FILE

OPOWER 1515 N Courthouse Rd. Sixth Floor, Arlington, VA 22201 · Tel 703.778.4544 · Fax 703.778.4547 · opower.com

FAX 30

RECEIVED-DOCKETING DIV

2010 NOV -3 PM 3: 12

PUCO

****TRANSIMTTED VIA FACSIMILE****

November 3, 2010

Public Utilities Commission of Ohio
Docketing Division
180 East Broad Street
Columbus, Ohio 43215-3793

Phone: (614) 466-4095 | Fax: (614) 466-0313

| | | | |
|---|---|---|---|
| To: | Public Utilities Commission of Ohio Docketing Division | From: | OPOWER, Inc. |
| Fax: | (614) 466-0313 | Pages: | 29 |
| Date: | November 3, 2010 | Case No.: | 09-512-GE-UNC |
| Re: | In the Matter of the Protocols for the Measurement and Verification of Energy Efficiency and Peak-Demand Reduction Measures | | |

To Whom It May Concern:

OPOWER, Inc. ("OPOWER"), an energy efficiency and smart grid software company, respectfully submits the enclosed objections to the Public Utilities Commission of Ohio ("the Commission") with respect to the draft Technical Reference Manual ("TRM") submitted in Case 09-0512-GE-UNC and pursuant to the procedures established in the June 24, 2009 entry in that case.

Please do not hesitate to call with any questions or concerns regarding this submission.

Sincerely yours,

Michael Sachse
Sr. Director of Government Affairs & General Counsel | OPOWER, Inc.
Phone: (571) 384-1257 | Fax: (703) 778-4547

**BEFORE**

**THE PUBLIC UTILITIES COMMISSION OF OHIO**

| | |
|---|---|
| **In the Matter of the Protocols for the Measurement and Verification of Energy Efficiency and Peak-Demand Reduction Measures** | **Case No. 09-512-GE-UNC** |

**OBJECTIONS TO THE DRAFT TECHNICAL REFERENCE MANUAL SUBMITTED BY OPOWER, INC.**

## Table of Contents

## I. Introduction and Statement of the Case

OPOWER, Inc. ("OPOWER"), an energy efficiency and smart grid software company, respectfully submits the following objections to the Public Utilities Commission of Ohio ("the Commission") with respect to the draft Technical Reference Manual ("TRM") submitted in Case 09-0512-GE-UNC and pursuant to the procedures established in the June 24, 2009 entry in that case. It is OPOWER's opinion that the inclusion of an evaluation, measurement and verification ("EM&V") protocol for behavioral energy efficiency programs, based on current and accepted best practices, will result in greater transparency in the measurement and reporting of results.

Currently partnering with 42 utilities across 21 states, including seven of the ten largest U.S. utilities, OPOWER's software has become the customer engagement platform of choice in leading behavioral energy efficiency programs. The OPOWER Home Energy Reporting program offers a cost-effective way to convert data into insights that deliver potential energy efficiency gains to the customer directly. OPOWER's platform – with results verified by the rigorous experimental design proposed in the attached protocol – has been consistently effective in driving energy consumption reductions in each deployment to date.

As an energy efficiency company with extensive experience in applying behavioral science and data analytics to drive energy savings, OPOWER encourages the Commission to include the protocol provided in Appendix A, or similar, in the TRM. It is OPOWER's experience and opinion that the adoption of such a protocol will allow for transparent, verifiable results for the state's behavioral energy efficiency programs.

## II. Argument

OPOWER offers the following points for consideration by the Commission:

- Behavior-based programs have been proposed before the Commission and are already underway in Ohio.

- Providing a protocol that utilizes experimental design and ex-post measurement in the TRM will allow for greater transparency and regulatory certainty moving forward.

- Behavior-based programs are proven to generate *significant, cost-effective* energy savings. Through experimental design, energy savings generated through such programs have been rigorously measured and independently evaluated in various large-scale pilots across the country. This record of success has led commissions in 14 states to accept behavior-based savings as a measurable source of energy efficiency.

- Best practices to evaluate behavior-based programs have been established by an extensive body of research on efficiency programs such as those administered by OPOWER and from other disciplines. These tools are assembled in a model EM&V protocol for behavior-based programs, attached as Appendix A, and a list of relevant resources is provided in Appendix B.

- The experimental design described in these comments allows behavior-based programs to isolate and properly allocate behavior-based savings from energy savings that could be claimed by other programs ("double counting").

### A. Behavior-based programs have been proposed and implemented in Ohio.

The inclusion of a protocol for measuring behavior-based programs is of particular importance to Ohio's ratepayers and utilities, considering that American Electric Power ("AEP") Ohio is currently administering a behavior-based program and Duke Energy Ohio ("Duke") has proposed a behavior-based efficiency program before the Commission.

AEP has successfully deployed the behavior-based Home Energy Reporting Program using OPOWER's platform to 150,000 customers with traditional meters and to an additional 65,000 AMI-enabled households.[1] Both deployments employ experimental design and results are measured using the methodology described in Appendix A. Furthermore, by leveraging experimental design, results from the deployment to the AMI-enabled households will allow greater understanding of the effect of behavior-based messaging on peak demand.

Similarly, Duke Energy Ohio proposed a "Home Energy Comparison Report" (HECR) program for residential customers in their Energy Efficiency and Peak Demand Reduction Programs Portfolio filing with the PUCO on December 29, 2009.[2] As described in this filing, the HECR is a pilot program designed to drive behavior-based energy efficiency by providing customers with information on opportunities for potential savings through periodic comparative usage data reports that include specific energy saving recommendations.

Two of the largest utilities in Ohio have turned to behavior-based efficiency to meet their reduction targets, marking a great opportunity for further energy conservation in the state – provided standards are adopted for the evaluation, measurement, and verification of these savings. Including a protocol reflecting best practices in EM&V for behavior-based energy efficiency programs will achieve this very purpose.

## B. Establishing a clear protocol for measuring behavior-based savings will provide transparency of results and regulatory certainty.

The current draft of the TRM does not identify an official measurement method for behavior-based programs. Given the interest by utilities in Ohio to deploy behavior-based programming on a larger scale, it is in the interest of all parties to officially establish a clear and rigorous measurement methodology. By doing so, the Commission will remove

[1] As described on page 24 of the AEP Ohio Testimony, Case No. 09-1999-EL-POR
[2] As described on pages 28 – 30 of the Duke Energy Ohio, Inc.'s Energy Efficiency and Peak Demand Reduction Programs Portfolio, Case No.09-1089-EL-POR and 09-1090-EL-POR

uncertainty over the measurement of such programs and foster innovative approaches that achieve the dual goals of increased efficiency and greater customer engagement.

OPOWER proposes herein an EM&V protocol for measuring savings attributable to behavior-based programs (Appendix A), which applies an experimental design approach already well established in the field.

The protocol contains several key components:

- *Experimental Design*

The attached protocol is organized around the use of experimental design to measure the impact of behavioral messaging on energy efficiency. This means that for each program, statistically identical control and treatment groups need to be established from a population of eligible customers. The control group provides a baseline against which energy savings in the test, or treatment, group are measured. Because the comparisons are made in real time, nuisance factors, such as variations in weather and energy prices, are neutralized because the same conditions apply to both treatment and control. This simple yet robust treatment-and-control methodology provides a strong foundation upon which to measure the impact of behavioral messaging.

For example, consider Sacramento Municipal Utility District's ("SMUD's") behavior-based program. Together with OPOWER, SMUD launched its behavior-based program to 35,000 homes, while maintaining a 50,000 home control group. The two groups were randomly selected and had no statistically significant difference in their energy consumption prior to deployment. Since deployment, the impact has been clear – over twenty months, behavior-based messaging has decreased consumption by 2.5% in the test group. Because the groups are, in the aggregate, identical—save for the fact that one group receives the reports while the other does not—the difference in energy savings may safely be attributed to the Home Energy Reporting program. Relevant to this proceeding, the very same methodology is used to measure results in the AEP program.

This test and control methodology, widely used in other analytical fields, is explicitly endorsed for energy efficiency use by the California Public Utilities Commission (CPUC),[3] in

[3] California Public Utilities Commission, D.10-04-029

the California Evaluators Protocols,[4] and the guidelines for the National Action Plan for Energy Efficiency,[5] which was jointly produced by the US Department of Energy and the Environmental Protection Agency.

- *Ex-Post Measurement*

*Ex-post* measurement is both the most accurate methodology to detect savings generated by behavior-based programs and the best way to hold program providers accountable for the efficacy of their programs.

*Ex ante* measurement is inappropriate for behavior-based approaches because the key variable in behavior-based efficiency is not the number of homes reached, but the level of energy savings actually achieved. That savings level is highly dependent upon characteristics that are unique to each program administrator. An administrator that is effective in its messaging and micro-targeting – for example, recommending pool pump replacement to homes with pools – should achieve greater savings than one that sends out only generic messages. That relative efficacy can only be accurately measured *ex post*.

- *Billing Analysis*

Like experimental design, billing analysis is an ideal tool for evaluating behavior-based savings. Traditional billing analysis techniques are directly applicable to behavior-based savings – the only notable difference is that larger sample sizes need to be used to allow for broad-based programs with smaller percentage in energy savings per customer (e.g. <5%) to be measured with the appropriate statistical rigor. The results of a billing analysis on a behavioral change program are just as reliable as for traditional efficiency programs provided that the behavioral program is experimentally designed with sufficient sample size.

- *Opt-out program design*

Opt-out program design allows for rigorous EM&V by assigning customers to the participant and non-participant groups at random. The randomization procedure ensures

---

[4] California Public Utilities Commission. *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, April 2006

[5] National Action Plan for Energy Efficiency. *Model Energy Efficiency Program Impact Evaluation Guide*. Prepared by Steven R. Schiller, Schiller Consulting, Inc., December 2007

that any unobservable characteristics – like attitudes, beliefs, behaviors, attention paid to direct mail, etc. – are balanced between the participant and non-participant groups. As a result, one can draw a causal, unbiased inference about the impact of the program. Furthermore, opt-out program design allows for the creation of very large sample sizes required to establish the statistical significance necessary when detecting relatively small savings impacts (1-3%) on a per home basis.

By contrast, an opt-in program is difficult to measure with certainty. This is because the most significant challenge when measuring an opt-in program is the creation of a relevant and unbiased comparison group. Although there are a variety of statistical techniques one can use to match participants with non-participants based on observable characteristics – such as housing data, demographic data, and census data – none of these methods address differences in those unobservable characteristics. While a "matched" comparison group may appear to be similar to the treatment group, it is likely that undetected biases will render the measured savings invalid. This is especially true in the case of opt-in programs: the act of opting-in signals a difference from those who did not opt-in, otherwise known as responder (or selection) bias.[6]

### C. Behavior-based programs are proven to generate measureable, cost-effective savings when employing experimental design and measuring results *ex post*.

In recent years, many utilities have partnered with companies like OPOWER to run large-scale pilots proving that behavior-based savings can be cleanly measured with proper program design. These implementations have demonstrated that behavior-based efficiency is cost-effective and, if implemented using randomly selected test and control populations, is measureable at high confidence levels. The results from these pilot programs have been independently evaluated and confirmed by Summit Blue (d/b/a Navigant), Power System Engineering, Yale University economist Ian Ayers, and Massachusetts Institute of

[6] This responder bias (also known as "survey responder bias") has made opt-in programs disfavored for EM&V purposes. As the Electric Policy Research Institute observed, "Matching methods by themselves are to be used sparingly because they are prone to the introduction of bias that cannot be anticipated or measured. The calculated estimates of differences (or difference of differences) are biased (they cannot be inferred to reflect the real values) and inconsistent (the variance is large and unknown, so we cannot make statements about the confidence interval around the estimate). These constitute a strong cautionary." Electric Policy Research Institute. *Guidelines for Designing Effective Energy Information Feedback Pilots: Research Protocols*, p. 3-18.

Technology Professor Hunt Alcott.[7] In short, there is consensus that behavior-based savings are significant and measurable, and that experiments using randomized treatment and control groups is the program design that best allows for transparent EM&V. By applying the principles of experimental design, the Commission can ensure that the impacts of behavior-based programs in Ohio will be reliably and accurately measured.

As stated previously, each of OPOWER's Home Energy Reporting programs are designed using a simple test and control methodology and results are measured *ex post*, just as they are in each independent evaluation. By using test and control groups and *ex-post* measurement, OPOWER is able to isolate and cleanly evaluate the impact of its program.

### a. OPOWER Results

OPOWER's Home Energy Reporting program has been consistently effective in each deployment to date. Every utility with at least twelve months of results has achieved energy savings between 1.5% and 3.5% (see Figure 1). These results have been consistent across electric and gas utilities, as well as in winter-peaking, summer-peaking, and mild climates. Furthermore, programs deliver savings when ratepayers –and utilities– need them most.

[7] See Appendix B for a list of relevant resources.

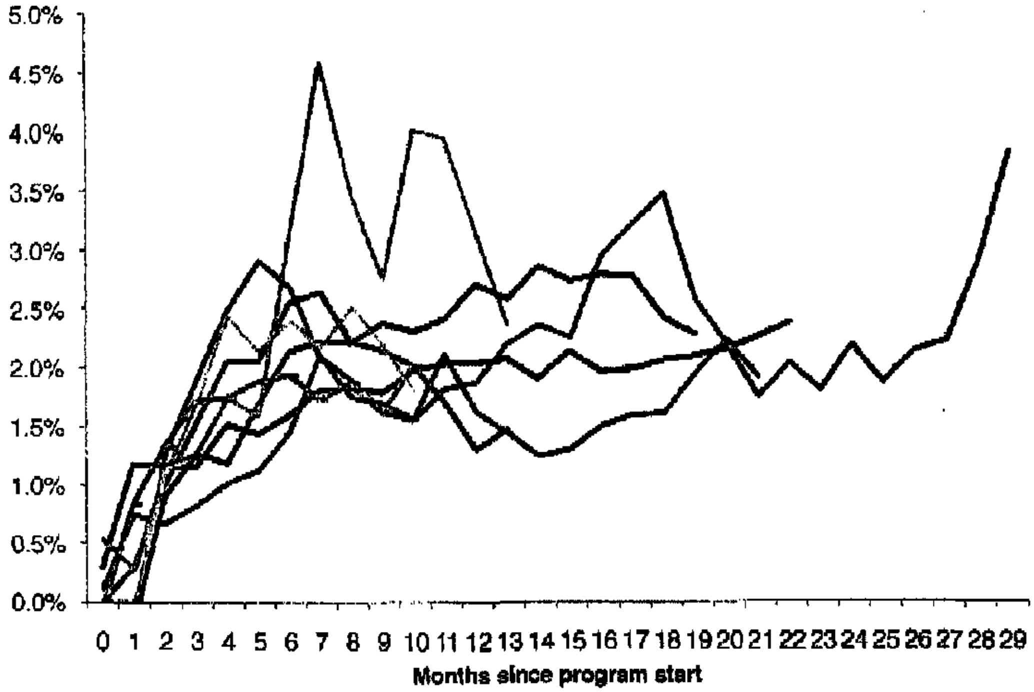


**Figure 1: Consistency of savings achieved through OPOWER's program**

As Figure 1 shows, results from these behavior-based programs are often seen in the first one or two months. Furthermore, in each of these deployments, energy savings have consistently increased over time. For example:

- With SMUD, the savings in the second year of the program have been greater than the savings in the first year. After 30 months, these savings are not only continuing, but are increasing – in the second year of the program, SMUD customers were saving 22% more energy than the year before.[8]

- Electricity savings in Puget Sound Energy, a large IOU in Washington, are now more than 2% with nearly two years of results. OPOWER works with this dual-fuel utility to use behavioral messaging to target electricity and natural gas use on the same report. Savings for the last six months have been 2.04% and 1.43% for electricity and natural gas, respectively.

[8] Summit Blue Consulting. *Impact Evaluation of OPOWER SMUD Study*, September 2009. http://www.opower.com/LinkClick.aspx?fileticket=naU7NN5-430%3d&tabid=72

- Connexus Energy, a large electric distribution cooperative in rural Minnesota, has seen savings of more than 2%.[9] Cumulative savings after 15 months are more than 2.3%, and averaged more than 2.6% in the winter of 2010.

## D. Other states have already accepted behavior-based programs evaluated with experimental design.

The strong, verified results from behavior-based programs have led other states to accept behavior-based programs as efficiency resources. Regulatory authorities in Massachusetts, Minnesota, and California have expressed support for behavior-based programs evaluated with experimental design. Enabled by their regulators, utilities in these states have moved forward with their energy efficiency portfolios to include behavior-based programs.

- California - On Thursday, April 8, 2010, The California Public Utilities Commission (CPUC) adopted a protocol to count energy savings from behavior-based energy efficiency programs in a decision regarding EM&V of energy efficiency programs for 2010 through 2012 (D.10-04-029). The decision made California the third state to count behavior-based results towards energy efficiency goals, joining Minnesota and Massachusetts. Key elements of the decision include:
    - Ex-post measurement methodology: The results from OPOWER's program will be measured only after the savings have been incurred.
    - No limits on type or size of deployment: Utilities are able to count the savings from both residential and non-residential deployments, at any size.

  These two requirements create an environment that rewards rigor while encouraging innovation. Ex post evaluation ensures that ratepayer dollars are spent wisely, while unlimited deployment capacity leaves California's utilities free to choose the most cost effective efficiency resources.

- Massachusetts – The Massachusetts Department of Energy Resources ("DOER") and program administrators from its IOUs have decided to include behavior-based

[9] Alcott, Hunt. *Social Norms and Energy Conservation.* February 2010. Available online at: http://web.mit.edu/allcott/www/Allcott%202010%20-20Social%20Norms%20and%20Energy%20Conservation.pdf

programming as part of the state's three-year efficiency plan. DOER noted, "One successful organization upon whose work the Program Administrators would like to build is Positive Energy [now OPOWER], a corporation that is committed to persuading consumers to save energy through a combination of technology, analytic direct marketing, and behavioral science."[10] Accordingly, National Grid has made behavior-based programming one of its largest sources of efficiency.

- <u>Minnesota</u> - The Office of Energy Security (OES) has approved OPOWER's Home Energy Reporting program for Centerpoint Energy, one of Minnesota's largest utilities. Indeed, OES was effusive in its praise of behavior-based programming: *OES Staff are pleased to see that CPE [Centerpoint Energy] will be starting the Residential Home Energy Reports project in 2010. Recent evaluations of programs across the country and in Minnesota suggest that home energy reports are a cost-effective way to educate customers and encourage energy saving behavior. CPE plans to include 225,000 residential customers, approximately 30 percent of the Company's residential customers, in this program by the third year of its triennial plan. This project is also expected to be one of the largest drivers of new energy savings in the Company's Residential Segment. [...] In future filings, the energy savings claimed by the Company should reflect the actual energy savings associated with the project based on measurement and verification by Positive Energy [now OPOWER].*[11]

Furthermore, commissions in 14 states, including Ohio, have approved the OPOWER platform for utilities to help them achieve their efficiency goals. See Appendix C for a list of those states and utilities.

### E. Existing studies and evaluations indicate established best practices regarding EM&V of behavior-based programs.

An extensive body of research and evaluations indicates that ex-post measurement of program effects, using randomized test and control groups, is the established best practice

[10] Massachusetts Joint Statewide Three-Year Electric Efficiency Plan: 2010-2012, p. 238

[11] Office of Energy Security, Proposed Decision, October 1, 2009, p. 23. Behavior-based programming was approved in the Final Decision dated November 23, 2009.

for evaluating behavior-based programs. A select list of existing studies – provided in Appendix B – demonstrates that leading economists, social scientists, and professional program evaluators have indicated that using experimental design and ex-post measurement is the preferred methodology in randomized field studies, particularly those that examine changes in behavior.

Professors Hunt Allcott, of MIT, and Sendhil Mullainathan, of Harvard University, recently published a peer-reviewed discussion of behavior change and energy use in *Science*, the leading journal of the natural sciences. [12] Their review of existing evaluations and academic literature brought them to the following conclusion:

> "Although laboratory studies and small-scale pilots demonstrate academic insights and proofs-of-concept, scalable behavioral interventions require in situ testing. OPOWER illustrates this: It would be difficult to predict the effects without randomized, controlled field trials in a representative population. [...] In our own work testing behaviorally informed interventions, we have seen how the long-understood insight of randomization can be made practical. Useful techniques include randomizing letter content across groups, encouragement designs that simultaneously evaluate program marketing and the program itself, and phased implementation. In some settings, outcomes can be measured with little additional cost; utilities, for example, already record their customers' energy consumption. In the OPOWER example, it is straightforward to send letters to a study group and not to a group of controls, and effects are measured simply by comparing the two groups' electricity bills."

Several other economic reviews and studies provided in Appendix B indicate that randomized experiments and field trials are the extremely valuable and accurate,

[12] Alcott, Hunt and and Sendhil Mullainathan. Behavior and Energy Policy. *Science*. March 2010. Available online at:
http://web.mit.edu/allcott/www/Allcott%20and%20Mullainathan%202010%20-%20Behavioral%20Science%20and%20Energy%20Policy.pdf

particularly when measuring changes in behavior.[13] A number of the other studies included in Appendix B also point to the cost effectiveness of behavior-based efficiency programs.[14]

### F. A methodology utilizing experimental design is able to isolate savings attributable to behavior-based programs.

The EM&V protocol proposed in Appendix A fully addresses potential concerns over double counting of savings associated with other IOU programs, including rebated measures, upstream programs such as CFLs, and the savings claimed in AMI business cases, as well as questions about the persistence of the behavioral measure. Each of those issues is discussed in detail below.

However, it is important to note that it is likely that most of the savings resulting from behavior-based programming do not overlap with other efficiency measures. In surveys of homes that have received behavior-based messaging, the most reported energy saving actions are turning off lights, adjusting thermostats and unplugging appliances. These findings are consistent with a study by Wilhite and Ling, which found that savings were sustained over the course of a 3-year informative billing study, but participants could not recall the energy saving actions they had taken; the authors of the study confirm, "Our impression from the interviews is that after three years the changes people made had become so routine that they had trouble identifying them." [15] Thus, while it is necessary to

---

[13] See: Banerjee, Abhijit and Esther Duflo, 2008. "The Experimental Approach to Developmental Economics." NBER Working Paper.
Levitt, Steven D, 2008. "Field experiments in economics: The past, the present, and the future." European Economics Review. Vol. 53 (1)
Davis, Lucas, 2008. "Durable goods and residential demand for energy and water: Evidence from a field trial." RAND Journal of Economics.
Imbens, Guido, et al., 2009. "Recent Developments in the Econometrics of Program Evaluation." Journal of Economic Literature.

[14] See: Ayres, Ian, et al., 2009. "Evidence From Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage." NBER Working Paper.
Allcott, Hunt, 2010. "Social Norms and Energy Conservation." Working Paper, Massachusetts Institute of Technology's Center for Energy and Environmental Policy Research.

[15] Wilhite H and R Ling (1995) "Measured energy savings from a more informative energy bill." *Energy and Buildings* 22 pp145-155

take steps to avoid double counting of results, the actual scope of savings overlap between behavior-based programs and other efficiency measures is likely to be small.[16]

OPOWER recognizes that correct attribution of savings is critical to the fair accounting of portfolio efficiency standards and offers the following protocol for addressing double-counting related issues. For most efficiency programs, double counting can be addressed through these two steps: (1) measure program participation in treatment and control groups; and (2) measure the overlap effect – that is, attribute savings coming from any additional program participation (vs. control group) in the treatment group (e.g. exercising a rebate) to the programs that finance the rebate – not to the behavioral program.

(1) *Measure program participation in treatment and control groups*

There are two ways to establish other program participation across the population participating in the behavioral program. The correct method depends on whether or not the program is individually tracked.

- *Individually tracked programs*

For individually tracked programs, the utility can track specific customer participation. This scenario covers the vast majority of programs implemented in the residential sector and range from air conditioner rebates to home energy surveys. To avoid "double counting," utilities simply must continue to track the participation in these programs on an individual household basis, and the difference in frequency of participation can be compared from the treatment to the control. Then the utility may choose to either (a) subtract the deemed savings from the additional installed measures in the treatment group, or (b) add the costs of the additional installed measures to the cost of the behavior change program and count the savings.

- *Non-Individually tracked programs*

In the case of "upstream" subsidies the method to assess double counting is to perform surveys that measure the increase in the installation of the subsidized measures in both the treatment and control groups. The survey should be done in a statistically rigorous fashion,

[16] In an analysis performed after 18 months of the SMUD program, OPOWER estimated that less than 2% of the reported energy savings are overlapped with savings reported by other efficiency programs at SMUD.

with the results achieving a minimum precision of 90% and power of 0.8. Once these rates of use are established, the energy savings stemming from the increase in installed measures in the treatment group can then be accounted for in the same fashion increases from *individually tracked programs are handled.*

(2) *Measure the overlap effect*

Experimental design allows for a clear view of the impact that behavior-based programs have on other efficiency measures and limits the potential for double counting. For example, if 100 homes in the control group install efficient furnaces, and 120 homes in the treatment group do the same, the savings from the additional 20 furnaces installed will be reflected in the overall energy savings reported by the behavioral program, but can be easily identified, allowing the Commission to account for those energy savings accordingly, i.e. attributed to either the behavioral program or the furnace rebate program, but not both.

Figure 2 illustrates an example in which the reports lead to increased participation in utility programs. The savings generated from installations that occur in both groups ("A" and "B" in the figure) cancel each other out and are not reflected in overall savings measured as the difference in energy use between the treatment and control groups. However, the incremental installations that occurred as a result of receiving the behavior-based program ("C") do show up in the overall savings estimates. The total kWh or therms associated with the incremental installations can be estimated using the deemed savings for each type of installed measure. This process can be repeated across each type of measure offered by the utility.

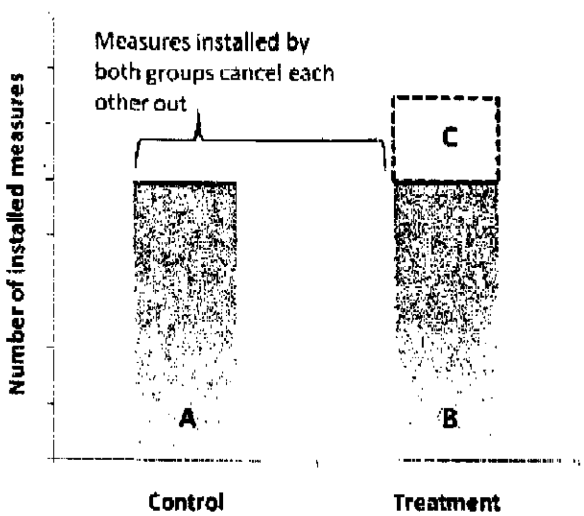


Figure 2: Double counting mechanism

1. Savings measured as a difference between treatment and control groups do not include measures installed by both groups (areas A + B)
2. The potential for double counting only exists when the Treatment group installs additional measures (area C)
3. Savings due to incremental measures are easily identified, allowing utilities to account for them accordingly

Because of the experimental approach used for program design and measurement, the potential for double counting is limited to the difference in participation between the two groups, not the absolute level of participation. Thus, utilities must decide how to account for this component ("C") in their internal accounting.

There are two ways the Commission can account for energy savings that were partly achieved as a result of behavioral messaging, and partly due to the financial incentive provided via another energy efficiency program (e.g. a rebate). The first is to subtract the incremental energy savings from the program providing the financial incentive. The second is to subtract the same savings from the total impact estimate of the behavioral program.

Regardless of the option chosen, the approach described above provides a rigorous procedure for identifying and accounting for energy savings, and ensures that ratepayers are not paying twice for the same savings.

## III. Conclusion

For the foregoing reasons, it is OPOWER's opinion that the Commission should order that a protocol for measuring behavior-based efficiency programs, recommending the use of experimental design and ex-post measurement of savings, be included in the Ohio Technical Reference Manual. By doing so, the Commission will put a methodology into place that yields clear and unambiguous savings attributable to the state's behavioral energy efficiency programs.

Respectfully submitted,

Michael Sachse
Senior Director of Government Affairs & General Counsel
OPOWER, Inc.
1515 N. Courthouse Rd.
Arlington, VA 22201
(571) 384 1257
Michael.sachse@opower.com

# Appendix A:
# Measurement & Verification Protocol for Behavior-Based Efficiency Programs

## Description of Measure

Behavior-based programs are proven to generate significant, cost-effective energy savings. Through experimental design, energy savings have been rigorously measured and independently evaluated in numerous large-scale pilots across the country. There are a significant number of evaluations supporting the methodology described in the following protocol that have been performed by academics and professional evaluators.[1] This protocol reflects the best practices established through that body of work.

This evaluation protocol describes a method for evaluating behavior-based savings for residential utility customers. The methods specified here allow for rigorous evaluation of behavior-based savings by applying techniques already applied in a number of states, including Ohio. Specifically, the methodology described in this protocol:

- Allows behavioral programs to achieve the definition of verified savings as specified by the Ohio *Green Rules* as "an annual reduction of energy usage or peak reduction from an energy efficiency or peak-demand reduction program directly measured or calculated using reasonable statistical and/or engineering methods consistent with approved measurement and verification guidelines";[2]
- Follows the guidelines for Billing Regression Analysis specified in the IPMVP for whole-facility measurement;[3]
- Is endorsed by the National Action Plan for Energy Efficiency guidelines under the described methodology for "Large Scale Data Analysis";[4]
- Fully accounts for double-counting of savings with current efficiency programs and AMI-enabled conservation; and

---

[1] See: Allcott, Hunt, 2009. Social Norms and Energy Conservation. MIT Center for Energy and Environmental Policy Research working paper.

Allcott, Hunt, and Sendhil Mullainathan, 2010. Behavior and Energy Policy. *Science*. Vol. 327

Ayers, Ian, Sophie Raseman, and Alic Shih, 2010. Evidence from Two Large Field Experiments that Peer Comparison Feedback can Reduce Residential Energy Usage. NBER working paper No. 15386.

Levitt, Steven D. and John List, 2008. Field Experiments in Economics: The Past, the Present, and the Future. NBER working paper 14356.

Power System Engineering, 2010. Measurement and Verification Report of OPOWER Energy Efficiency Pilot Program (Connexus Energy)

Summit Blue Consulting, 2009. Impact Evaluation of OPOWER SMUD Pilot Study

[2] Ohio State Rule Code 4901:1-39-01 1-39-06

[3] *International Performance Measurement & Verification Protocol (IPMVP); Concepts and Options for Determining Energy and Water Savings: Volume I.* Section 4.9.4 and Appendix B-2. Efficiency Valuation Organization, September 2009. EVO 10000 – 1:2009

[4] *NAPEE Model Energy Efficiency Program Impact Evaluation Guide.* Section 4.4, p. 4-10. 2007

- Fully accounts for double-counting of savings with current efficiency programs and AMI-enabled conservation; and
- Can be executed by utilities in a cost-effective and timely fashion, using existing measurement protocols and software packages.

The types of programs that this protocol will apply to include residential energy efficiency behavioral programs that promote efficient behavior, customer engagement, and individual energy management. Behavior-based programs may include one or more of the following characteristics:

- Normative comparison of a customer's usage against comparable customers in the same geographical area
- Targeted conservation and peak reduction tips based on an analysis of a customer's past usage and individual profile
- Alerts and tips to reduce usage during peak events
- Encouraging participation in other programs in a utility's efficiency portfolio based on previous usage patterns and individual consumer profile

Information from behavioral programs may be delivered to the customer through direct mail, a utility or vendor website, and/or a display in the consumer's home.

## Measure Life

While there is evidence that behavior-based program results persist, behavior-based programs only require a single-year measure life, thereby reducing any risk associated with uncertain future performance. No assumptions are made regarding the full "lifetime" savings of behavior-based program beyond the actual measurements. Likewise, any costs associated with the program (including measurement and verification) are attributed to the program in the year they are incurred. There is no amortization of program costs beyond the program length, nor are any future efficiency savings considered part of the behavioral intervention. As a result, this measurement strategy can be considered as a series of single years of actual measurement, being summed for as long as the program is being run and results are being measured.

## Definition of Efficient and Baseline Cases

The baseline case is defined first by collecting energy usage information for both the test and control groups to establish a pre-treatment baseline, and then observing energy use among the control group to establish a post-treatment baseline after the program has begun. The efficient case will be determined by measuring the energy savings in the test group – i.e., those customers receiving the treatment – versus the control group.

## Calculation of Savings

This protocol may be applied to programs administered by either natural gas or electric utilities and provides a methodology for measuring energy savings for individual utility customers. The protocol occurs in three distinct phases:

1. **Phase 1: Program Setup.** Describes the setup needed to employ experimental design to accurately evaluate the impact of behavior-based programs.

2. **Phase 2: Billing and Survey Analysis.** Outlines the statistical methods required to accurately measure energy savings as well as the data needed to properly attribute savings where there is overlap with another efficiency program.

3. **Phase 3: Reporting and Accounting of Savings.** Provides guidelines for applying survey and billing data to properly report and attribute program savings.

## Phase 1: Program Setup

### Step 1: Identify target population

Program setup work must be conducted prior to launching the behavior-based program and, while Steps 1-3 are not directly descriptive of the evaluation methodology, these steps are critical to measuring and verifying the resulting savings in an accurate and transparent manner.

Indentifying the universe of participants is the first step in the program setup process. Participants will vary depending on the goal of the implementing utility. For example, a utility could choose to focus on high usage homes, small commercial enterprises, or low-income populations. Any of the following factors could be used to determine potential participants:

- Fuel type (electric and/or natural gas)
- Customer demographics
- Availability and quality of billing or consumption data
- Participation in other efficiency programs
- Presence of specific technologies (AMI, HAN, electric vehicle, customer-owned generation, etc)
- Historical energy consumption
- Other criteria (income level, usage patterns, etc)

Inclusion and exclusion criteria must be applied from the start, before participants are assigned to treatment or control groups. The resulting population of eligible customers must be large enough to yield a statistically significant result as determined by the power analysis outlined in Step 2.

### Step 2: Match program size to expected magnitude of impact

Once the potential participant universe has been defined, statistical power analyses must be conducted to determine the sample sizes required to achieve the required level of precision. The sample sizes will depend upon the expected impact of the program, the required level of statistical significance, the desired power for the experiment, and the coefficient of variation in the target variable (consumption, peak demand, etc). For example, a residential program expected to deliver a 10% reduction in energy consumption needs roughly 800 participants in each group (split evenly between the treatment and control groups) to achieve an 80% power.[5] A program expected to deliver 2% savings will need at least 19,600 participants in both treatment and control groups to achieve the same power.[6]

Most behavior-based programs will have heterogeneous treatment effects – that is, the program will work better in some customer segments than others. If the program designer wishes to evaluate the program results for specific population segments, the appropriate power analyses must be conducted at the segment level. To extend the example above, if the program goal was to measure the results across five equally sized demographic segments (such as income), then a program expecting 10% savings would need roughly 5*800 = 4,000 participants, while a program expecting 2% savings will require at least 5*19,600 = 98,000 participants.

Given that behavioral programs can be easily scaled, it is recommended that an enhanced level of statistical precision[7] only possible with large deployments be required. In practical terms this means that for every level of expected impact, there is a minimum number of program participants required in order to achieve the desired statistical precision in the billing analysis described in Step 4. Table 1 below can be used as a guide for minimum program size requirements for different levels of expected demand reduction, ranging from 1% to 10%.[8]

**Table 1: Minimum required sample size for expected level of impact**

| Expected Impact | Sample size required for 90% precision | Sample size required for 95% precision |
|---|---|---|
| 1% | 61,826 | 78,490 |
| 2% | 15,458 | 19,624 |
| 5% | 2,474 | 3,140 |
| 10% | 620 | 786 |

**Step 3: Establish valid test and control groups**

After the target population is identified, participants should be randomly assigned to treatment and control groups, rendering them statistically identical. Randomization is the only assignment algorithm guaranteed to ensure internal validity and allow program evaluators to draw causal linkages between the treatment and the measured effect.

## Implementation

[5] Power analysis, in this case, is used to calculate the minimum sample size required to accept the outcome of the statistical test with a particular level of confidence.

[6] Both examples assume an alpha of 0.05 (corresponding to 95% confidence intervals) and a coefficient of variation of 0.5, which is typical for residential programs.

[7] It is recommended that the program achieve 90% precision and a power of 0.8, at a minimum.

[8] Calculations assume a power of 0.8 and a coefficient of variation of 0.5. Reported sample sizes include participants in both the treatment and control groups

Once the treatment and control groups have been randomly selected from the target population identified in Step 1, the program is ready to be administered. Note that it is critical that the program is made available only to those customers in the treatment group and not to those in the control. If the control group is contaminated the validity of any measured impact can be called into question.

**Adjusting Control Group As Program Expands**

Successful programs will often be expanded to non-participants over time. In order to maintain robust measurement, a control group must be maintained. The control group, however, does not need to grow as the treatment group grows; so long as the new participants come from the same population, the original control group remains a valid basis of comparison. There are two situations in which the control group may need to change in order to accommodate an expanded program:

1. **Additional participants differ from the original test group** - If the program is expanded to participants outside the initial target population, the selection process for the program expansion must follow the protocols laid out in Step 1. The expansion will require a new determination of inclusion/exclusion criteria, new power analysis, and a new randomization procedure to assign homes into treatment and control.

2. **Additional participants come from the original control group** – A utility may desire to take homes in the control group and place them in the treatment group. It may do so without jeopardizing the effectiveness of the experimental design so long as the control group remains large enough to continue robust measurement as determined by a power analysis (Step 2).

## Billing and Survey Analysis:

***Step 4: Perform Statistical Billing Analysis***

Performing a billing analysis using properly specified regression models is the preferred approach when evaluating a large-scale, experimentally designed behavior program, as specified by NAPEE.[9] Billing analysis is the preferred methodology when:

1. Both pre and post-treatment billing data are available;
2. Expected program impacts can be expected to be observed in a billing analysis; and
3. The analysis is of a program with larger numbers of participants that are more homogeneous.

Any program that follows the principles laid out in the Program Setup section above should satisfy these criteria to perform a randomized control trial. If the appropriate power calculations have been performed, experimentally designed programs of sufficient sample size can use billing analysis to detect changes in consumption as small as 0.5%.

In order to implement a randomized control trial, the sample of customers eligible to participate in the program must be carefully selected, as outlined in Step 1 above. If participants have been randomly

[9] *NAPEE Model Energy Efficiency Program Impact Evaluation Guide.* Section 4.4, p. 4-10. 2007

assigned to the treatment and control groups prior to the launch of the behavioral program, there is virtually no risk of selection bias and the results of the regression analysis will have internal validity.

Several regression techniques can be used for billing analysis. Roughly, all such models should have functional forms similar to:

$$E_{it} = \alpha_i + \boldsymbol{X}_{it}\beta + \delta_1 \boldsymbol{T}_i + \delta_2 \boldsymbol{P}_{it} + \delta_3 \boldsymbol{T}_i \boldsymbol{P}_{it} + \varepsilon_{it}$$

Where

| | | |
|---|---|---|
| $E_{it}$ | = | Average daily energy consumption for customer *i* in period *t* |
| $\alpha_i$ | = | Household fixed effects |
| $\boldsymbol{X}_{it}$ | = | Matrix of time-varying household coefficients, including heating and cooling degree days |
| $\boldsymbol{T}_i$ | = | Vector of treatment indicator variables, 1 if household *i* is in the treatment group, otherwise 0 |
| $\boldsymbol{P}_{it}$ | = | Matrix of post-treatment indicators, 1 if period *t* is after the program launch for household *i*, otherwise 0 |
| $\varepsilon_{it}$ | = | Statistical error term for unexplained variation in observed energy consumption |
| $\delta_k$ | = | Average difference between treatment and control groups in the pre- and post time periods |

Functionally, this model compares the average usage of the treatment and control households while adjusting for other factors that may influence energy consumption (household characteristics, weather, etc). Models of this form produce unbiased estimates of the energy savings for a program with homes that were randomly assigned to the treatment group at the outset of the program. The critical coefficients are $\delta_1$, $\delta_2$, and $\delta_3$, which represent the average difference between the test and control groups before the test started (which should be statistically insignificant under randomization), the average difference between the before and after consumption levels (which captures macro effects), and lastly, the average difference between the test and control groups after the start of the program (which is the impact of the program), respectively. This model can also be used to estimate the impact of the program in different population segments by adding various interaction terms.[10]

It should be noted that billing analysis must be carefully performed to be effective. Evaluators must take care to look to current best practices for the most accurate methodologies. Furthermore, evaluators must address issues such as model misspecification, autocorrelation, serial correlation, heteroscedasticity, collinearity, and influential or missing data.

**Step 5: Perform Program Participation Survey**

The experimental design described so far uses regression analysis to determine the net energy savings resulting from a behavior-based program as measured by the average difference in energy consumption between the treatment and control groups. This measure avoids the need to estimate traditional net-

[10] Adding treatment by post by segment dummies will accomplish the former, while replacing the post variable with time period dummies will accomplish the latter.

to-gross effects such as free-ridership or spillover. However, additional analysis is required to obtain a true net energy impact.

Even though some increase in other program participation is attributable to the behavioral program, it is *important that these savings be reported separately in order to prevent double counting of benefits in* approved energy efficiency portfolios.

There are two types of other programs for which participation rates must be measured: individually tracked incentive programs such as mailed rebates, and so-called "upstream" programs providing subsidies for energy efficiency products, such as CFLs. In the case of the individually tracked programs, the utilities should simply continue to track the participation in these programs on an individual customer basis in both the test and control groups. In the case of "upstream" products, a customer survey must be performed to assess participation levels in both test and control groups. Participation levels for both groups are needed to properly attribute energy savings to the various, contributing energy efficiency programs as describe in Step 6 below.

**Step 6: Calculate Savings Attributable to Other Programs**

Savings from rebates or "upstream" subsidies must be distinguished to prevent double counting. Thus, the evaluator must first separate these savings from the total savings achieved through a behavioral program. Once the program participation levels are correctly established as described in Step 5 above, this becomes relatively straightforward.

For example, if 100 homes in the control group install efficient furnaces, and 120 homes in the treatment group do the same, the savings from the additional 20 furnaces installed can be easily identified and accounted for by reporting them as part of the behavioral program or as part of the furnace rebate program, but not both.

Figure 2 illustrates an example in which the reports lead to increased participation in a furnace rebate program run by the utility. The savings generated from installations that occur in both groups ("A" and "B" in the figure) cancel each other out and do not contribute to the overall savings measured as a difference in energy use between treatment and control groups. However, the incremental installations that occurred as a result of receiving the behavioral messaging ("C") do show up in the behavioral program's overall savings estimates. The total kWh or therms associated with the incremental installations can be estimated using the deemed savings for each type of installed measure. This process can be repeated across each type of measure offered by the utility.

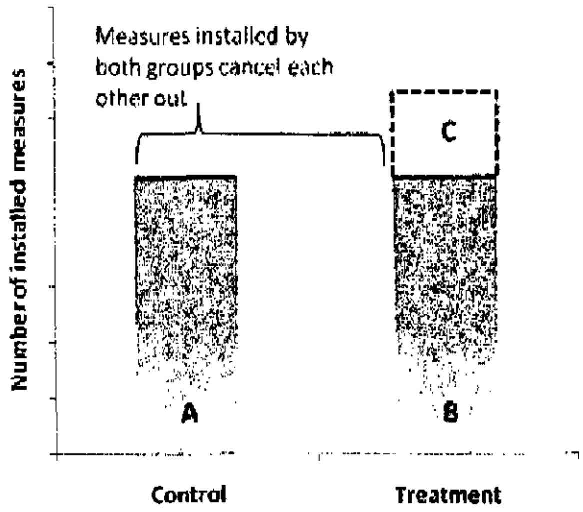


Figure 2. Double-counting mechanics

1. Savings measured as a difference between treatment and control groups do not include measures installed by both groups (areas A + B)
2. The potential for double-counting only exists when the Treatment group installs additional measures (area C)
3. Savings due to incremental measures are easily identified, allowing IOUs to account for them accordingly

A simple example is given in Table 2. The example assumes an energy efficiency portfolio consisting of programs actively promoting three installed measures in addition to the behavioral program: an Energy Star refrigerator incentive, a CFL incentive a program supporting installations of Home Area Networks and in-home displays. Participation rates in each of the three programs for both the treatment and control groups can be determined using the process described in Step 5, with the results listed in Columns 2 and 3. The difference in participation (Column 4) can then be multiplied by the deemed savings for each measure (Column 5) to arrive at the energy savings attributable to the refrigerator, CFL and HAN programs respectively.

Table 2: Example incremental savings calculations

| **Measure Type** (Column 1) | **Treatment group participation** (Column 2) | **Control group participation** (Column 3) | **Incremental participation** (Column 4) | **Deemed Savings** (Column 5) | **Double-counted savings to be accounted for** (Column 6) |
|---|---|---|---|---|---|
| ES Refrigerator | 1,100 units | 1,000 units | 100 units | 130 kWh | 13 MWh |
| CFL | 15,000 bulbs | 14,000 bulbs | 1,000 bulbs | 30 kWh | 30 MWh |
| HAN / IHD | 100 devices | 50 devices | 50 devices | 500 kWh | 25 MWh |
| **Subtotal** | | | | | **68 MWh** |

Note the because of the experimental approach used for program design and measurement, the potential for double-counting is limited only to the difference in participation between the test and control groups shown in Column 4, not the absolute level of participation shown in Column 2. The IOUs must decide how to report for incremental savings, in this case the 68 MWhs shown in Column 6.

**Conduct a Survey to Assess "Upstream" Participation Rates**

For energy efficiency programs that are not tracked at the individual customer level, estimates of participation rates must be constructed using other quantitative and qualitative data. Surveys are tools well suited to this task: they can be administered to sample populations from the treatment and control groups without polluting the results of the experiment. Specifically, these surveys should include questions that identify participation in the "upstream" programs of interest, such as CFLs. Because the goal of the survey is to estimate the difference in program participation rates between the treatment and control groups, the survey must be administered to both groups in order for the results to be useful.

Surveys are frequently used in the EM&V process for exactly this purpose; however, they must be carefully designed, administered, and analyzed in order to obtain reliable, unbiased results. For example, customers typically respond to these programs by making small, daily changes to their behavior and inaccurate or leading questions could lead to inconclusive results. A carefully designed survey administered to a substantial number of customers from both the test and control groups will work to avoid such inaccuracies.

## Reporting and Accounting of Savings

**Step 7: Reporting Savings to the Public Utilities Commission of Ohio ("the Commission")**

There are two ways to account for energy savings that were partly achieved as a result of behavioral messaging, and partly due to the financial incentive provided via another energy efficiency program, e.g. a rebate. The first is to subtract the incremental energy savings from the program providing the financial incentive. The second is to subtract the same savings from the total impact estimate of the behavioral program. In the example provided in Table 2 above, this would require reducing the savings claimed for the refrigerator, CFL, and HAN programs by 13 MWh, 30 MWh, and 25 MWh respectively be reported only ones, as part of the behavioral program, or the respective rebate programs, but not both.

Once the Commission has determined the preferred reporting methodology, savings should be attributed to the behavior-based program or other efficiency measure as appropriate. It is important to note that, although there is some overlap between behavior-based programs and other efficiency measures, behavioral programs that utilize experimental design have been shown to achieve greater aggregate energy savings than rebate programs. This is due to the typically high rates for customer engagement typically observed in behavior-based programs. As a result, the level of overlap with other efficiency programs is likely to be only a small portion of the total energy savings reported by a behavioral program.[11]

It is recommended to report program results to the Commission on a regular, annual basis beginning once the program has been deployed for 12 months. These interim results can be easily generated using standard statistical analysis software, and are critical to ensuring ongoing accurate measurement and accounting of savings and thereby ensure cost-effectiveness.

[11] In an analysis done with data from the Sacramento Municipal Utility District (SMUD) Home Energy Reporting program, OPOWER estimated that only 3% of total savings were attributable to financial incentives provided by other SMUD programs, while it was found that approximately 85% of treatment households changed their behavior as a result of the program.

## Appendix B:
## Sources Supporting the Measurement & Verification Protocol for Behavioral Programs

---

1. Allcott, Hunt and Sendhi Mullainathan, 2010. "Behavior and Energy Policy." *Science*. Vol. 327

   *Summary:* This article in *Science* advocates for the use of research and design processes to develop basic behavioral science into large-scale business and policy innovations in the area of energy efficiency.

2. Allcott, Hunt, 2010. "Social Norms and Energy Conservation." *Working Paper, Massachusetts Institute of Technology's Center for Energy and Environmental Policy Research.*

   *Summary:* This evaluation of a large-scale pilot program in Minnesota offers further evidence that non-price, social comparison messaging can substantially affect consumer behavior. This study also advocates for the randomized natural field experiment approach employed in this pilot program.

3. Ayres, Ian, et al., 2009. "Evidence From Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage." *NBER Working Paper.*

   *Summary:* This analysis of the SMUD and PSE field experiments concludes, "By providing feedback to customers on home electricity and natural gas usage with a focus on peer comparisons, utilities can reduce energy consumption at a low cost."

4. Banerjee, Abhijit and Esther Duflo, 2008. "The Experimental Approach to Developmental Economics." *NBER Working Paper.*

   *Summary:* This working paper discusses the strengths and limitations of randomized experiments as a tool in development economics research. It is concluded that the main benefit of randomized experiments is that they allow the estimation of parameters that otherwise would be outside the scope of evaluation. Although some of the concerns that are highlighted—including environmental dependence, compliance issues, randomization issues, equilibrium effects, heterogeneity in treatment effects, relationship with structural estimation, and relation to theory—are real, this study concludes they are not specific to experiments.

5. Davis, Lucas, 2008. "Durable goods and residential demand for energy and water: Evidence from a field trial." *RAND Journal of Economics.*

   *Summary:* This study advocates for random field trials as the ideal approach for observing and measuring household behavior.

6. Gillingham, Kenneth, 2006. "Energy Efficiency Policies: A Retrospective Examination." *Annual Review of Environment and Resources.*

   *Summary:* This literature review concludes in part that recent evidence suggests that techniques for measuring both energy savings and cost have improved markedly.

7. Goldstein, Noah, 2008. "A Room with a Viewpoint: Using Norm-Based Appeals to Motivate Conservation Behaviors in a Hotel Setting." *Journal of Consumer Research.*

   *Summary:* This study argues for the benefits of normative messaging in affecting behavior, and confirms the value of field experiments in the areas of behavioral economics and psychology.

8. Grinblatt, Mark, 2008. "Social Influence and Consumption: Evidence from the Automobile Purchases of Neighbors." *Review of Economics and Statistics.*

   *Summary:* Along with providing further evidence of the persuasive normative affect of neighbors on consumption, this article points to the limitations of observational studies.

9. Imbens, Guido, et al., 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature.*

   *Summary:* This article concludes that randomized experiments, though traditionally rare in economics, are extremely influential when they are conducted.

10. Ivanov, Chris, 2010. "Measurement and Verification Report of OPOWER Energy Efficiency Pilot Program." *Power System Engineering.*

    *Summary:* This third-party report evaluates, measures and verifies the one-year results of OPOWER's energy efficiency pilot program in Minnesota with Connexus Energy.

11. Klos, Mary, 2009. "Impact Evaluation of OPOWER SMUD Pilot Study." *Summit Blue Consulting, LLC.*

    *Summary:* This third-party report by Summit Blue Consulting evaluates, measures and verifies the results of OPOWER's Home Energy Reporting program in the Sacramento Municipal Utility District.

12. LaLonde, Robert, 1986. "Evaluating the Econometric Evaluations of Training Programs." *American Economic Association.*

    *Summary:* This paper takes the results of an employment and training program that was run as a randomized field experiment, and compares these results to the estimates that might have been produced by econometric evaluation procedures. This comparison shows that many of these econometric procedures fail to replicate the experimentally determined results, and suggests that researchers should be aware of the potential for specification errors in other nonexperimental evaluations.

13. Lee, David and Thomas Lemieux, 2009. "Regression Discontinuity Designs in Economics." *NBER Working Paper.*

    *Summary:* This evaluation compares RD design to randomized experiments, noting the aspects that make them similar yet still set randomized experiments apart in their ability to provide actionable results.

14. Levitt, Steven D, 2008. "Field experiments in economics: The past, the present, and the future." *European Economics Review.* Vol. 53 (1)

    *Summary:* This study explores the history and validity of conducting economic field experiments, advocating for their use when possible and articulating a number of ways in which they can be effectively deployed in the future.

15. Rubin, Donald, 2009. "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies." *Journal of Education Psychology.*

    *Summary:* This comparison of study design concludes that randomization has significant benefits and should be employed whenever possible.

16. Schultz, Wesley, et al., 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Journal of Psychological Science.*

    *Summary:* This study examines the results of a field experiment in which normative messages were used to promote energy conservation.

17. Shippee, Glenn, 1980. "Energy Consumption and Conservation Psychology." *Environmental Management.*

    *Summary:* This review of methodologies that have been employed in studies of conservation psychology—including the survey study, field experiment and laboratory investigation—concludes that several directions can be generalized that span across methodological approaches. Regarding field experiments in particular, the review notes that external validity of this type of experimental design has been high and delineates its importance in contributing to the development of feedback consumption research and the extent to which encouraging behavior changes can result in energy conservation.

## Appendix C:
## List of States and Utilities Using the OPOWER Behavioral Platform to Meet Efficiency Goals

| State | Partner(s) |
|---|---|
| Arizona | Arizona Public Service |
| California | Select IOUs |
| Colorado | Xcel |
| Florida | Progress, Gulf Power, FPL |
| Indiana | NIPSCO, Vectren, Indiana Michigan |
| Illinois | Commonwealth Edison |
| Maryland | BGE |
| Massachusetts | National Grid, NSTAR |
| Minnesota | Austin & Owatonna, Centerpoint, Connexus, Lake Country Power, MERC, Xcel Energy |
| New Jersey | New Jersey Natural Gas |
| New York | National Grid, Central Hudson |
| Ohio | AEP Ohio |
| Oregon | Energy Trust of Oregon |
| Washington | Puget Sound Electric |