December 12, 2011

Mr. Gary Martin
The Commission for Energy Regulation
The Exchange
Belgard Square North
Tallaght,
Dublin 24

Dear Mr. Martin,

Opower, Inc. ("Opower"), a behavioural energy efficiency and smart grid software company, would like to thank the Commission for Energy Regulation ("CER" or the "Commission") for the opportunity to comment on its consultation for Phase 1 of the Proposed National Rollout of Electricity and Gas Smart Metering (the "Rollout", "Programme", or "Plan") that was released November 15, 2011.

Opower works with over 60 utilities in the United States, including 8 of the 10 largest, and with First Utility in the United Kingdom, to deliver energy savings to residential households.[1] Similar to the proposed energy usage statements, Opower motivates customers to use less energy and save money on their monthly bills by providing customers with better information about their energy use and personalized energy savings advice. This year Opower will deliver its personalized behavioural efficiency programme to over 10 million residential customers through mail, email, websites, smart phone applications, and text messages.

Opower's results have been independently verified by twelve separate evaluations.[2] Exemplary of these evaluations, a recent article published in the *Journal of Public Economics* by Dr. Hunt Allcott of New York University ("NYU") evaluated Opower's 17 longest running deployments reaching 600,000 households.[3] Allcott concluded that Opower's program generated savings of 1.4 – 3.3%, with an average of 2%, across all geographies, and that these savings persist over time. To date, Opower has saved households more than 525 GWh or the equivalent of over €47 million in bill savings. Opower's programmes are also cost-effective, achieving these savings at an average of €0.02 per kWh saved. Opower also increases the rate of participation in other efficiency programmes, such as home insulation rebate programmes, by as much as 60%.

Recommendations

Opower supports CER's goal of creating a robust and secure market for energy efficiency and delivering smart meter benefits to households in Ireland. The Commission has used experimental design and *ex-post* measurement in its smart metering customer behavioural trials to assess the potential for enabling technologies such as energy usage statements (statements) and in-home displays (IHDs). This rigorous, results-based approach accurately identified statements as a valuable investment with net benefits for Irish households.

Along with this approach to the evaluation of customer behaviour programs, Opower supports the following components of the Rollout plan:

- **National rollout of smart meters**. As an enabling technology, smart meters offer the potential for households in Ireland to realize vast benefits from energy conservation and operational efficiencies.
- **Mandatory energy usage statements**. Opower's results support CER's decision to invest in energy usage statements as a means to enhance and lock in the benefits of smart meters.
- **Energy efficiency and peak load management are the first and second objectives.** There is vast potential for benefits to accrue to households and suppliers from increased efficiency and reduced peak load. Achieving these two objectives will help Ireland realize significant net benefits from investment in smart meters.
- **Access to half-hourly data**. Along with operational efficiencies and dynamic pricing, more granular data provides the opportunity for companies like Opower to provide more relevant information via energy usage statements. Access to half-hourly data also allows companies like Opower to provide unique products such as peak time alerts to help curb wasteful consumption during times when it is most expensive to generate. All of these benefits can accrue with full protection of household privacy.

Opower does not support the national rollout of IHDs:

- **IHDs should not be rolled out nationally.** Existing evidence of the impact from IHDs is limited by small sample sizes, short trial duration, variable savings rates, and selection bias resulting from opt-in design. There is also evidence that IHDs are cost ineffective relative to other forms of feedback, and may not provide considerable incremental benefits when provided in conjunction with other types of customer engagement. Further, CER's electricity trials raise questions about the persistence of IHD savings over time. Given these issues, the Commission should reconsider a national investment in IHDs.

Finally, Opower provides the following recommendations beyond the scope of this initial Programme:

- **Experimental design should be used for evaluation of savings from behavioural programmes.** Experimental design is consistent with the approach used by CER in its customer behavioural trials. This methodology is characterized by *ex-post* measurement of savings, billing analysis, and, where possible, opt-out design. Consistent with best practices in the United States for evaluating savings from behaviour-based energy efficiency programs, experimental design is the most rigorous approach to evaluating savings – which are the primary benefits that will accrue directly to households.
- **Energy saving goals and incentives should be implemented to encourage results-based investment by suppliers.** The Commission should also explore developing an energy saving goal and require results to be measured and verified, where possible using ex post, experimental design. Performance incentives for utilities that exceed goals should also be considered. This approach will encourage a results-based over a compliance-based response.

Responses to specific questions

Q1: Respondents are invited to comment on the proposed decision by the CER to proceed with the national rollout of electricity and gas smart metering as outlined in Section 2. Are you in favour of this proposal? Outline reasons for agreement or disagreement.

Opower supports both the national rollout of smart meters and mandatory energy usage statements for all residential households. The net benefits of smart meters are potentially much greater than the significant initial costs of €1 billion. Smart meters provide both utility and customer benefits. Among other customer benefits, such as dynamic pricing, smart meters provide much more granular energy usage data. While this data alone is not helpful to customers, it can be translated into insights about their energy usage through innovations like energy usage statements. By providing the building blocks necessary for other technologies and services to provide households with personalised information, feedback, and advice, smart meters can lead to significant energy savings. These savings in part justify the considerable initial cost. The importance of these benefits led the Environmental Defense Fund, an influential American advocacy organization, to include "Empower Consumers" as the first goal in its "Evaluation Framework for Smart Grid Deployment Plans."[4]

Opower has proven the benefits of energy usage statements in the United States. In a recently published article in the *Journal of Public Economics*, Dr. Hunt Allcott determines that 17 Opower deployments involving over 600,000 households achieved average savings of 1.4 – 3.3%.[5] These savings were generated without access to more granular data from smart meters, which Opower expects will increase both savings and peak reductions. Exemplary of this potential, two of the largest utilities in the US, Pacific Gas & Electric and Baltimore Gas & Electric, have contracted with Opower to engage their customers in the context of their full smart meter deployments.[6]

To take full advantage of these benefits, it is necessary that suppliers have access to this granular half-hourly data. Vendors like Opower and billing service providers operate as secure data processors. In this role, they are extensions of the utility, granted access to energy usage data in order to deliver specifically identified services – such as energy usage statements. With access to half-hourly data for both electric and gas meters, these companies can provide the most relevant information to customers. In turn, this information will encourage more savings and ensure that households take full advantage of smart meters.

Q2: Respondents are invited to comment on the proposed objectives of the National Smart Meter Programme outlined in Section 3. Are you in favour of the proposals? Outline reasons for agreement or disagreement.

Energy efficiency is the lowest cost resource. For example, in the US it costs €0 - €38 per MWh saved through increased efficiency, compared with a cost that ranges from €45 – 250 per additional MWh generated, depending on the energy source.[7] For every dollar invested in energy efficiency, CER will lower the overall cost of energy.

Reducing peak demand can have an even greater effect on the cost of energy. Generating facilities that are usually offline are operated to meet the few hours of peak demand. There is high marginal cost of building, operating and maintaining this peak supply for the relatively few MWhs that it actually generates each year. It is thus the most expensive to use. Given this high cost, reducing peak demand can lead to significant cost savings and increased system reliability, the benefits of which are passed along to customers.[8]

For these reasons, Opower supports CER's assertion that energy efficiency and peak load management should be the first and second objectives for the Rollout. Opower also recommends that CER implement an energy savings goal for suppliers to ensure they invest in results-based efficiency and peak reduction. A savings goal will help achieve these top two objectives. Providing incentives to suppliers to exceeding the goal will also ensure that they maximize their investment in efficiency rather than having a compliance response. CER should examine similar goal-based efficiency schemes in other countries, such as the Energy Efficiency Resource Standards in multiple U.S. states and the Energy Savings Scheme in New South Wales in Australia for examples of frameworks that may work in the Irish market.

Q3: Respondents are invited to comment on the proposed working assumptions, outlined in Section 4 relating to data ownership, display, and provision. Are you in favour of the proposals? Outline reasons for agreement or disagreement.

Opower supports the rollout of smart meters, supplier access to half-hourly data, and mandatory energy usage statements. We also recommend that experimental design be used for evaluation of savings from behavioural programs such as statements. Finally, Opower recommends that CER not deploy IHDs in conjunction with smart meters.

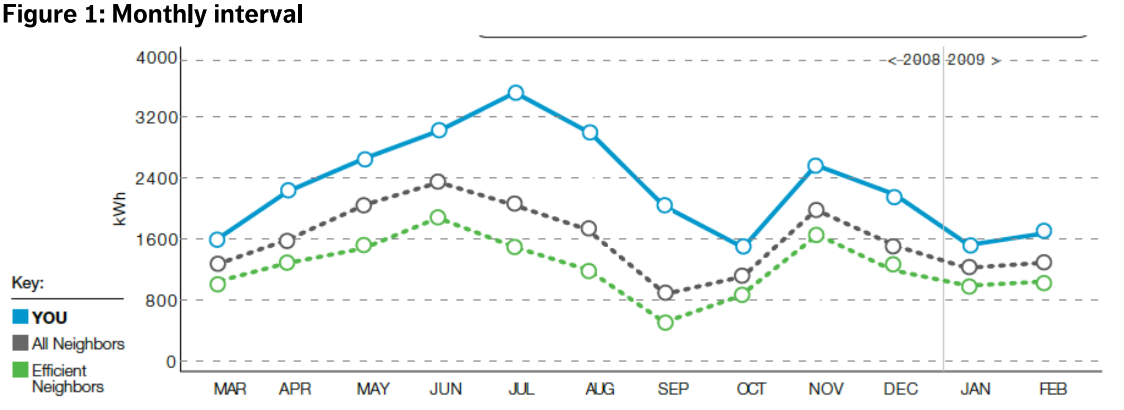
Support for the rollout, data access, and energy usage statements

Behavioural energy efficiency companies encourage energy-saving behaviour change through analysis of energy usage information and other publicly available data. As the amount of data available increases, this data analysis becomes more insightful and targeted. Greater insight leads to greater savings. These savings are some of the benefits that accrue to households from the implementation of smart meters.

More granular data leads to greater saving opportunities

To give a sense of the type of information that can be gleaned from more granular data intervals, and the corresponding insights that can be derived from that information, it is useful to look at several examples.

Figure 1: Monthly interval



With monthly data, customers can see how their energy usage compares to their neighbours and how it varies over seasons.

Monthly intervals allow energy information companies to show the aggregate amount of energy used by a given household in a given month. While it is possible to make some hypotheses about energy use by comparing months (i.e. if use goes up during very hot or very cold months, this can likely be attributed to heating or cooling), it is impossible to determine the actual amount of monthly usage attributable to heating and cooling, or to any distinct aspect of energy use. With access only to monthly data, a behavioural efficiency program largely focuses on:

- providing an accurate peer comparison (i.e. comparative consumption),
- showing progress month-to-month, and
- presenting personalised advice based largely on parcel data (e.g. size and date of construction of the home), demographic data, and seasonal variation.

The majority of Opower's programs in the US have operated with monthly data. Each of Opower's programs begun prior to the end of 2009 generated average savings of 1.4 - 3.3%, as verified by Dr. Hunt Allcott of NYU.[9]

Figure 2: Daily interval

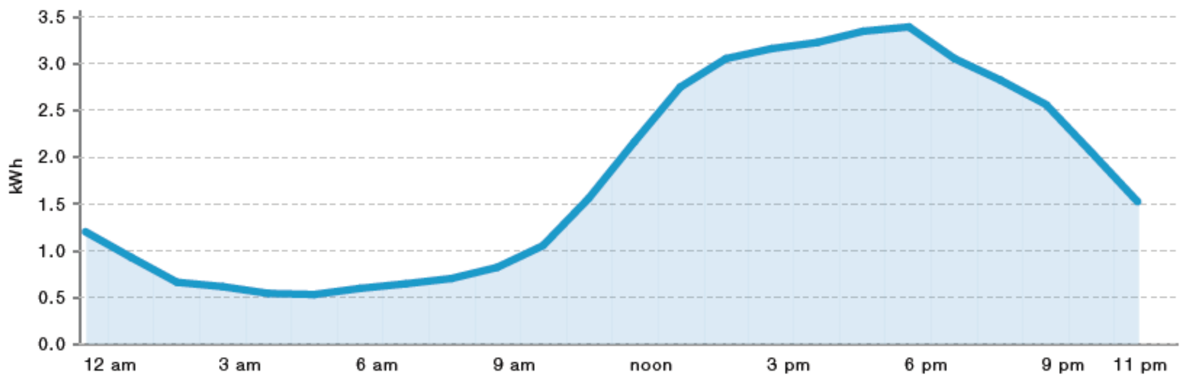
With daily meter reads, customers can receive high bill alerts and compare their weekday usage to their weekend usage.

At daily intervals, it is possible to disaggregate to some degree the amount of energy use dedicated to heating and cooling, and to offer consumers useful information about the ways they can save energy by modifying their behaviour (e.g. turning the thermostat down a few degrees when leaving the house for the day). Better insights can be developed as well, including by:

- building better profiles and tips by looking at weekday versus weekend usage,

- providing bill high bill alerts that let people know if they are on track for a particularly high bill at the end of the month; and,
- understanding how weather and behaviour are interacting.

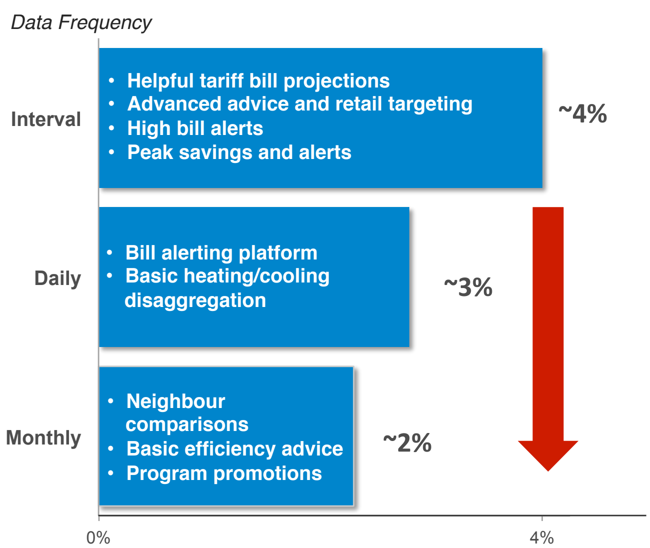
Figure 3: 30-minute interval



With 30-minute intervals, customers see how their energy usage changes over different periods of the day.

With smart meter data at a frequency of at least every 30-minute interval, behavioural efficiency programs can present customers with insights on the ways they use energy at different times of day, enabling them to adjust their usage and lessen strain on the grid during peak times. Interval data of at least this level of specificity is essential to offer truly dynamic pricing (i.e. critical peak pricing).[10] This level of data enables more accurate heating and cooling disaggregation and the construction of profiles identifying classes of usage.

Figure 6: Less data means fewer benefits



Behavioural measures can use this data to empower households

As illustrated in **Figure 6**, Opower estimates that Ireland could lose significant energy savings potential benefits from energy usage statements if it only allows suppliers to access monthly as opposed to sub-daily data. Sub-daily intervals allow companies like Opower to providehigh bill alerts and tariff bill projections, to name just two examples of capabilities that are not available with data at the daily or monthly level.

Energy usage statements are not the best delivery mechanism for time sensitive and dynamic messages like these. Rather, online web portals, text messages, phone calls, and smartphone applications are specifically designed for this purpose. By layering these technologies with energy usage statements, companies like Opower deliver these messages through the medium that is best able to empower customers to improve their energy usage behaviour.

Experimental design should be used to evaluate savings from all behavioural measures

Behavioural programs have a distinct advantage over installed measures: Impact from these programs can be evaluated *ex-post* via data analysis. As is consistent with The Brattle Group's "Measurement and Verification Principles for Behaviour-Based Efficiency Programs," rigorous M&V for these programs should include experimental design, statistically equivalent control and treatment groups, *ex-post* measurement of savings, and panel data regression billing analysis.[11] This approach is consistent with the Commission's methodology for the customer behavioural trials, which measured savings with 90% confidence.

Opower recommends that the Commission use experimental design to measure savings from energy usage statements, IHDs and other behavioural efficiency programs. Since savings from these programmes are mainly a function of discretionary changes in behaviours—which are unpredictable—it is more appropriate to measure savings after they occur rather than to assume or deem a future rate of savings based on past performance. Such an approach is used widely and cost effectively in the United States to measure the savings from behavioural measures.

For a more thorough discussion of experimental design and ex-post measurement, please refer to **Appendix A: M&V Protocol for Behavioural Energy Efficiency Programs**.

CER should not deploy IHDs in conjunction with smart meters

The results from Ireland's trials indicate that IHDs do provide a marginal benefit to energy usage statements. But, the trial also demonstrated a statistically significant decrease in the savings from IHDs in the electricity trial in the last six months compared to the first six. With 90% confidence, CER's electricity trials resulted in 2.6% savings in the first six months and 2.8% savings in the last six months when households were exposed to energy usage statements only. When electricity monitors were included, households saved 4.0% in the first six months and 2.4% in the last six months. These results provide evidence that savings from IHDs may decay over time.[12] With similar overall results, the gas trials resulted in 2.2 – 2.8% savings when households were exposed to energy usage statements in isolation. When households received IHDs as well as energy usage statements, savings were 2.9 – 3.6%.[13]

Lack of evidence that savings persist

The electricity results indicate that the persistence of IHD savings may diminish over time. These results are consistent with other IHD trials. One of the largest and longest running IHD trials included 30,000 households in Ontario over two years. In this trial, only 29% of households continued to use the IHDs after the two years of the trial. Further, only 56% used it even one month into the trial. [14]

There is also generally a lack of evidence regarding the impact of IHDs over time. In a global meta-analysis of IHD trials in Victoria, Australia, consultant Accenture evaluated 60 IHD trials that demonstrated energy savings. The longest running of these trials were 2 years in length, and only 3 of the 60 trials were this long. 45 of the 60 trials ran less than a year.[15] Given the short duration of the available studies, there is simply not enough evidence of persistent savings to justify the considerable investment in IHDs. Indeed, as discussed above, there is some evidence that the effectiveness of IHDs actually diminishes over time.

Opt-in design of IHD trials complicates accurate M&V for savings

IHD trials tend to be opt-in, which can lead to selection bias when it comes to measuring results. Although there are a variety of statistical techniques one can use to match participants with non-participants based on observable characteristics – such as housing, demographic, and census data – none of these methods address differences in unobservable characteristics like attitudes and beliefs. 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 itself of opting-in signals a difference from those who did not opt in. In the world of surveys, this is known as survey responder (or selection) bias.[16] Given this selection bias, it is unlikely that energy savings anticipated based on opt-in studies of IHDs will be realized in a nation-wide rollout of IHDs.

Opt-out design, by contrast, eliminates selection bias and other biases, and is essential for establishing rigorous experimental design, which is key to accurate and reliable M&V of program savings. Specifically, opt-out design allows for the creation of treatment and control groups that are statistically equivalent so that the effect of a program on the treatment group's energy usage can be measured with statistical confidence.

Small sample sizes diminishes the statistical confidence in results

Of the 60 trials evaluated in the same meta-analysis in Victoria, 45 had fewer than 1,000 participants and, of these forty-five, seventeen had fewer than 100. The most thorough of these trials is discussed in Case Study 1.1 on page 24 of the study. This Hydro One study concluded that IHDs in isolation reduce energy usage by 6.5% without referencing sample size, statistical confidence, or duration of this study. According to Appendix A1, this Hydro One pilot program had a total of 81 participants and was less than one year in duration.[17] This is neither a large enough sample, nor a long enough study to convincingly make the case for the persistent value of IHDs.

Further research is needed on IHDs prior to a full-scale rollout

There is insufficient evidence to accurately and rigorously assume the savings impact of IHDs over time. CER performed rigorous randomized controlled trials that demonstrated marginal incremental benefits of IHDs beyond energy usage statements, but these savings did not demonstrate persistence over time. Evidence from other IHD trials confirms this finding.

Given this lack of convincing evidence, and the considerable expense of IHD deployment, Opower recommends that CER evaluate IHDs further using randomized controlled trials before deploying them nationwide.

Conclusion

Opower appreciates the opportunity to comment on this important investment in Ireland's energy infrastructure. CER has largely developed a deployment plan consistent with best available evidence regarding smart meters. Opower broadly supports the Commission's nationwide rollout of smart meters, mandatory deployment of energy usage statements, and decision to allow access to half-hourly data with strict privacy protections in place. Opower also recommends the Commission implement experimental design to measure savings from behavioural programmes, and establish an energy savings goal -- coupled with incentives for suppliers who exceed it -- to encourage results- rather than compliance-based responses from suppliers. Finally, we hope the Commission decides to re-evaluate a nationwide rollout of IHDs given the lack of consistent evidence.

Sincerely,

Jim Kapsis
Director of Market Development and Strategy
Opower | jim.kapsis@opower.com

Appendix A: M&V Protocol for Behavioral Energy Efficiency Programs

Description of Measure

Behavioral 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 programs 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, as well as the best practices adopted by The Brattle Group's "Measurement and Verification Principles for Behavioral Efficiency Programs."[2]

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

---

[1] See the following: (i) Allcott, Hunt, October 2011. "Social Norms and Energy Conservation." *Journal of Public Economics*, Vol 95(9-10), pp. 1082 – 1095;
(ii) Davis, Matt, May 2011. "Behavior and Energy Savings: Evidence from a Series of Experimental Interventions." *Environmental Defense Fund*;
(ii) Davis, Matt, May 2011. "Behavior and Energy Savings: Evidence from a Series of Experimental Interventions." (iii) Cooney, Kevin, February 2011. "Evaluation Report: Opower SMUD Pilot Year 2." *Navigant Consulting*; (iv) Todd,
Environmental Defense Fund;
(iii) Cooney, Kevin, February 2011. "Evaluation Report: Opower SMUD Pilot Year 2." *Navigant Consulting*; (iv) Todd, Annika, Steven Schiller, and Charles Goldman, October 2011.[1] "Analysis of PSE's Pilot Energy Conservation Project: "Home Energy Reports." *Lawrence Berkeley National Laboratory;*
(v) Ivanov, Chris, July 2010. "Measurement and Verification Report of Opower Energy Efficiency Pilot Program." *Power System Engineering*;
(vi) Macke, Rich, June 2010. "Measurement and Verification Report of Lake Country's Opower Energy Efficiency Pilot Program." *Power System Engineering*;
(vii) Allcott, Hunt and Sendhil Mullainathan, March 2010. "Behavior and Energy Policy." *Science*, Vol. 327; (viii) Allcott, Hunt, February 2010. "Social Norms and Energy Conservation." *Working Paper, Massachusetts Institute of Technology's Center for Energy and Environmental Policy Research*;
(ix) Ayres, Ian, et al., September 2009. "Evidence From Two Large Field Experiments That Peer Comparison Feedback Can Reduce Residential Energy Usage." *NBER Working Paper*;
(x) Klos, Mary, September 2009. "Impact Evaluation of Opower SMUD Pilot Study." *Summit Blue Consulting, LLC*;
(xi) October 2010. "Puget Sound Energy's Home Energy Reports Program." *KEMA*;
(xii) Dougherty, Anne, June 2011. "Massachusetts Cross-Cutting Behavioral Program Evaluation," *Navigant Consulting and Opinion Dynamics*

[2] Sergici, Sanem, and Ahmad Faruqui, May 2011, "Measurement and Verification Principles for Behavioral Efficiency Programs," *The Brattle Group*, available here: http://www.brattle.com/_documents/UploadLibrary/Upload955.pdf

- 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
- 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. Behavioral 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, phones, and/or a display in the consumer's home.

Measure Life

While there is evidence that behavioral program results persist, behavioral 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 behavioral 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

[3] *International Performance Measurement & Verification Protocol (IPMVP); Concepts and Options for Determining Energy and Water Savings: Volume 1.* 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

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 behavioral 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 behavioral 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 behavioral 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

[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.

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.

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

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).

Phase 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.[8] 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 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*
α_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
ε_{it}	=	Statistical error term for unexplained variation in observed energy consumption

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

δ_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 d_1, d_2, and d_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.[9]

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 behavioral 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-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

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

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 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.

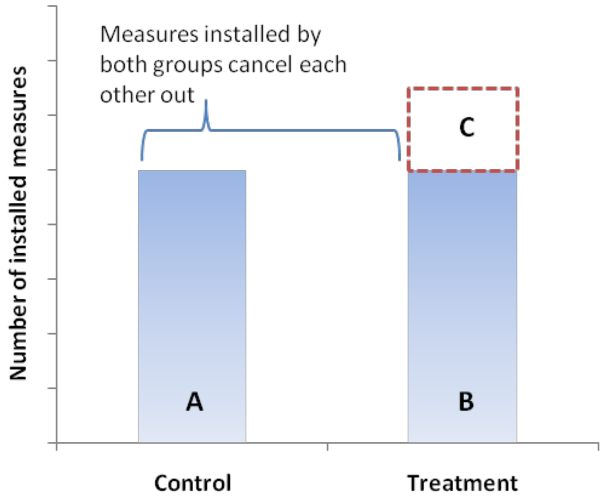


Figure 2: Double-counting mechanics

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 1: 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 that 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.

Phase 3: Reporting and Accounting of Savings

STEP 7: REPORTING SAVINGS

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 1 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 regulator has determined the preferred reporting methodology, savings should be attributed to the behavioral program or other efficiency measure as appropriate. It is important to note that, although there is some overlap between behavioral 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 behavioral 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.[10]

It is recommended to report program results 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.

[10] 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.

Endnotes

---

[1] See UK expansion press release here: http://opower.com/company/news-press/press_releases/28

[2] See the following: (i) Davis, Matt, May 2011. "Behavior and Energy Savings: Evidence from a Series of
See the following: (i) Davis, Matt, May 2011. "Behavior and Energy Savings: Evidence from a Series of
Experimental Interventions." *Environmental Defense Fund*;

(ii) Cooney, Kevin, February 2011. "Evaluation Report: OPOWER SMUD Pilot Year 2." *Navigant Consulting*;

(iii) Todd, Annika, Steven Schiller, and Charles Goldman, October 2011.[2] "Analysis of PSE's Pilot Energy Conservation Project: "Home Energy Reports." *Lawrence Berkeley National Laboratory;*

(iv) Ivanov, Chris, July 2010. "Measurement and Verification Report of OPOWER Energy Efficiency Pilot Program." *Power System Engineering*;

(v) Macke, Rich, June 2010. "Measurement and Verification Report of Lake Country's OPOWER Energy Efficiency Pilot Program." *Power System Engineering*;

(vi) Allcott, Hunt and Sendhil Mullainathan, March 2010. "Behavior and Energy Policy." *Science*, Vol. 327;

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

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

(ix) Klos, Mary, September 2009. "Impact Evaluation of OPOWER SMUD Pilot Study." *Summit Blue Consulting, LLC*;

(x) October 2010. "Puget Sound Energy's Home Energy Reports Program." *KEMA*

[3] Allcott, Hunt, October 2011, "Social Norms and Energy Conservation," *Journal of Public Economics*, available here: http://opower.com/uploads/library/file/1/allcott_2011_jpubec_-_social_norms_and_energy_conservation.pdf

[4] Herter, Karen, June 2011, "Evaluation Framework for Smart Grid Deployment Plans," *Environmental Defense Fund*, available here: http://www.smartgridinformation.info/pdf/4570_doc_1.pdf

[5] Allcott, Hunt, October 2011, "Social Norms and Energy Conservation," *Journal of Public Economics*, available here: http://opower.com/uploads/library/file/1/allcott_2011_jpubec_-_social_norms_and_energy_conservation.pdf

[6] See, e.g., St. John, Jeff, 7 April 2011, "Opower Lands Utilities PG&E, BG&E," *GigaOM*, available here: http://gigaom.com/cleantech/opower-lands-utilities-pge-bge/

[7] See, "Levelized Cost of Energy analysis – Version 3.0," February 2009, *Lazard*, available here: http://blog.cleanenergy.org/files/2009/04/lazard2009_levelizedcostofenergy.pdf

[8] For a discussion of this concept, see, e.g., Nadel, Steve, et al., November 200, "Using Targeted Energy Efficiency Programs to Reduce Peak Electrical Demand and Address Electric System Reliability Problems," *ACEEE*, available here: http://www.aceee.org/sites/default/files/publications/researchreports/U008.pdf

[9] Allcott, Hunt, October 2011, "Social Norms and Energy Conservation," *Journal of Public Economics*, available here: http://opower.com/uploads/library/file/1/allcott_2011_jpubec_-_social_norms_and_energy_conservation.pdf

---

[10] For a discussion of dynamic pricing, see, e.g.: Faruqui, Ahmad, and Sanem Sergici, 14 June 2011 "Dynamic Pricing: Past, Present, and Future," *The Brattle Group*, available here: http://www.brattle.com/_documents/UploadLibrary/Upload956.pdf

[11] Sergici, Sanem and Ahmad Faruqui, May 2011, "Measurement and Verification Principles for Behavior-Based Efficiency Programs," *The Brattle Group*, available here: http://www.brattle.com/_documents/UploadLibrary/Upload955.pdf

[12] CER, 16 May 2011, "Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report," p. 69, available here: http://www.cer.ie/en/information-centre-reports-and-publications.aspx?article=5dd4bce4-ebd8-475e-b78d-da24e4ff7339

[13] CER, 11 October 2011, "Gas Smart Metering Customer Behaviour Trials (CBT) Findings Report," p. 46, available here: http://www.cer.ie/en/information-centre-reports-and-publications.aspx?article=5dd4bce4-ebd8-475e-b78d-da24e4ff7339

[14] Darby, Sarah, "Literature review for the Energy Demand Research Project," *Environmental Change Institute of Oxford University*, p. 13, available here: http://www.ofgem.gov.uk/Sustainability/EDRP/Documents1/SD%20Ofgem%20literature%20review%20FINAL%20081210.pdf

[15] "In Home Displays in the Victorian Energy Saver Incentive Scheme," 22 November 2011, *Accenture*, available here: https://www.veet.vic.gov.au/Public/Public.aspx?id=Consultations

[16] This responder bias (also known as "survey responder bias") has made opt-out programs preferred 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.

[17] Id., p. 95. This assumes that the pilot referenced in this case study is recognized as "(IHD Only Pilot)" in the table. If it is one of the other pilots, the sample size could be 153, 382, or 30,000.