OEB Pilot Plan
Technical Manual

PREPARED FOR
Ontario Energy Board

PREPARED BY

Ahmad Faruqui, Ph.D.
Neil Lessem, Ph.D.
Sanem Sergici, Ph.D.

July 16, 2016

THE Brattle GROUP

Acknowledgement: We acknowledge the valuable contributions of many individuals to this report and to the underlying analysis, including Brian Hewson, Roy Hrab and Anthony Ionno of the Ontario Energy Board, Tom Aagaard of the IESO, the RPP Working Group, and The Brattle Group for peer review.

Table of Contents

I. Introduction

A. Purpose of OEB Pilot Plan Technical Manual

The OEB's Regulated Price Plan (RPP) Roadmap sets out a five-point plan to be implemented over the course of the next 3 to 5 years. This multi-year process will undertake a comprehensive revamping of the RPP that will make incremental changes over the course of the plan to provide consumers with an adequate amount of time to adjust to the changes. Through this plan, the OEB intends to redesign the RPP to better respond to policy objectives, improve system efficiency, and give greater consumer control.

As part of the RPP Roadmap, the OEB established a working group with distributors, the IESO and consumer representatives to discuss potential options for price and non-price pilots as part of implementing the OEB's new RPP policy. These options and the application procedure for the new wave of pilots are described in the OEB Pilot Guidelines.

By incorporating input from the Working Group, the intention of this technical manual is to supplement the OEB Pilot Guidelines by explaining what the OEB's pilot objectives are, and laying out a framework for LDCs to design, implement and evaluate these pilots. These guidelines lay out the best practices in experimental design, recruitment, survey design and pilot outputs that need to be incorporated into pilot design and implementation, which will differ for the price and non-price pilots.

B. Background

The OEB recently undertook an initiative – summarized in the RPP Roadmap – to assess whether the RPP is meeting public policy objectives, to determine if improvements can be made, and to ensure that RPP customers are maximizing the value of the program, with the overarching goal that the RPP is able to meet future challenges. The current stable state of Ontario's electricity system, as indicated by the IESO's most recent forecasts, provides an ideal opportunity to make adjustments to the RPP.[1] To help determine future policy directions, the OEB commissioned studies exploring how the RPP has performed and how consumers perceive the RPP, as well as studies exploring potential changes in pricing and consumer information.

- **Navigant Consulting** undertook a two-part study of TOU rates, and how TOU could be improved through structural changes (part one, part two).
- **Ipsos Reid** conducted (1) exploratory qualitative research, using both online and traditional focus groups; and (2) a second phase of quantitative research using a telephone survey, which was developed based on the results of the focus groups
- **BEworks** applied a behavioural economics approach to review and assess the ways in which consumers are, and are not, responding to the current pricing structure in Ontario
- **Power Advisory** conducted a review of dynamic pricing in leading jurisdictions in North America and beyond

The conclusion reached through these studies and in the general review of the RPP is that:

- **Of the four objectives originally set out in the RPP, two are not being achieved:**[2]
  - *Create a price structure that is easily understood by consumers.* There is a need for improved provision of information, such as improvements to bills, fostering better understanding of energy usage, and providing advice regarding which actions to take and when to take them. In other words, consumer response could be improved through education.
  - *Set both prices and a price structure to give consumers incentives and opportunities to reduce their electricity bills by shifting their time of electricity use***.** While there is a focus on conservation, the current time-of-use pricing structure has not provided sufficient incentives for consumers to shift and/or reduce use and does not contribute to long-term goals outlined in Ontario's Long Term Energy Plan (LTEP), such as targeting infrastructure deferral.

---

1 While Ontario is expected to remain long on supply for the next several years, capacity constraints are likely to become binding again at some stage in the period 2020 to 2025.

2 The four objectives were: i) set prices to recover the full cost of RPP supply; ii) set the price structure to reflect RPP supply costs; iii) set both prices and the price structure to give consumers incentives and opportunities to reduce their electricity bills by shifting their time of electricity use; and iv) create a price structure that is easily understood by consumers.

- **The existing RPP objectives do not reflect current public policy objectives**. While the new LTEP policy focuses on long term system benefits, cost-effectiveness, and reducing peak demand, this focus is not reflected in the current objectives of the RPP as developed by the OEB. The RPP needs to take these policy priorities into account so that the OEB can take a long term view when setting and designing prices.
- **Optimal pricing structures can only be achieved with greater flexibility**. For example, there is limited ability to set efficient rates that could reward response, due to restrictions on off-peak hours (7pm to 7am) required by regulation. Or, the way in which the global adjustment (GA) is recovered from Class B consumers (residential and small business customers with a peak demand between five kilowatts and five megawatts) limits flexibility in pricing, as does the practice of setting prices based on short-run forecast system costs.

To address these issues, the RPP Roadmap sets out a **five-point plan** to be implemented over the course of the next three to five years:

1. **Renewing the RPP objectives.** The OEB will immediately update the RPP objectives as follows (changes and additions to the original objectives are *italicized*):
   - Set prices to recover the full cost of RPP supply, on a forecast basis, from the consumers who pay the prices;
   - Set the price structure to reflect *current and future* RPP supply costs;
   - Set the price structure to support the achievement of efficient electricity system operation and investment;
   - Set both prices and the price structure to give consumers incentives and opportunities to reduce their electricity bills by shifting their time of electricity use *and reducing their peak demand;*
     - Create a price structure that is easily understood by consumers;
     - *Provide fair, stable and predictable commodity prices to consumers.*
2. **Empowering Consumers: enhancing energy literacy and non-price tools**. Non-price tools and pilots (e.g., the electricity bill, automated load controls, benchmarking, and data collection) will be used to help make TOU pricing more easily understood by consumers.
3. **Implementing price pilots**. The OEB will work with LDCs to undertake several pricing (and non-price) pilots over the next 18 months. The pilots will run for at least one calendar year to assess whether there is persistence in the impact of the intervention.

4. **Engaging with low volume business consumers**. The OEB will seek to engage low-volume business consumers through data collection surveys and a series of meetings to discuss high level concerns for this consumer class to help inform the design of future pricing options.

5. **Working with government to reduce barriers**. The OEB is committed to working with the government and the IESO to address the regulations which fix the TOU time periods and the recovery of GA costs which limit the OEB's ability to set optimal TOU prices.

The OEB has retained The Brattle Group to assist with the implementation of the third objective listed above, "Implementing Pricing Pilots". This report details the proposed plan for meeting this objective.

II. Pilot Objectives

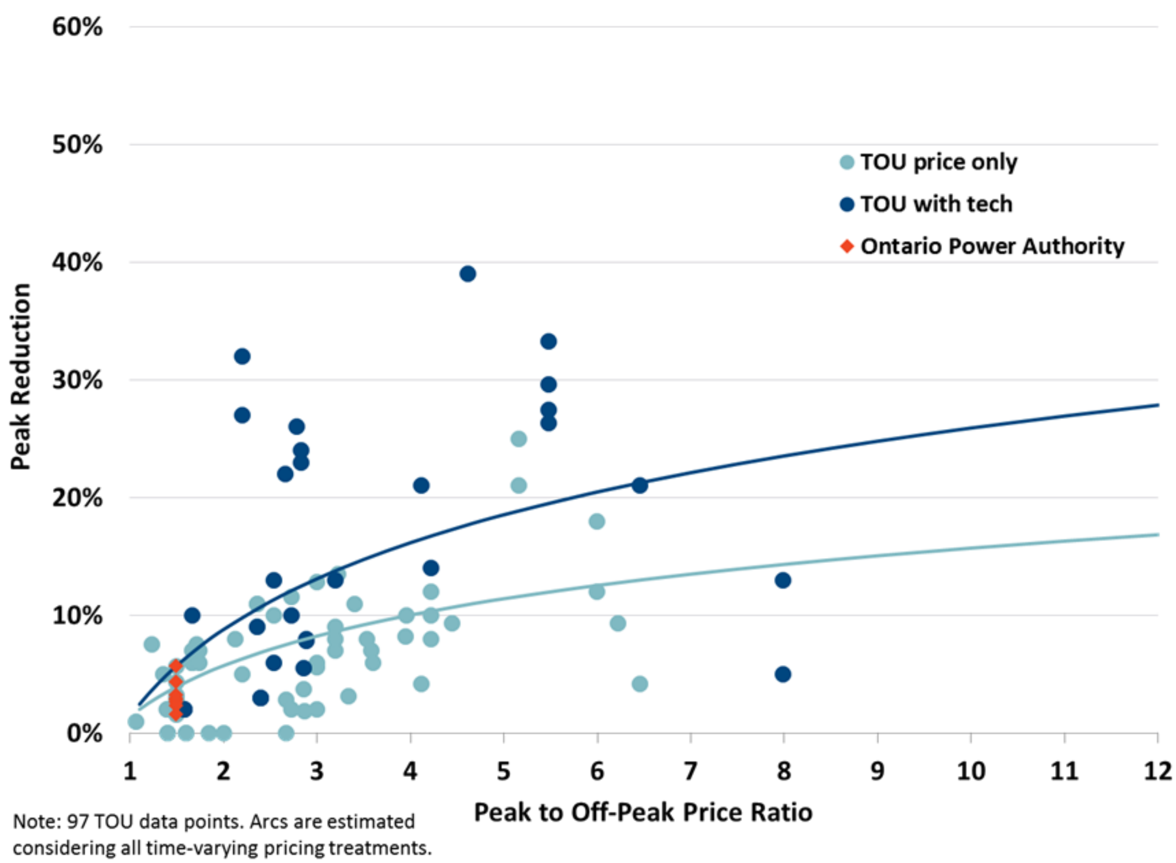
The OEB is calling for proposals for a series of pilots to test price and non-price programs that may eventually be scaled up province-wide. The objectives of these pilots are (as indicated in the RPP Roadmap):

1. Set the price structure to support the achievement of efficient electricity system operation and investment;
2. Set both prices and the price structure to give consumers incentives and opportunities to reduce their electricity bills by shifting their time of electricity use and reducing their peak demand;
3. Empowering Consumers: Enhancing energy literacy and non-price tools.

The IESO has identified that in the future, fast and flexible energy resources will be important to help integrate the increasing amount of renewable energy being placed on the grid. Due to the intermittent nature of most renewable resources, cost-effective resources that can instantly be called on to ramp down (or up) will become increasingly important. Customer demand response through prices and other mechanisms can possibly provide some of these resources more cost-effectively than building generation and distribution system assets. In addition, Ontario is in a stable supply situation for the next several years and capacity constraints are not expected to become binding again until at least the mid-to-late 2020s, according to recent planning outlooks by the IESO. Reducing peak demand, through load shifting or peak clipping, can reduce and further delay the need for future generation capacity investments.

The current RPP for most consumers in Ontario is a Time of Use (TOU) rate. TOU rates encourage customers to shift electricity usage from the periods where it is generally more expensive to produce, to the periods where it is generally cheaper (both in terms of generation costs and future capacity investments). The Brattle Group has collated the results from 97 TOU programs across the globe, including the TOU rate in Ontario, in the Brattle Arc of Price Responsiveness. The TOU rate in Ontario is the only system-wide TOU deployment. The Arc, which is shown in Figure 1, illustrates the relationship between the peak to off-peak price ratio and peak reductions. The higher the peak to off-peak price ratio, the larger the peak reduction. Ontario, shown in red, has a relatively low peak to off-peak price ratio, but peak reductions are largely in line with results elsewhere. The Arc also shows that customer price responsiveness is increased when customers are given access to enabling technologies.

Figure 1: The Brattle Arc of Price Responsiveness for Time of Use (TOU) Rates



Source: Faruqui, Ahmad. "Arcturus", The Brattle Group.

While TOU addresses general trends in electricity pricing across time, it does not address specific intra or inter-day variation. Electricity cannot be cheaply stored leading to large variation in wholesale prices, even over a single day. Dynamic pricing policies can be designed to encourage customer response to these specific price changes rather than the general trends that TOU addresses.

Dynamic pricing, where prices reflect the time-varying marginal costs of generating electricity, is inherently more efficient than "static pricing". By providing a price signal that reflects real-time market conditions and costs, dynamic pricing enhances the efficiency of consumer consumption decisions, promoting allocative efficiency. Consumers consume electricity only if they value it at least equal to its cost. Dynamic pricing also promotes more efficient system operations. Higher prices during peak periods when higher-cost resources are operating cause consumers to reduce consumption and their reliance on these resources. Dynamic pricing can also be a strong deterrent to the exercise of market power, as consumers respond to high prices by reducing demand, which reduces the sales of the supplier that is exercising market power by withholding output, thus offsetting the benefit from these higher prices. Conversely, lower prices during low demand periods or when there is additional output available from variable output renewable energy resources will result in higher demand, which in turn can facilitate the integration of these resources. In the long term, productive efficiency is promoted whereby the development of electricity infrastructure is efficient. For example, dynamic pricing can reduce requirements for peaking generation resources and transmission and distribution investment through higher prices during peak periods or when the electricity supply infrastructure is stressed." - Power Advisory (2014).

The priority pricing pilots identified in the OEB Pilot Guidelines reflect both modifications to the existing TOU rate and the introduction of dynamic pricing

Finally, empowering customers involves educating them about their energy consumption habits, offering choices so that they can choose rates that allow them to balance out their energy bills and lifestyle needs, and providing tools for them to make more informed energy consumption decisions. The BEworks report found that consumer understanding of how to respond to TOU pricing was low, but that this could be improved by providing consumers with better visuals and clearer information about TOU periods and prices. The Ipsos Reid research also indicated that a majority of consumers want more information about which appliances use the most electricity and also how much money they would save if they changed when they use these appliances. The importance of consumer education was further supported by the Power Advisory report.

Moreover, the BEworks research indicated that some consumers are more likely to respond to TOU pricing for social and/or environmental reasons rather than financial reasons.

Pilot design and implementation will differ for the price and non-price pilots. Pricing pilots are intended to assess new price structures that replace the current RPP default option (Default RPP), or be implemented as a tariff choice to augment a future default RPP offering (RPP Choice). This future pricing regime will be undertaken at the provincial level and thus it is important to have a thorough understanding of its likely impacts across the province. Non-price pilots will lead to programs that LDCs can individually opt-in to implement, although some may be required as part of an RPP energy literacy program. These interventions are unlikely to have the risk of negative bill impacts inherent in a change in pricing. For this reason and to encourage greater creativity, these pilots can be run individually by LDCs or in conjunction with pricing pilots.

For both pricing and non-pricing pilots, the OEB and its advisors will provide guidance and advice on the adequacy of the experimental design, sampling, survey and measurement and verification approach. The OEB will also be requiring regular updates from participating LDCs throughout the life of the pilot.

III. Pilot Implementation Guidelines

This section outlines best practices for deploying pilots. All proposed pilot studies must adhere to these general guidelines. Where necessary, there will be more specific details given in the pricing and non-pricing sections, where implementation differs from the guidelines outlined below.

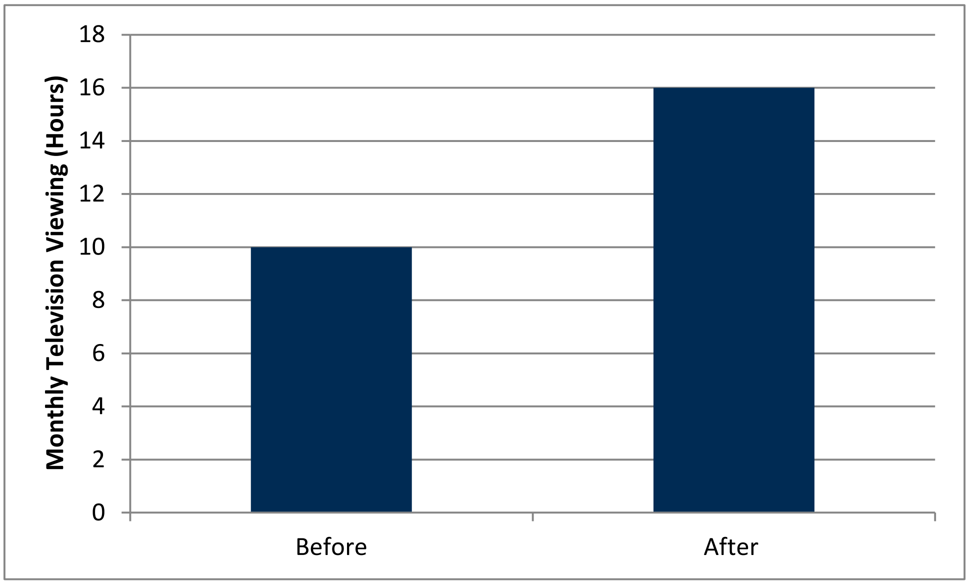
A. Experimental Design

1. Internal Validity

Once a pilot treatment or combination of treatments has been selected and the outcome(s) of interest for the particular program defined, the pilot program should be designed such that a cause and effect relationship can be identified between the two. This characteristic of well-defined programs, known as **internal validity**, means that the estimated effect can be attributed to the intervention as opposed to any other factors. All confounding factors must be controlled for so that the true treatment effect may be isolated.

For example, imagine if we wanted to evaluate the impact that offering several new television channels would have on the amount of time that the average household spent watching television. Figure 2 shows the number of hours that the average household spent watching television before and after the intervention.

Figure 2: Impact of new channels on hours spent watching television (illustrative example)



By merely looking at the period before the new channels were offered and the period afterward, we may falsely conclude that the introduction of new television channels increased the number of hours spent watching television by 60 percent (6 hours per month). However this ignores the fact that other factors may have changed in the period after the introduction of the new channels. For example, we may find that the Olympics were screening in this period, causing much of the increase in television viewing.

Figure 3: Impact of new channels on hours spent watching television (illustrative example)

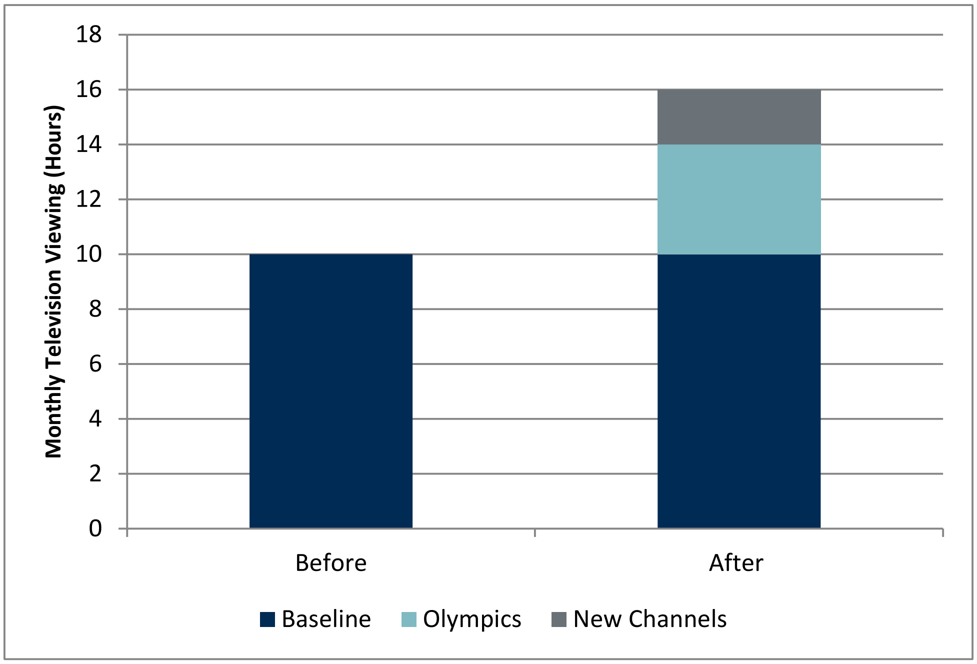


Figure 3 shows that two thirds of the increase in television watching hours was caused by the Olympics and the intervention of more channels actually only caused a 20 percent increase in viewing hours.

Of course, there may be multiple factors that change after the intervention, some of which may be unknown to the evaluator. The best means of accounting for these factors and ensuring internal validity is to associate every treated group of customers with a statistically indistinguishable control group. The control group, which does not receive the intervention, represents the outcomes that would have been observed among the treatment group "but for" the intervention. Deviations in outcomes between the treated group and an appropriately designed and constructed control group constitute the true impact of the intervention.

In addition, to further validate the measurements, it is best to have pre-treatment data on both the control and treatment groups as well as the treatment-period data. Such data enable a *difference-in-differences* approach. In Figure 4 below, the values of the outcomes of interest, such as annual usage, are represented by C1 and T1 for the control and treatment customer groups respectively. The treatment period outcomes are represented by C2 and T2. The difference between the change in treated customer behavior and the change in control customer behavior, represented by {(T2-T1)-(C2-C1)}, is the estimated effect of the program.

Figure 4: Treatment and Control Group Observations

	Control Group	Treatment Group
Pre-Treatment	C1	T1
Post-Treatment	C2	T2

Without a control group in the program design (i.e., relying on only (T2-T1)), it is impossible to control for external factors that may change between the pre-treatment and treatment periods, such as macroeconomic changes or general changes in attitudes toward energy use brought about by other exogenous factors. Without pre-treatment data (i.e., relying on only (C2-T2)), it is difficult to verify that the treatment and control groups were truly comparable before the treatment was introduced. When systematic pre-treatment differences exist, there may be selection bias or other problems in the sample that will bias the estimated effect size. Panel studies, with treatment and control customers and pre-treatment and treatment period data, are thus most powerful and most likely to produce internally valid estimates.

Random assignment of the customer sample into the treatment and control groups, as in a **randomized controlled trial (RCT)**,[3] is widely regarded as the best means of meeting the internal validity requirement. Even with random assignment, however, some *ex-post* comparative analysis should be undertaken to ensure that the treatment and control groups are statistically indistinguishable.

In some instances, depending on the research question and program context, mandatory assignment of customers to the treatment group is not feasible or appropriate. If the program involves opt-in enrollment, for example, customers may not comply with their assigned group. A

[3] For further discussion of RCT experiments, see the Ontario Power authority document *Protocols for Evaluating Behavioral Programs*, section 4.3.1 beginning on page 17. http://www.powerauthority.on.ca/sites/default/files/Behaviour-Based-Evaluation-Protocols.pdf

randomized encouragement design (RED)[4] allows the researcher to construct a valid control group in these situations, maintaining the benefits of an RCT. This approach still involves random assignment of customers to treatment and control groups, but in this case the treated group is *encouraged* to apply for the intervention rather than provided with it automatically.

In contrast to a RCT, in a RED researchers indirectly manipulate program participation using an encouragement "instrument". The instrument may be as simple as extending an offer to opt-in to the program. Administering the instrument only to the treatment group produces the variation in participation needed to generate a causal estimate of the treatment effect. As discussed above, the experimental design should reflect the eventual scale-up of the pilot to the entire target population. If it will be offered on an opt-in basis, the pilot should likewise be conducted in that manner.

2. External Validity

Internal validity ensures that the estimated effect can be causally attributed to the intervention. **External validity**, on the other hand, ensures that the estimate can be extrapolated to customers outside the sample. The external validity of an experimental design depends on how customers are sampled from the target population about whom inferences are to be made and how well the sample treatments match real world treatments. An appropriate methodology for extrapolating estimated program impacts to future years is less established, and it is generally not valid to assume that a program will produce exactly the same level of savings over time. Valid estimates of treatment effects can nonetheless provide insight to the type of interventions that are worth scaling up.

The gold standard in terms of external validity, as well as in terms of internal validity, is an RCT design. In this design, the study sample is drawn randomly from the entire population that would be included in full policy or program implementation. The sample reflects the full target population and covers the full range of characteristics that may affect energy use among the relevant customers. Moreover, an adequately sized random sample will reflect customer characteristics in proportion to that of the full population, so the estimated treatment effect will be valid for the target population.

---

[4] For further discussion of RED experiments, see *Protocols for Evaluating Behavioral Programs*, section 4.3.2 beginning on page 19.

While a randomly selected sample is the most straightforward approach, *stratification with proportional allocation* offers the same advantages as well as some additional benefits. Stratification divides the target population into subgroups of interest, sometimes for administrative or operational convenience. Proportional allocation means that samples from each subgroup are drawn in proportion to each subgroup's presence in the full target population. This procedure reduces the risk that the mix of customers in the sample may be distorted along lines of interest due to random chance.

In the context of the OEB pilot plan, pilots programs will be run by LDCs at the LDC level. By their nature, no single LDC will be able to select a sample from its service population that is representative of the entire province. *Stratification with non-proportional sampling* may be useful in achieving better representation of the overall Ontario population. Alternatively, or in addition, high-priority pilots may be run across several LDCs to test and extend the external validity of estimated treatment effects. Stratification with non-proportional sampling across subgroups may also be useful more generally to ensure adequate coverage of particular customer groups of interest, such as large or low-income customers. If large customers are of particular interest, for example, because they may have more scope for energy savings, a disproportionately high number of large customers may be included in the sample to ensure that the treatment effect estimated for them is sufficiently statistically precise. Higher sampling rates for larger customers also tend to provide improved accuracy for regression analysis.

The disadvantage of stratification is a more complex and potentially more costly design and recruitment process. With non-proportional sampling across subgroups in particular, the experimental design and evaluation may be less transparent. In terms of the analysis in this situation, sampling weights may be needed in regression analysis to control for the over- and under-representation of particular subgroups. Both probability-weighted and unweighted regression coefficients should be calculated and the model should be investigated if the two sets of coefficients are very different, as a significant difference could indicate that the model is missing an important covariate.[5] In addition, treatment effects for each sampling cell or subgroup may be estimated separately, but this approach requires sufficient sample size within each cell in order to reach statistically significant conclusions.

---

[5] This is particularly true if the direction or magnitude of the differential impacts for each subgroup are not accounted for by theory or common sense.

In order to ensure that results are generalizable to the population of interest, the experimental treatment should match the likely real world treatment as closely as possible. As explained in the OEB Pilot Guidelines, the manner in which the *pilot is deployed should mirror the likely deployment scenario if the pilot were to be implemented as a province-wide policy going forward.* Another important dimension *is whether the customers are offered bill protection or not*. Bill protection ensures that the customers would not pay for more than what they are currently paying under the current tariff. If a customer is operating under the bill protection framework and knows that in the worst-case scenario their bills will not be higher than what they would normally pay under their current tariff, their incentives to respond to price and information signals would likely diminish. Since it will not be practical to offer customers bill protection under a new baseline RPP tariff, pilots assessing the impact of changing baseline RPP tariff must not offer bill protection. This requirement also holds true for all other price pilots.

Although this guidance also holds true for pilots offering rate choice/RPP menu, the potential impact of "bill protection" on customer incentives can be expected to matter less. This is simply because the customers make a choice and opt-in to these rates in the first place, signaling their potential engagement level.

3. Sample Size

In addition to the selection of an appropriate sampling methodology, statistical power analyses must be conducted in advance of recruitment to determine the required sizes of the treatment and control groups. The factors that determine the required number of treatment and control customers to be included in the analysis are:

- The desired level of statistical precision
- The expected magnitude of the treatment effect to be estimated
- Whether hypothesis testing will be one- or two-sided
- The ratio of treatment and control group sizes[6]
- The number of pre-treatment observations per customer
- The number of treatment period observations per customer

[6] Treatment customers can be traded off for control customers, but as the ratio of treatment to control customers decreases, more and more additional control customers will be needed to compensate for each additional treatment customer.

- The variance-covariance structure of the data (how correlated certain sets of observations are with one another)

Larger sample sizes are required to achieve greater statistical precision and a smaller confidence interval around the estimated effect. As the detection threshold (the minimum treatment effect size to be detected) becomes smaller, the sample size must increase to maintain the same level of statistical precision. In other words, the smaller the effect size that the researcher wishes to be able to identify at a given level of precision, the larger the sample needs to be. One sided hypothesis testing, where the research question is whether the treatment either increases or decreases the outcome of interest but not both, requires a smaller sample than two-sided testing.

It is also important to note that the treatment and control groups do not need to be equal in size within a pilot program. Inclusion of more customers in the sample, regardless of whether they are assigned to the control or treatment group, increases the amount of available information, improving statistical precision. However, there are diminishing returns to increasing the size of either treatment or control group. The optimal tradeoff will depend on the extent to which it is less costly, or otherwise more desirable, to increase the number of control group customers instead of the number of treatment group customers.

Most pilots involve a random sample of customers that are representative of a given population to ensure that the average impact observed in the pilot can be applied to the population at large. Designing, deploying and running a pilot program is costly, and there is usually a trade-off between the cost of a pilot and the number of customers included the pilot. If one of the objectives of the pilot is to obtain statistically significant impact results for certain sub-groups in the population (i.e., low-income customers, senior citizens, etc.), then the sample should be stratified to ensure each sub-group has a sufficiently large sample size to yield statistically significant results. This will increase the overall sample size, increasing both the scale and cost of the pilot.

Ideally customers should only be participating in one pilot at a time; however, participation in multiple pilots is acceptable if the treatments are unlikely to impact each other. For example, the impact of a pricing pilot may be larger if the customer has access to an IHD through another pilot study. In this case we would want to ensure that they do not participate in both treatments. However, a window glazing pilot is unlikely to impact a customer's response to a critical peak price and having a customer participate in both pilots would be acceptable, if information about participation in both pilots is available and can be accounted for in the impact evaluation.

After excluding customers who are ineligible due to other pilots, who do not have sufficient amounts of historical data (usually one year before treatment begins), and other non-representative customers (extreme high/low usage, seasonal occupation, etc.), the pool of pilot-eligible customers may be far smaller than the LDC's total customer base. This will vary across LDCs based on their customer base's demographics. For instance, a large student population will negatively impact the percentage of accounts with at least one year of historical data.

The Working Group have identified several issues based on their past pilot experience that decrease sample sizes. Over sampling can help maintain statistically significant results in the presence of such issues as:

- Attrition, including the control group if applicable
- Low participation: only a fraction of customers will take advantage of any single hardware feature or technology option.
- Hardware or communication issues: not all treatments customers will effectively receive treatment

B. Pilot Outputs

1. Measurement and Verification

Each LDC is expected to submit a Measurement and Verification Plan along with their application to ensure that the results of the pilot are evaluated using widely accepted statistical techniques.[7] The experimental design of each pilot dictates the optimal evaluation method. Estimation of treatment impacts may be done via differences-in-differences (ANOVA or ANCOVA), panel regressions (fixed-effects, random-effects, or instrumental variables), or individual customer regressions, depending on experimental design. Typically fixed effects and instrumental variables panel regressions yield the most robust and reliable impact evaluations.

In general, impact metrics fall into two categories: those regarding electricity consumption and demand (annual consumption, average hourly consumption, system peak demand); and those

[7] Guidelines for developing an evaluation plan are discussed in more detail by the Ontario Power Authority *Evaluation, Measurement, and Verification (EM&V) Protocols and Requirements* document. See especially "Step 7: Evaluation Plan Development Guidelines" beginning on page 26. http://www.powerauthority.on.ca/sites/default/files/conservation/Conservation-First-EMandV-Protocols-and-Requirements-2015-2020-Apr29-2015.pdf

regarding demand elasticities (own-price, and substitution elasticities). There might be variation in terms of the metrics reported by each of the pricing pilots to the extent that their designs differ, but at a minimum each pricing pilot should aim to report each of the following, where applicable:

- Peak, mid-peak, and off-peak impacts
- Critical event peak period impact
- Average daily conservation impact
- Own/daily price elasticity
- Inter-period substitution elasticity

Since non-price pilots are largely heterogeneous in terms of their definition, the metrics should be defined on a case-by-case basis.

C. Data Requirements

1. Survey Data

High-quality surveys can provide invaluable information to complement and aid in the analysis of pilot program data. Such surveys fall into four basic categories: i) market response surveys to assess customers' reactions to enrollment mechanisms; ii) pre-treatment surveys to study baseline household conditions; iii) within-experiment surveys to study conditions during experiments; and iv) post-treatment surveys to study household conditions after the intervention.

Assessment of changes in customer energy literacy, and the extent to which any observed effects persist, will require a panel survey approach where individual households are surveyed multiple times. As in the analysis of program treatment effects, customer attrition must be monitored and addressed in the analysis of survey responses if it is found to be significant.

In addition to behavioral changes, it is also important to measure customer acceptance and satisfaction. Two essential questions in post-treatment surveys are:

1. How satisfied are you with [program name]? (5 to 7 Likert scales)
2. Would you recommend [program name] to friends? (5 to 7 Likert scales)

To facilitate the aggregation of survey results across LDCs, pre-defined questions will be provided by the OEB. Useful survey questions in general regard household characteristics, appliance holdings, business characteristics, or customer awareness and education. When possible, information on customer characteristics and appliance holdings should be collected as part of the process of enrolling experimental subjects into treatment and control groups.

Surveys should also rely on consistent protocols for minimizing and analyzing non-response. Response rates should be similar across the treatment and control groups, and analysis of survey data should address any potential non-response bias. Response rates in excess of 60% are achievable with reasonable expenditure of resources, and should be targeted when discussing expected performance with survey providers.

The guidelines that apply to sample selection for pilot programs also apply to the design of survey samples. Samples should be randomly drawn from study cells or subgroups, and from treatment and control. The size of each sample should be sufficient to estimate parameters of interest at pre-determined levels of statistical confidence.

2. Ongoing Data Requirements

LDCs will need both hourly metering data from the IESO's Meter Data Management Repository (MDMR) and customer account data. Hourly metering data stored in the MDMR is linked to a particular premise rather than a particular customer. If two customers inhabit the same address at different times within the evaluation timeframe, this transition would not be apparent using just the hourly metered data. For this reason, metering data needs to be supplemented with customer account data that includes move-in and move-out dates.

Customer account data can also help increase precision in the study by accounting for other utility programs in which the customer participates. If particular subpopulations are of interest, these customers will need to be identified using the account level data. Alternatively, if these data do not exist, customer membership in particular subgroups can be identified through surveys tied to account data, or through geo-locational data that can be linked to census data. Matching customer data across data sources can be challenging, but constitutes a very critical part of the impact evaluation efforts. For example in one pilot, matching customers that identified themselves as participants of the IESO Province-Wide Heating and Cooling program ("HVAC") with the LDC report on HVAC participants resulted in delayed validation of program enrollment.

There are several important considerations that need to be taken into account when dealing with the data process:

1. **Timing**: obtaining large data pulls from the MDMR requires substantial effort on the IESO's part. Depending on their ability to fulfill these requests, this can result in unanticipated delays. Coordination with the IESO in advance can avoid these delays.[8]
2. **Estimated data**: estimates for missing data in the MDMR are interpolations and represent a "false" data point. They should potentially be disregarded in analysis.
3. **Move outs**: customers should be removed from the analysis when they are no longer in the original premise.
4. **Time zones**: data should be reported in the correct time zone with daylight savings accounted for. Data in the MDMR are stored in Coordinated Universal Time (UTC) and is hour ending. Data requests to the MDMR should be made in local Eastern Standard Time.
5. **Time Stamp:** data should be reported using a consistent hour beginning or hour ending convention.

Obtaining customer information may entail a legal review, and the implementation of processes related to customer data access, obtaining consent, and access to information can take much longer than anticipated. To expedite this process ensure that the evaluator will have access to all necessary data, including that of treatment and control group customers prior to launch. The policies concerning customer energy and non-energy data vary significantly by LDC; following the privacy guidelines laid out in "Privacy by Design," which is endorsed by the Ontario Privacy Commissioner, can mitigate the risks of potential misuse of data, either knowingly or unknowingly.[9] Additionally, the IESO primer on privacy issues with social benchmarking programs is likely to be applicable to some of the proposed RPP pilots. This document is available on request from the IESO.

---

[8] If the LDC does not participate in the IESO's MDMR, LDCs should with their respective metering/data departments that they have the technical capability to provide the large data files required for pilot selection and EM&V.

[9] See Privacy by Design website, www.privacybydesign.ca

CAMBRIDGE
NEW YORK
SAN FRANCISCO
WASHINGTON
TORONTO
LONDON
MADRID
ROME
SYDNEY

THE Brattle GROUP