = 2$ during the second, etc.

- β_0 = The average energy use per unit of time for subjects in the control group.
- β_1 = The average treatment effect of the program. The energy savings per subject per period equals $-\beta_1$.
- Tr_i = An indicator for whether subject i received the treatment. The variable equals 1 for subjects in the treatment group and equals 0 for subjects in the control group.
- ε_{it} = The model error term, representing random influences on the energy use of customer i in period t .

In this simple model, the error term ε_{it} is uncorrelated with Tr_i because subjects were randomly assigned to the treatment or control group. OLS estimation of this model will result in an unbiased estimate of β_1 . The standard errors should be clustered on the subject.³⁴

This specification does not include subject fixed effects. Because the available energy use data apply to the treatment period only, the program treatment effect cannot be identified and subject fixed effects cannot be accounted for. However, as previously noted, because of the random assignment of subjects to the treatment group, any time-invariant characteristics affecting energy use will be uncorrelated with the treatment, so omitting that type of fixed effects will not bias the savings estimates.

Using Equation 1, however, more precise estimates of savings could be obtained by replacing the coefficient β_0 with time-period fixed effects. The model thus captures more of the variation in energy use over time, resulting in greater precision in the estimate of savings. The interpretation of β_1 , the average treatment effect per home per time period, is unchanged.

4.3.4 Simple Differences Regression Estimate of Heterogeneous Savings Impacts

Suppose that the evaluator still has energy use data that apply to the treatment period only, but wishes to obtain an estimate of savings from the treatment as a function of some exogenous variable such as preprogram energy use, temperature, home floor space, or pretreatment efficiency program participation (to determine, for example, whether high energy users save more or less energy than low energy users). If data for treatment and control group subjects on the exogenous variable of interest are available, the evaluator may be able to estimate the treatment effect as a function of this variable.

Let m_{ij} be an indicator that subject i belongs to a group j , $j = 1, 2, \dots, J$, where membership in group j is exogenous to receiving the treatment. Then the average treatment effect for subjects in group j can be estimated using the following regression equation:

³⁴ Although the methods recommended in this protocol minimize the potential for violations of the assumptions of the classical linear regression model, evaluators should be aware of—and take steps to minimize—potential violations.

Equation 2

$$y_{it} = \beta_0 + \sum_{j=1}^J \beta_{1j} * Tr_i * m_{ij} + \sum_{j=1}^{J-1} \gamma_j m_{ij} + \varepsilon_{it}$$

Where,

- m_{ij} = An indicator for membership of subject i in group j . It equals 1 if customer i belongs to group j and equals 0, otherwise.
- β_{1j} = The average treatment effect for subjects in group j . Energy savings per subject per period j equals $-\beta_{1j}$.
- γ_j = The average energy use per period for subjects in group j , $j = 1, 2, \dots, J-1$.

All of the other variables are defined as in Equation 1.

This specification includes a separate intercept for each group indicated by γ_j and the treatment indicator Tr_i interacted with each of the m_{ij} indicators. The coefficients on the interaction variables β_{1j} show average savings for group j relative to baseline average energy use for group j .

4.3.5 Simple Differences Regression Estimate of Savings During Each Time Period

To estimate the average energy savings from the treatment during each period, the evaluator can interact the treatment indicator with indicator variables for the time periods as in the following equation³⁵:

Equation 3

$$y_{it} = \sum_{j=1}^T \beta_t Tr_i * d_{jt} + \sum_{j=1}^T \theta_t d_{jt} + \varepsilon_{it}$$

Where,

- β_t = The average savings per subject specific to period t (e.g., the average savings per subject during month 4 or during hour 6).
- d_{jt} = An indicator variable for period j , $j = 1, 2, \dots, T$. d_{jt} equals 1 if $j = t$ (i.e., the period is the t^{th}) and equals 0 if $j \neq t$ (i.e., the period is not the t^{th}).
- θ_t = The average effect on consumption per subject specific to period t .

Equation 3 can be estimated by including a separate dummy variable and an interaction between that dummy variable and Tr_i for each time period t , where $t = 1, 2, \dots, T$. When the time period is in months, the time-period variables are referred to as month-by-year fixed effects. The coefficient on the interaction variable for period t , β_t , is the average savings per subject for period t . Again, because ε_{it} is uncorrelated with the treatment after accounting for the average

³⁵ If the number of time periods is very large, the number of time period indicator variables in the regression may overwhelm the capabilities of the available statistical software. Another option for estimation is to transform the dependent variable and all of the independent variables by subtracting time period-specific means and then running OLS on the transformed data.

energy use in period t , OLS estimation of Equation 3 (with standard errors clustered at the subject level) results in an unbiased estimate of the average treatment effect for each period.

Evaluators with smart meter data can use this specification to estimate BB program demand savings during specific hours of the analysis period. The coefficient β_t would indicate the demand savings from the treatment during hour t . Examples of research that estimates savings during hours of peak usage include Stewart (2013) and Todd (2014).

4.3.6 Difference-in-Differences Regression Model of Energy Use

This section outlines a D-in-D approach to estimating savings from BB field experiments. This protocol recommends D-in-D estimation to the simple differences approach, but it requires information about the energy use of treatment and control group subjects during the pretreatment and treatment periods. These energy use data enable the evaluator to:

- Include subject fixed effects to account for differences between subjects in time-invariant energy use.
- Obtain more precise savings estimates.
- Test identifying assumptions of the model.

Assume there are N subjects and $T + 1$ periods, $T > 0$, in the pretreatment period denoted by $t = -T, -T+1, \dots, -1, 0$, and T periods in the treatment period, denoted by $t = 1, 2, \dots, T$. A basic D-in-D panel regression with subject fixed effects could be specified as:

Equation 4

$$y_{it} = \alpha_i + \beta_1 P_t + \beta_2 P_t * Tr_i + \varepsilon_{it}$$

Where,

α_i = Unobservable, time-invariant energy use for subject i . These effects are controlled for with subject fixed effects.

β_1 = The average energy savings per subject during the treatment period that was not caused by the treatment.

P_t = An indicator variable for whether time period t occurs during the treatment. It equals 1 if treatment group subjects received the treatment during period t , and equals 0 otherwise.

β_2 = The average energy savings due to the treatment per subject per unit of time.

The model includes fixed effects to account for differences in average energy use between subjects. Including subject fixed effects would likely explain a significant amount of the variation in energy use between subjects and result in more precise savings estimates. The interaction of P_t and Tr_i equals one for subjects in the treatment group during periods when the treatment is in effect, and 0 for other periods and all control subjects.

Equation 4 is a D-in-D specification. For control group subject i , the expected energy use is α_i during the pretreatment period and $\alpha_i + \beta_1$ during the treatment period. The difference in expected energy use between pretreatment and treatment periods, also known as *naturally occurring savings*, is β_1 . If that same subject i had been in the treatment group, the expected energy use would have been α_i during the pretreatment period and $\alpha_i + \beta_1 + \beta_2$ during the treatment period. The expected savings would have been $\beta_1 + \beta_2$, which is the sum of naturally occurring savings and savings from the BB program. Taking the difference yields β_2 , a D-in-D estimate of program savings. OLS estimation results in an unbiased estimate of β_2 .

4.3.7 D-in-D Estimate of Savings for Each Time Period

By respecifying Equation 4 with time-period fixed effects, savings can be estimated during each period and the identifying assumption tested to determine that assignment to the treatment was random. Consider the following D-in-D regression specification:

Equation 5

$$y_{it} = \alpha_i + \sum_{j=-T}^T \theta_j d_{jt} + \sum_{j=-T}^{-1} \beta_j Tr_i * d_{jt} + \sum_{j=1}^T \beta_j Tr_i * d_{jt} + \varepsilon_{it}$$

Savings in each period are estimated by including a separate dummy variable and an interaction between the dummy variable and Tr_i for each time period t , where $t = -T, -T+1, \dots, -1, 0, 1, 2, \dots, T$. The coefficient on the interaction variable for period t , β_t^T , is the D-in-D savings for period t .

Unlike the simple differences regression model, this model yields an estimate of BB program savings during all periods except one, i.e., $t = 0$, for a total of $2T-1$ period savings estimates. Figure 4 shows an example of savings estimates obtained from such a model. The dotted lines show the 95% confidence interval for the savings estimates.

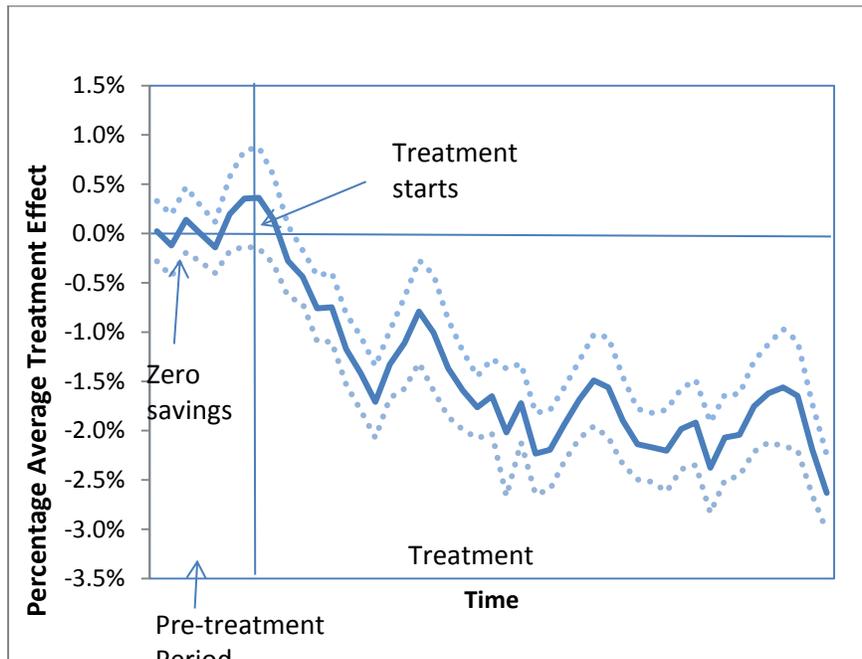


Figure 4. Example of D-in-D regression savings estimates

Estimates of pretreatment savings can be used to test the assumption of random assignment to the treatment. Before utilities administer the treatment, statistically significant differences in energy use between treatment and control group subjects should not be evident. BB program pretreatment saving estimates that were statistically different from zero would suggest a flaw in the experiment design. For example, an error in the randomization process may result in assignments of subjects to the treatment and control groups that were correlated with their energy use.

As with equation 3, this specification can be used to estimate demand savings during specific hours. Energy use data for hours before the treatment are required, however.

4.3.8 Randomized Encouragement Design

Some field experiments involve an RED in which subjects are only encouraged to accept a BB measure, in contrast to RCTs in which a program administers a BB intervention. This section outlines the types of regression models that are appropriate for REDs, how to interpret the coefficients, and how to estimate savings from RED programs.

Evaluators can apply the model specifications previously described for RCTs to REDs. The model coefficients and savings are interpreted differently, however, and an additional step is required to estimate average savings for subjects who accept the behavioral intervention. Treatment in an RED is defined as receiving encouragement to adopt the BB intervention, rather than actually receiving the intervention as with RCTs.

Consider a field experiment with an RED that has energy use data for treatment and control group subjects available for the pretreatment and treatment periods. Equations 1 through 4 can be used to estimate the treatment effect, or the average energy use effect on those receiving encouragement. The estimate captures savings from compliers only, because never takers never accept the intervention, and always takers would accept the intervention with or without encouragement.

To recover the LATE, the savings from subjects who accept the treatment because of the encouragement, scale the estimate of β_2 must by the inverse of the difference between the percentage of subjects in the treatment group who accept the intervention and the percentage of subjects in the control group who accept the intervention (which is zero if control group subjects are prohibited from accepting the intervention). Estimate this as:

Equation 6

$$\beta_2 / (\pi_T - \pi_C)$$

Where,

π_T = The percentage of treatment group subjects who accept the intervention.

π_C = The percentage of control group subjects who accept the intervention.

4.3.9 Models for Estimating Savings Persistence

A utility offering a residential BB program may want to know what happens to savings during the second or third year of the program or after the treatment stops. There are two kinds of savings effects to measure: (1) the effect of continuing the intervention on consumption called *savings during treatment*; and (2) the effect on consumption after discontinuing the intervention called *post-treatment savings*.

Suppose a utility implemented a BB program as an RCT and wants to measure the persistence of savings after the BB intervention stops. The utility started the treatment in period $t = 1$ and administered it for t^* periods. Beginning in period $t = t^* + 1$, the utility stopped administering the intervention for a random sample of treated subjects. Evaluators can estimate the average savings c for subjects who continue to receive the treatment (continuing treatment group) and for those who stopped receiving the treatment after period t^* (discontinued treatment group).

Assuming that pretreatment energy use data are available, the following regression equation can be used to estimate *savings during treatment* and *post-treatment savings*:

Equation 7

$$kWh_{it} = \alpha_i + \tau_t + \beta_1 P_{1,t} * Tc_i + \beta_2 P_{1,t} * Td_i + \beta_3 P_{2,t} * Tc_i + \beta_4 P_{2,t} * Td_i + \varepsilon_{it}$$

Where,

τ_t = The time-period fixed effect (an unobservable that affects the consumption of all subjects during time period t). The time period effect can be estimated by including a separate dummy variable for each time period t , where $t = -T, -T+1, \dots, -1, 0, 1, 2, \dots, T$.

β_1 = The average energy savings per continuing subject caused by the treatment during periods $t = 1$ to $t = t^*$.

$P_{1,t}$ = An indicator variable for whether subjects in the continued and discontinued treatment groups received the treatment during period t . It equals 1 if period t occurs between periods $t = 1$ and $t = t^*$ and equals 0 otherwise.

Tc_i = An indicator for whether subject i is in the continuing treatment group. The variable equals 1 for subjects in the continuing treatment group and equals 0 for subjects not in the continuing treatment group.

β_2 = The average energy savings per discontinuing subject caused by the treatment during periods $t = 1$ to $t = t^*$.

Td_i = An indicator for whether subject i is in the discontinuing treatment group. The variable equals 1 for subjects in the discontinuing treatment group and equals 0 for subjects not in the discontinuing treatment group.

- β_3 = The average energy savings from the treatment for subjects in the continuing treatment group when $t > t^*$.
- $P_{2,t}$ = An indicator variable for whether continuing treatment group subjects received the treatment and discontinued treatment group subjects did not receive the treatment during period t . It equals 1 if period t occurs after $t = t^*$ and equals 0 otherwise.
- β_4 = The average energy savings for subjects in the discontinued treatment group when $t > t^*$.

OLS estimation of equation 7 yields unbiased estimates of *savings during treatment* (β_3) and *post-treatment savings* (β_4) because original treatment group subjects were assigned randomly to the continuing and discontinued treatment groups. Evaluators can expect that $\beta_4 \geq \beta_3$, that is, the average savings of the continuing treatment group will be greater than that of the discontinued treatment group. To estimate savings decay after treatment stops, evaluators can take the difference between savings during treatment (β_2) and post-treatment savings (β_4) for subjects in the discontinued treatment group.

Evaluators can test the identifying assumption of random assignment to the discontinued treatment group by comparing the savings of continuing and discontinuing treatment group subject between period $t = 1$ and t^* . If assignment was random, their savings during this period are expected to be equal.

4.3.10 Standard Errors

Panel data have multiple energy use observations for each subject; thus, the energy use data are very likely to exhibit within-subject correlations. Many factors affecting energy use persist over time, and the strength of within-subject correlations usually increases with the frequency of the data. When standard errors for panel regression model coefficients are calculated, these correlations must be accounted for. Failing to do so will lead to savings estimates with standard errors that are biased downward.

This protocol strongly recommends that evaluators estimate robust standard errors clustered on subjects (the randomized unit in field trials) to account for within-subject correlation. Most statistical software programs, including STATA, SAS, and R, have regression packages that output regression-clustered standard errors.

Clustered standard errors account for having less information about energy use in a panel with N subjects and T observations per subject than in a dataset with $N \cdot T$ independent observations. Because clustered standard errors account for these within-subject energy-use correlations, they are typically larger than OLS standard errors. When there is within-subject correlation, OLS

standard errors are biased downward and overstate the statistical significance of the estimated regression coefficients.³⁶

4.3.11 Opt-Out Subjects and Account Closures

Many BB programs allow subjects to opt out and stop receiving the treatment. This section addresses how evaluators should treat opt-out subjects in the analysis, as well as subjects whose billing accounts close during the analysis period.

As a general rule, evaluators should include all subjects initially assigned to the treatment and control groups in the savings analysis.³⁷ For example, evaluators should keep opt-out subjects in the analysis sample. Opt-out subjects may have different energy use characteristics than subjects who remain in the program, and dropping them from the analysis would result in nonequivalent treatment and control groups. To ensure the internal validity of the savings, opt-out subjects should be kept in the analysis sample.

Sometimes treatment or control group subjects close their billing accounts after the program starts. Account closures are usually unrelated to the BB program or savings; most are due to households changing residences. Subjects in the treatment group should experience account closures for the same reasons and at the same rates as subjects in the control group; evaluators can thus safely drop treatment and control group subjects who close their accounts from the analysis sample.

However, if savings are correlated with the probability of an account closure, it may be best to keep subjects with account closures in the analysis sample. For example, if young households, which are the most mobile and likely to close their accounts, are also most responsive to BB programs, dropping these households from the analysis would bias the savings estimates downward,³⁸ and evaluators should keep these households in the analysis.

If evaluators drop customers who close their accounts during the treatment from the regression estimation, they should still count the savings from these subjects for periods during the treatment before customers closed their accounts. To illustrate, when estimating savings for a 1-year BB program, evaluators can estimate the savings from subjects who closed their accounts and from those who did not as the weighted sum of the conditional average program treatment effects in each month:

Equation 8

$$\text{Savings} = \sum_{m=1}^{12} -\beta_m * \text{Days}_m * N_m$$

³⁶ Bertrand et al. (2004) show when D-in-D studies ignore serially correlated errors, the probability of finding significant effects when there are none (Type I error) increases significantly.

³⁷ This protocol urges evaluators not to arbitrarily drop outlier energy use observations from the analysis unless energy use was measured incorrectly. If an outlier is dropped from the analysis, the reasons for dropping the outlier and the effects of dropping it from the analysis on the savings estimates should be clearly documented. Evaluators should test the sensitivity of the results to dropping observations.

³⁸ See State and Local Efficiency Action Network (2012), p. 30.

Where,

m	=	Indexes the months of the year
$-\beta_m$	=	The conditional average daily savings in month m (obtained from a regression equation that estimates the program treatment effect on energy use in each month)
$Days_m$	=	The number of days in month m
N_m	=	The number of subjects with active accounts receiving the treatment in month m or in a previous month

This approach assumes that savings in a given month for subjects who close their accounts are equal to savings of subjects whose accounts remain open.

4.4 Energy Efficiency Program Uplift and Double Counting of Savings

BB programs may increase participation in other utility energy efficiency programs; this additional participation is known as *efficiency program uplift*. For example, many energy reports programs encourage report recipients to adopt efficiency measures, such as furnaces, air conditioners, wall insulation, windows, and compact fluorescent lamps, in exchange for cash rebates. A utility may want to quantify savings from efficiency program uplift. Also, when a household participates in an efficiency program because of this encouragement, the utility might count their savings twice: once in the regression-based estimate of BB program savings and again in the estimate of savings for the rebate program. To avoid double counting savings, evaluators must estimate savings from program uplift and subtract them from the efficiency program portfolio savings.³⁹

Estimating the savings from BB program uplift with the experimental research designs recommended in this protocol is conceptually straightforward. To illustrate, suppose that a utility markets energy efficiency Measure A to treatment and control group subjects identically through a separate rebate program. Subjects in the treatment group also receive behavioral messaging encouraging them to adopt efficiency measures, including Measure A. Because customers were randomly assigned to the treatment and control groups, and the groups are equivalent except for whether they received the behavior treatment, evaluators can attribute any differences between the groups to the uptake of Measure A because of the BB program.

Figure 5 illustrates this logic for calculating behavior program savings from efficiency program uplift. Behavior program savings from adoption of Measure A is the difference between the treatment group and the control group in savings from Measure A. Savings can be estimated as the difference in rate of adoption of measure A between treatment and control group subjects

³⁹ BB program savings from efficiency program uplift were caused by the BB program: the savings would not have occurred in the program's absence. The level of participation in other utility efficiency programs caused by the BB program will depend on the efficiency program incentive amount, however. Although the BB program is necessary to cause the uplift, it may or may not be sufficient on its own. Because the incentive amount is typically not randomized, it is unclear whether the incentive program is necessary to cause the uplift; however, it alone is certainly not sufficient. Program uplift may be greater with larger rebates.

multiplied by the number of treatment group subjects multiplied by Measure A's per-unit savings.

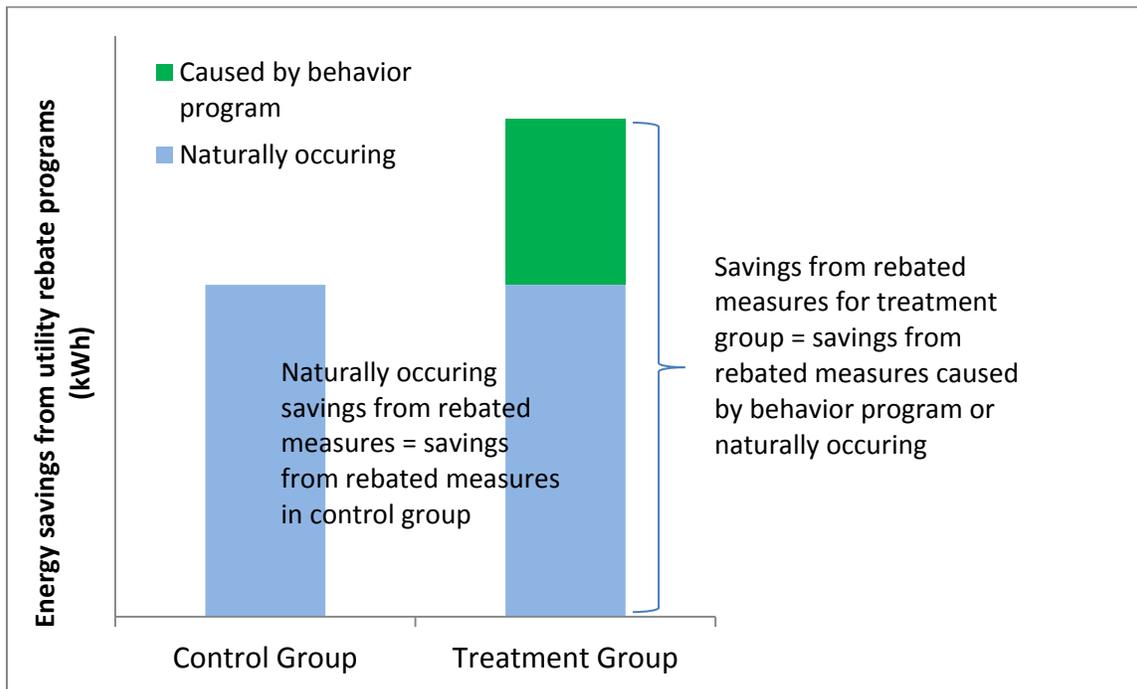


Figure 5. Calculation of double-counted savings

Evaluators can estimate BB program savings from efficiency program uplift for efficiency measures that the utility tracks at the customer level. Most measures that utilities rebate—such as high-efficiency furnaces, windows, insulation, and air conditioners—fit this description.

To estimate BB program savings from efficiency program uplift, evaluators should:

1. Match the BB program treatment and control group subjects to the utility energy efficiency program tracking data.
2. Calculate the savings per treatment group subject from efficiency uplift as the difference between treatment and control groups in average efficiency program savings per subject, where the savings are obtained from the utility tracking database. (The average should be calculated over all treatment group subjects, not just those that participated in efficiency programs.)
3. Multiply that difference by the number of subjects who are in the treatment group to see the savings from efficiency program uplift.

Evaluators should be mindful of specific reporting conventions for efficiency program measures in utility tracking databases. For example, many jurisdictions require utilities to report weather-normalized and annualized measure savings, which ignore both when measures were installed during the year and the actual weather conditions that affect savings. In contrast, the regression-based estimate of energy savings will reflect installation dates of measures and actual weather. Also, many utility tracking databases report gross savings instead of net savings.

To achieve accurate estimates of program uplift, evaluators may therefore need to adjust the measure savings in the database before taking differences between treatment and control groups. Otherwise, the measure savings in the tracking database and the regression-based estimate of savings will be inconsistently measured and the estimate of savings from program uplift may be biased.

Estimating savings from program uplift for measures that the utility does not track at the customer level is more difficult. The most important such measures are high-efficiency lights such as compact fluorescent lamps and light-emitting diodes that are rebated through utility upstream programs. Most utilities provide incentives directly to retailers for purchasing these measures, and the retailers then pass on these price savings to utility customers in the form of retail discounts. Data on the purchases of rebated measures by treatment and control group subjects must be collected to estimate BB savings in upstream efficiency programs. Evaluators can use household surveys for this purpose.⁴⁰ However, because the individual difference in the number of upstream measure purchases between treatment and control group subjects may be small, a large number of subjects must be surveyed to detect the BB program effect.

⁴⁰ For an example of the approach required to estimate BB program savings from adoption of compact fluorescent lamps, see PG&E (2013).

5 Reporting

BB program evaluators should carefully document the research design, data collection and processing steps, analysis methods, and plan for calculating savings estimates. Specifically, evaluators should describe:

- The program implementation and the hypothesized effects of the behavioral intervention
- The experimental design, including the procedures for randomly assigning subjects to the treatment or control group
- The sample design and sampling process
- Processes for data collection and preparation for analysis, including all data cleaning steps
- Analysis methods, including the application of statistical or econometric models and key assumptions used to identify savings, including tests of those key identification assumptions
- Results of savings estimate, including point estimates of savings and standard errors and full results of regressions used to estimate savings.

A good rule-of-thumb is that evaluators should report enough detail such that a different evaluator could replicate the results with the study data. Every detail does not have to be provided in the body of the report; many of the data collection and savings estimation details can be provided in a technical appendix.

References

- Allcott, H. (2011). Social Norms and Energy Conservation. *Journal of Public Economics*, 95(2), 1082-1095.
- Allcott, H.; Rodgers, T. (forthcoming). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*.
- Bertrand, M.; Duflo, E.; Mullainathan, S. (2004). “How Much Should We Trust Difference-in-Differences Estimates?” *Quarterly Journal of Economics* 119 (1), 249–275.
- The Brattle Group (2011). Measurement and Verification Principles for Behavior-Based Efficiency Programs. Prepared by Sanem Sergici and Ahmad Faruqui for Opower.
- The Brattle Group (2012). Persistence of Impacts from Opower’s Home Energy Report Programs: A Case Study of Two Utilities. Prepared by Ahmad Faruqui and Sanem Sergici for Opower.
- Cappers, P.; Todd, A.; Perry, M.; Neenan, B.; Boisvert, R. (2013). Quantifying the Impacts of Time-based Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines.
- Costa, D.L.; Kahn, M.E. (2010). Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. NBER Working Paper 15939. <http://www.nber.org/papers/w15939>.
- Consortium for Energy Efficiency Database (2013). <http://library.cee1.org/content/2013-behavior-program-summary-public-version>
- Davis, M. (2011). Behavior and Energy Savings: Evidence from a Series of Experimental Interventions. Environmental Defense Fund Report.
- EPRI (2010). Guidelines for Designing Effective Energy Information Feedback Pilots: Research Protocols. Electric Power Research Institute: Palo Alto, CA. 1020855.
- Greene, W. (2011). *Econometric Analysis*. New Jersey: Prentice Hall.
- Harding, M.; Hsiaw, A. (2012). Goal Setting and Energy Efficiency. Stanford University working paper.
- Hilbe, J.M. (1993). “Sample Size Determination for Means and Proportions.” *Stata Technical Bulletin* 11, 17–20.
- Ignelzi, P.; Peters, J.; Dethman, L.; Randazzo, K.; Lutzenhiser, L. (2013). Paving the Way for a Richer Mix of Residential Behavior Programs. Prepared for Enernoc Utility Solutions. CALMAC Study SCE0334.01.
- List, J.A. (2011). “Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off.” *Journal of Economic Perspectives* (25), 3–16.

List, J.A., Sadoff, S.; Wagner, M. (2010). So You Want to Run an Experiment, now What? Some Simple Rules of Thumb for Optimal Experimental Design. National Bureau of Economic Research Paper 15701.

Mazur-Strommen, S.; Farley, K. (2013). ACEEE Field Guide to Utility-Run Behavior Programs. American Council for an Energy Efficient Economy, Report Number B132.

PG&E (2013). *2012 Load Impact Evaluation for Pacific Gas and Electric Company's SmartAC Program*. Prepared by Freeman, Sullivan & Co. <http://fscgroup.com/reports/2012-smartac-evaluation.pdf>

PSE (2012). Home Energy Reports Program: Three Year Impact Behavioral and Process Evaluation. Puget Sound Energy. Prepared by DNV KEMA.

Rosenberg, M.; Kennedy Agnew, G.; Gaffney, K. (2013). Causality, Sustainability, and Scalability – What We Still Do and Do Not Know about the Impacts of Comparative Feedback Programs. Paper prepared for 2013 International Energy Program Evaluation Conference, Chicago.

SMUD (2011). Sacramento Municipal Utility District Home Energy Report Program Impact and Persistence Evaluation Report, Years 2008-2011. Prepared by May Wu, Integral Analytics.

SMUD (2013). SmartPricing Options Interim Evaluation. Prepared for the U.S. Department of Energy and Lawrence Berkeley National Laboratory by Sacramento Municipal Utility District and Freeman, Sullivan, & Co.

Seed, P.T. (1997). "Sample Size Calculations for Clinical Trials with Repeated Measures Data." *Stata Technical Bulletin* 40, 16-18.

Shadish, W.R., Cook, T.D.; Campbell, D.T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cenage Learning.

Stewart, J. (2013). Peak-Coincident Demand Savings from Behavior-Based Programs: Evidence from PPL Electric's Behavior and Education Program. UC Berkeley: Behavior, Energy and Climate Change Conference. Retrieved from: <http://escholarship.org/uc/item/3cc9b30t>.

SEE Action (2012). Evaluation, Measurement, and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations. State and Local Energy Efficiency Action Network. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman, Lawrence Berkeley National Laboratory. <http://behavioranalytics.lbl.gov>.

Todd, A. (2014). Insights from Smart Meters: The Potential for Peak-Hour Savings from Behavior-Based Programs . Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory. LBNL Paper LBNL-6598E. Retrieved from: <http://escholarship.org/uc/item/2nv5q42n>