A Randomized Experiment in Direct Load Control Using Smart Meter Data

Jessei Kanagarajan, Ontario Power Authority, Toronto, ON, Canada
Dries Berghman, Freeman, Sullivan & Co., San Francisco, CA, USA

ABSTRACT

As smart meter data has become increasingly available, large randomized controlled trials (RCTs) have become an affordable and highly accurate alternative to logger data for direct load control (DLC) impact estimation.

The RCT method was applied to a provincial residential and small commercial load control program. This program involves the installation of programmable communicating thermostats (PCTs) and/or direct load control switches (switches) in households and small businesses with central air conditioning (CAC) in the service territory of more than forty local distribution companies (LDCs) across the province. In 2012, an RCT using a sample of 28,000 customers was used to estimate impacts for the program's 180,000 customers. Past evaluation efforts were only able to include a few hundred customers in two different LDCs, which only represent a small portion of the customer base.

During a load control event, a portion of the participant population received the treatment (in this case, load control) while the remainder of the participant population served as the control group and provides a reference load. The results of this direct comparison will be used to inform the choice of future load control strategies to improve program load impacts and cost effectiveness, while maintaining an acceptable level of comfort for program participants.

While the RCT method was able to yield accurate program impacts within a few hours of receiving smart meter data, evaluators did encounter a few challenges, which are described in this report.

Introduction

The Ontario Power Authority (OPA) was established in 2005 and is responsible for coordinating province-wide conservation efforts in Ontario, Canada. Since 2005 the OPA has sponsored conservation and demand management programs, which are generally delivered province-wide by over 75 Location Distribution Companies (LDCs). OPA sponsored the successful *peaksaver*® program from 2007 to 2011, which focused on demand response in southern Ontario. Launched in July 2012, *peaksaver* PLUS™, which incorporates elements of energy efficiency and demand response, is a flagship initiative of OPA's Consumer Program and has province-wide coverage.

The *peaksaver* PLUS program builds on the success of OPA's previous *peaksaver*® residential and small commercial program by allowing consumers to participate in demand response with an expanded list of eligible high demand end uses. End uses eligible for *peaksaver* PLUS are central air conditioners, electric water heaters and pool pumps, which are controlled by one-way communicating PCTs or switches. Participants of *peaksaver* PLUS also receive a free in-home display (IHD) to monitor their electricity consumption and cost in near real time. As of April 2013, over 180,000 customers across Ontario have enrolled in the *peaksaver* PLUS program.

To date, 99% of residential customers have installed smart meters in Ontario. This represents the largest smart meter fleet in North America. Data for over 90% of residential customers is sent to Ontario's Meter Data Management and Repository (MDM/R), a central smart meter data repository residing at the Independent Electricity Systems Operator (IESO). The high penetration of smart meters in Ontario allows us to leverage interval meter data for energy evaluation purposes. The use of interval meter data was incorporated into the planning of the 2012 *peaksaver* PLUS program evaluation. This paper focuses on the demand response impact evaluation component of the *peaksaver* PLUS evaluation; it describes the methods used to complete the evaluation, as well as the evaluation's results and lessons learned.

Methods

This section describes the methods used to implement the evaluation. First, it describes the theoretical motivation for using a randomized controlled trial evaluation design. Next, it describes the specific design used to complete this evaluation. Finally, it details the method used to calculate ex-post and ex-ante load impact estimates for the program.

Theoretical Background

The purpose of this evaluation is to calculate ex-post and ex-ante load impact estimates for *peaksaver* PLUS customers. These estimates must be produced for the entire province of Ontario and specific Ontario Independent Electricity Systems Operator (IESO) zones (electric grid transmission areas). As with many other evaluation efforts, the principal challenge when evaluating this program is to determine what the load *would have been* – in other words, what the reference load is – if the customer did not participate in a load control event. Once the evaluators know what the load would have been, they can simply subtract the actual load during the event from the reference load to develop a load impact estimate.

In past evaluations of the *peaksaver* program, evaluators relied on a within-subjects research design. This design uses the customer's own load when there is not an event to develop a reference load. A common within-subjects strategy is to use whole-house smart meter data or AC end-use loggers to develop a regression model. These models sometimes incorporate weather during (or immediately preceding) the event, as well as autoregressive terms (the whole-house or AC load preceding the event). The dependent variable is either whole-house or AC load.

However, in the context of an AC load control program, a within-subjects approach may not yield reliable load impact estimates.

First, customers may not have experienced enough hot non-event days to accurately create a reference load. AC load control programs are typically used on hot days, because this is when most customers use AC and when they are most effective. In Ontario, which has a relatively mild climate compared to the rest of North America, most very hot days are also AC load control days. In the absence of hot non-event days, the evaluator's risk extrapolating an inaccurate estimate from data that is not representative of actual event day conditions and this can affect the performance results.

A second problem facing a traditional within-subjects approach is that it is easy to confound event impacts with other factors, particularly when the evaluation relies on individual customer regressions. This is true whether or not the regression uses AC end use data or whole-house data. The number of potential confounding factors is endless. For example, a customer could be on vacation during a heat wave, making it seem like they provided very large load impacts. On the other hand, load impact estimates would seem very small for a customer who is newly unemployed and is spending more time at home.

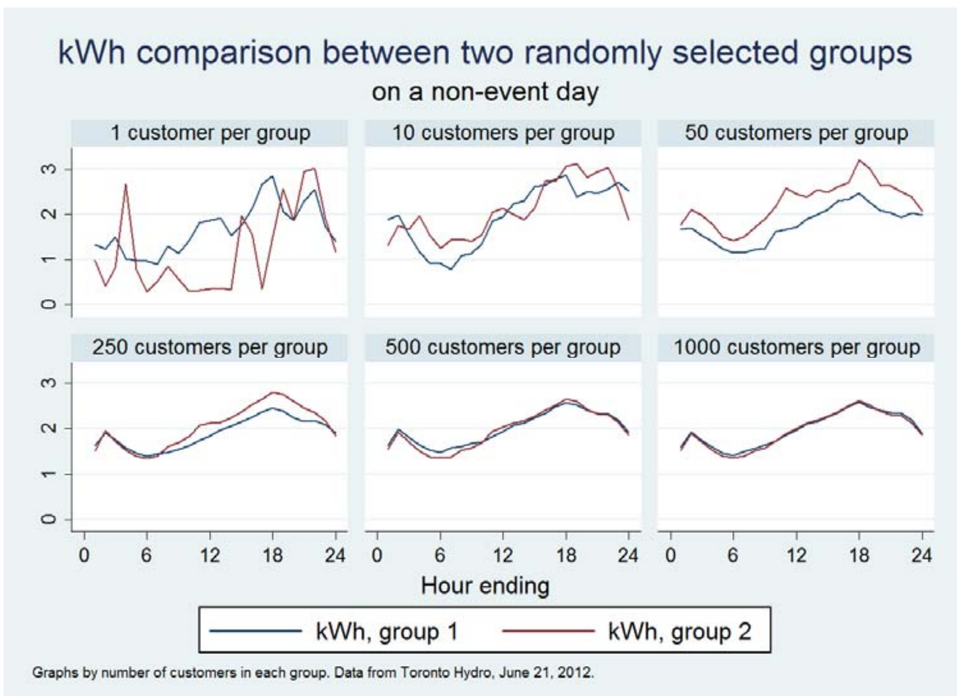
Finally, if an evaluation relies on AC end use data, the evaluation's sample size will likely be relatively small and not sufficiently representative of the general population. AC end use logging requires that a technician go out to a customer's property to install a logger, and then return once more to retrieve it at the end of the season. While the data collected from these loggers is very useful, because it directly isolates AC load, it is very difficult to apply this evaluation technique on samples that are large enough to draw statistically significant conclusions (i.e., more than 1,000 customers). Previous evaluations of the *peaksaver* program relied on a within-subjects approach using AC end use loggers.

Instead of relying on a within-subjects approach, our evaluation uses a between-subjects approach with whole-house data. The between-subjects approach uses data from customers that do not participate in the event to generate a reference load. Customers that participate in the event are part of the treatment group, whereas customers that do not participate in the event are part of the control group.

The main benefit of the between-subjects approach is that if the treatment and control groups are large enough, and randomly selected, it generates nearly perfect reference loads using whole-house smart meter data. Because a customer does not have to furnish a reference load for their own event, confounding factors are not an issue. Furthermore, because groups are large and randomly selected, unobservable differences between them can be ignored. When averaging the load for the control and treatment groups on an event day, any differences should be due to the event. If there is no event, the average treatment and control group load should be identical.

Figure 1 compares the average usage for randomly selected groups of various sizes on a non-event day. Because there is no event, their loads are expected to be identical. Indeed, as the group sizes increase, the average load for the two groups begins to approximate one another. When each group contains 500 customers, the differences are very small, and once each group contains 1,000 customers, their average load is basically identical. This indicates that one group could furnish an excellent reference load for the other group.

Figure 1: Effect of Increasing Group Size



An additional benefit of using whole-house smart meter data instead of AC end use loggers, is that it is cheaper. The marginal cost of adding additional customers to the evaluation sample is essentially negligible. For the 2012 *peaksaver* PLUS evaluation, six utilities were involved, and results were produced for five utilities. In total, approximately 26,000 customers were part of the evaluation sample, whereas less than 500 were included in previous years. The sample of 26,000 customers is made possible by the fact that 99% of residential customers in Ontario have smart meters. The 2012 evaluation is more representative of actual program participants than ever before.

A final benefit of using the between-subjects approach using smart meter data – as presented in this paper – is that one control group can be used for multiple treatments. The same control group can be used to furnish a reference load. The 2012 *peaksaver* PLUS evaluation exploited this possibility and tested four different AC cycling treatments. Further details are provided in the Experimental Design section.

Experimental Design

The evaluation involved six LDCs in Ontario. LDCs are electric utilities that provide distribution services. Customers at these LDCs were spread throughout the province of Ontario, although the majority was located in the greater Toronto metropolitan area. The six LDCs participating in the evaluation were selected to take part because they had large *peaksaver* PLUS populations and because they had the operational capability to provide smart meter data on relatively short notice.

In total, nearly 28,000 residential and small business customers were included in the evaluation. Customers were randomly selected and randomly assigned to groups of 1,000 customers. Depending on the total number of customers participating in the evaluation at each LDC, the number of groups per LDC varied from three to six. At each LDC, at least one group functioned as a control group, whereas the remaining groups experienced various cycling strategies (treatments) on event days. Cycling strategies tested included:

- 50% Simple Cycle – the compressor is allowed to run normally for 15 minutes and is forced to turn off for 15 minutes. This is the default strategy and the main focus of the evaluation.
- 60% Simple Cycle – the compressor is allowed to run normally for 12 minutes and is forced to turn off for 18 minutes.
- 50% TrueCycle I – an algorithm determines when to turn off the compressor and attempts to turn it off when it would have been running.
- 2°C ramping – the temperature setting on the customer's thermostat is increased by 2°C during the event.

As noted above, the main focus of the evaluation was to determine load impacts from 50% cycling, which was the default cycling strategy. The other cycling strategies were included to determine whether they could provide larger load impacts without sacrificing customer comfort.

A key feature of the experimental design was that load control events are called on a rotating schedule, meaning that a customer may be in the control group for one event, but in one of the treatment groups for another. This avoids using the same customers to create a reference load for every single event and makes sure that customers are not overburdened by more intrusive cycling strategies.

Table 1 shows a simplified example of an event-calling schedule used in the evaluation. It shows group assignments for four evaluation groups on four event days. Notice that on each event day, one group serves as a control group to generate reference loads for the various cycling strategies. In addition, cycling assignments rotate among groups, so that no one group is repeatedly exposed to the same cycling assignment.

Table 1: Simplified Event Schedule Example

Event Day #	Group #			
	1	2	3	4
1	Control	50% Simple	60% Simple	50% True Cycle I
2	50% True Cycle I	Control	50% Simple	60% Simple
3	60% Simple	50% True Cycle I	Control	50% Simple
4	50% Simple	60% Simple	50% True Cycle I	Control

Table 2 shows an overview of the events that were called in 2012. A total of seven load control events were called for the evaluation sample in 2012. Three of these events occurred in June 2012; three occurred in July 2012; and one occurred in August 2012. Two of these events were province-wide events and involved all customers in the province enrolled in *peaksaver* PLUS. Five events were EM&V test events and only involved customers in the evaluation sample. These EM&V test events were called to gather more data for the evaluation. The province-wide events were both four hours long, whereas the EM&V test events were anywhere from two to four hours long. Province-wide events tended to be slightly hotter than test events because they are called when needed for load relief.

Table 2: Event Schedule Overview

Type of Event	Event Date	Event Hours	Average Temperature During Event (°C)
Whole Province	6/20/2012	2 PM-6 PM	33
	7/6/2012	2 PM-6 PM	32
	Average	n/a	32
EM&V Test	6/21/2012	3 PM-5 PM	29
	6/29/2012	1 PM-5 PM	30
	7/4/2012	4 PM-6 PM	31
	7/17/2012	3 PM-5 PM	33
	8/31/2012	3 PM-5 PM	30
	Average	n/a	31
Total	**Overall Average**	n/a	31

Calculating Load Impacts

Using the between-subjects method described in this paper, the main inputs to calculate load impacts were the smart meter data for each customer participating in the evaluation, the customer's group assignment, and the group's scenario assignment (whether the group was a treatment group or a control group). Smart meter data was provided by each of the six LDCs participating in the evaluation, while group assignments and scenario assignments were determined by the evaluators. With these inputs, we constructed a database where each observation consisted of a single customer's ID number, their LDC, the date, their whole house load for each hour for that date, and their treatment assignment. Once this database was constructed, we simply collapsed it by date and treatment assignment to create an average load profile for each event date and treatment group. These load profiles form the basis for ex-post and ex-ante load impact calculations.

Ex-post impacts are average load impacts for the actual event days that were called in 2012, whereas ex-ante impacts are average load impacts for a set of idealized weather conditions. Ex-ante impacts are of most interest to stakeholders because they represent an estimate of the program's future load impacts and are used for planning purposes.

Ex-post load impacts for each event day and treatment type were calculated by subtracting the average load during the event window for the control group from the average load during the event window for each of the treatment groups. No additional manipulation was required to create an unbiased estimate of the ex-post impacts.

Creating an ex-ante load impact estimate is slightly more complicated because it involves a forecasting element. The Ontario Power Authority maintains a set of ex-ante weather conditions that are designed to represent 1-in-2 (average) conditions and 1-in-10 (extreme) conditions. The ex-ante weather conditions consist of a temperature value for each hour on the average week day and the monthly peak day for each summer month. The ex-ante load impact estimates were created by applying the coefficients from a simple regression of load impacts on temperature conditions preceding the actual 2012 events to the ex-ante weather conditions.

Ex-post and ex-ante load impact estimates were calculated for various geographic units. First, we calculated impacts for each participating LDC. These impacts were not used for reporting purposes, but were shared with each LDC participating in the study. The second set of impacts was calculated at the province-wide level. This was done by re-weighting customers in the evaluation sample to be more representative of the province-wide distribution of *peaksaver* PLUS customers. The final set of impacts was calculated at the settlement zone level. Settlement zones are a geographic unit representing transmission capacity areas and are used for planning purposes. As with the province-wide results, settlement zone results consist of a re-weighting of customers in the evaluation sample to be more representative of the geographic distribution of *peaksaver* PLUS customers in each zone.

Results

This section describes the results of the 2012 evaluation. It first describes the ex-post and ex-ante estimates calculated as part of the evaluation. The section concludes by describing lessons learned from the evaluation effort and future evaluation plans.

Ex-Post Results

Table 3 shows ex-post load impacts for each event day called in 2012.

Table 3: Ex-Post Load Impacts (All Customers)

Type of Event	Event Date	Event Hours	Average Reference Load (kW)	Average Event Impact (kW)	Average Percent Impact (%)	Average Temperature During Event (°C)
Whole Province	6/20/2012	2 PM-6 PM	2.51	0.42	17	33
	7/6/2012	2 PM-6 PM	2.51	0.43	17	32
	Average	n/a	2.51	0.42	17	32
EM&V Test	6/21/2012	3 PM-5 PM	2.49	0.39	16	29
	6/29/2012	1 PM-5 PM	2.34	0.40	17	30
	7/4/2012	4 PM-6 PM	2.57	0.47	18	31
	7/17/2012	3 PM-5 PM	2.66	0.44	17	33
	8/31/2012	3 PM-5 PM	1.97	0.28	14	30
	Average	n/a	2.40	0.40	16	31
Total	**Overall Average**	n/a	2.43	0.41	17	31

The average load impact for a province-wide event was 0.42 kW, whereas the average load impact for a test event was 0.40 kW. The average event impact, regardless of the event type, was 0.41 kW. These load impacts represent a decrease of approximately 17% compared to the whole building reference load. Load impacts were fairly consistent across the seven events. The one exception occurred on August 31; on this day, the impact was just 0.28 kW, or 14% of the whole building reference load. This is likely because August 31 was an isolated hot day, whereas the other events were called during a sustained heat wave during the first half of the summer. Load impacts tend to be larger if heat has been able to build up for several days.

Ex-Ante Results

Table 4 shows the average ex-ante load impacts by cycling strategy for an event lasting from 2 to 6 PM generated using data from the 2012 evaluation.

Table 4: Ex-Ante Load Impacts by Cycling Strategy (All Customers)

Cycling Strategy	Weather Year	Estimated Peak Impact (kW)				
		May Peak	June Peak	July Peak	August Peak	September Peak
50% Simple Cycle	1-in-2	0.31	0.34	0.37	0.38	0.32
	1-in-10	0.34	0.40	0.44	**0.47**	0.34
60% Simple Cycle	1-in-2	0.31	0.34	0.37	0.38	0.32
	1-in-10	0.34	0.44	0.51	**0.58**	0.34

Impacts for the 1-in-10 weather year are always greater than those for the 1-in-2 weather year, and impacts for the August peak are higher than those for the other months. This is because the potential load impacts from AC cycling are directly impacted by weather conditions: the program is able to deliver larger load reductions when it is hotter outside. 1-in-10 weather conditions are only expected to occur once every 10 years and are thus hotter than 1-in-2 weather conditions.

Likewise, impacts from 60% cycling are greater than those for 50% cycling. Again, this is unsurprising, because for every hour of program activation, 60% cycling curtails the AC unit six minutes longer than under 50% cycling. However, the result in Table 4 is important because it provides empirical proof that 60% cycling actually *does* provide larger load impacts than 50% cycling. The OPA is considering the adoption of 60% cycling as its default cycling strategy during future *peaksaver*PLUS program years.

The August 1-in-10 ex-ante impact is often used as the program's maximum possible impact. Because of this, it is the one figure resulting from the program's evaluation that is watched very closely by program stakeholders, it is shown in bold in Table 4. The 2011 evaluation, which was completed using a within-subjects method using AC data loggers, produced an ex-ante impact for 50% cycling of 0.56 kW. There are two potential reasons why the 2012 50% cycling figure is lower than the 2011 figure. First, the data used to complete the 2011 evaluation is actually dated from 2009 and 2010. Load control devices have a finite useful life and start to fail once they have been installed for several years, and since the last round of data collection in 2009 and 2010, a considerable amount of load control devices have potentially failed. Broken load control devices do not receive activation signals and are unable to provide load impacts. Non-functioning load control devices lower the average event impact for the program as a whole. A second reason why the 50% cycling figure shown in Table 4 may be lower than that created as part of the 2011 evaluation is that the between-subjects method used in the 2012 evaluation is believed to be more accurate than the method used in 2011. The 2012 results provides a clearer picture of actual performance across the province, whereas the 2011 evaluation relied on old data from just two urban LDCs. The fact that 60% cycling yields an ex-ante load impact that is closer to the 2011 ex-ante value for 50% cycling is one of the reasons why 60% cycling is recommended for adoption as the program's future cycling strategy.

Conclusions

Lessons Learned

Since the 2012 evaluation effort marked the first time the randomized experiment was used to estimate impacts for *peaksaver* and *peaksaver*PLUS, several useful lessons were learned. The first lesson is somewhat unique to OPA's situation, but is still applicable more generally: it is important to form solid relationships with multiple stakeholders and to coordinate fluid communication between stakeholders. The evaluation required buy-in from program officers and managers from each of the LDCs participating in the evaluation. LDC participation was critical, because they provided all needed customer data before, during, and after load control events were called. Another important stakeholder was the company responsible for sending paging signals to each of the devices. These paging signals set up the device groups that form the basis for various control and treatment groups and there is often extensive lead time involved in creating the groups in the head-end software.

A second lesson learned from the evaluation is that it can be difficult to obtain smart meter data in a timely manner. Most metering systems are optimized to provide data for billing purposes, but not for evaluation or research purposes. Billing data only needs to be extracted once a month, and depending on the utility, each month's usage only needs to be subdivided into a few time blocks. However, the data needed for evaluation purposes must be provided at the hourly level, and, even though smart meters are designed to provide data at this level, it was problematic for some LDCs to get adequate processes in place to provide hourly-level smart meter interval data. The main lesson learned from this particular problem was to give LDCs ample time to provide data and to relay very clear instructions about the evaluation's data requirements.

There is significant value in obtaining data shortly after a load control event to determine the success of a given activation dispatch. A timely analysis could prove valuable in identifying communication issues, broken devices and/or operational errors. Experience to date has shown LDCs require on average 4-6 weeks in order to fulfill a data request. These delays are often not a result of data unavailability, but are mainly due to the requirement of having skilled IT professionals query complex databases. This significant lag in fulfilling request leads to lost opportunities to improve program performance during an activation season. OPA is currently working with IESO to develop protocols and procedures to efficiently access smart meter interval data housed in the MDM/R, while maintaining data integrity and customer privacy.

Future Evaluation Plans

For the 2013 *peaksaver* PLUS evaluation, the OPA plans to maintain the randomized experiment design to develop load impact estimates for the various end-uses eligible for load control through *peaksaver* PLUS. The OPA is interested in continuing to develop it's understanding of the operations and impact of *peaksaver* PLUS. By using smart meter interval data, evaluators are able to work with information from thousands of participants across Ontario at a reasonable cost. By learning more about the program through large scale analysis, it allows OPA to enhance the problem to meet evolving needs.

The investment in building infrastructure to securely transfer and store data in additional to installing smart meters across Ontario has been significant. Exploiting smart meter interval data has been limited to date, however there are concrete steps being taken to make the data more accessible going forward.