



ONTARIO POWER AUTHORITY



Protocols for Estimating Load Impacts Associated with Demand Response Resources in Ontario

December 31, 2009

**Protocols for Estimating Load Impacts Associated
with Demand Response Resources in Ontario**

December 31, 2009

Prepared for:
Ontario Power Authority
120 Adelaide Street West, Suite 1600
Toronto, Ontario M5H 1T1

Prepared by:
Freeman, Sullivan & Co.
Stephen S. George, PhD
Josh Bode, MPP

TABLE OF CONTENTS

A.	INTRODUCTION	2
A.1.	LOAD IMPACT ESTIMATION OBJECTIVES	4
A.2.	CONTEXT AND USES FOR LOAD IMPACTS.....	4
A.3.	HOW TO USE THIS DOCUMENT	6
B.	OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS	8
B.1.	DEFINING LOAD IMPACTS	9
B.2.	CHARACTERISTICS OF DEMAND RESPONSE RESOURCES.....	11
B.3.	EX POST VESUS EX ANTE IMPACT ESTIMATION.....	13
B.4.	CHANGES IN PROGRAM ENROLMENT	15
B.5.	TIME PERIODS	16
B.6.	DAY TYPES	16
B.7.	EVENT WINDOW.....	17
B.8.	EXTREME CONDITIONS.....	17
B.9.	UNCERTAINTY	18
B.10.	LOCATION SPECIFIC IMPACTS.....	19
B.11.	CUSTOMER PRICE ELASTICITIES	20
B.12.	FREE RIDERS AND STRUCTURAL BENEFITERS	21
B.13.	DISTRIBUTIONAL IMPACTS	22
B.14.	PERSISTENCE AND LONG-TERM IMPACTS	23
B.15.	UNDERSTANDING WHY, NOT JUST WHAT	23
C.	LOAD IMPACT ESTIMATION PROTOCOLS	25
C.1.	PROTOCOL 1—EVALUATION PLANNING	25
C.2.	PROTOCOL 2—TIME PERIODS.....	27
C.3.	PROTOCOL 3—REPORTING FORMAT	27
C.4.	PROTOCOL 4—DAY TYPES AND EVENT CONDITIONS.....	30
C.5.	PROTOCOL 5—PORTFOLIO ANALYSIS	33
C.6.	PROTOCOL 6—STATISTICAL REPORTING AND VALIDATION	33
C.7.	PROTOCOL 7—ANALYSIS BASED ON SAMPLING	34
C.8.	PROTOCOL 8—REPORTING AND DOCUMENTATION.....	36
	GLOSSARY OF TERMS	40
	APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES	42
	APPENDIX 2: SAMPLING OVERVIEW	48
	APPENDIX 3: IMPACT EVALUATION METHODS	65
	APPENDIX 4: OVERVIEW OF REGRESSION METHODS	70

A. INTRODUCTION

In recent years, the Province of Ontario has adopted ambitious conservation goals¹ and assigned a leadership role to the Ontario Power Authority (OPA) in ensuring that they are met. In planning to meet the current goal of 6,300 MW of peak demand reduction by 2025, the OPA expects a significant contribution from customers enrolled in demand response (DR) resources. To date, this has primarily entailed DR programs administered by the OPA, but other options—such as DR programs offered by other entities including local distribution companies (LDCs) or the Independent Electricity System Operator (IESO), or time-varying retail rates approved by the Ontario Energy Board (OEB)—could also contribute to meeting future provincial targets.

The OPA is committed to the evaluation, measurement and verification (EM&V) of all conservation programs, including estimation of savings impacts based on data collected from actual program experience. This is necessary not only to assess progress toward meeting Provincial resource goals, but also to obtain information for improving program design and as input to resource planning. The OPA has developed a framework² and protocols³ for evaluating conservation programs that are designed to improve energy efficiency, based on methods developed over several decades of evaluating similar types of programs and initiatives across North America.

For demand response resources, evaluation methods are less well developed. Moreover, several key aspects of DR resources differ from energy efficiency in ways that impact their evaluation (see *How is Demand Response Different from Energy Efficiency?* on page 3). Recognizing these differences, the OPA initiated development of a separate DR evaluation framework that includes the following key elements:⁴

- a white paper⁵ that discusses issues and methods for estimating load impacts and cost-effectiveness of DR resources and provides recommendations on a DR evaluation framework—**completed in November 2008**;
- a set of protocols for estimating load impacts—**this document**; and
- a framework for determining the cost-effectiveness of DR programs—**to be completed**.

¹ “Conservation” is an umbrella term for four categories of demand-side resource: demand management/conservation behaviour; energy efficiency; fuel switching; and customer-based generation. Demand response is considered part of the first category.

² *Evaluation, Measurement & Verification Framework for Ontario Power Authority Conservation Programs*. Final Version 1.0, April 9, 2008. p.3-4.
http://www.powerauthority.on.ca/Storage/68/6326_OPA_EM&V_Framework_1.0.pdf.

³ *OPA Evaluation Protocols*. Undated.
http://www.powerauthority.on.ca/Storage/68/6325_FINAL_Evaluation_Protocols_-_Single_Document.pdf.

⁴ Process evaluation is another EM&V function that can inform program design and planning. This initiative does not include special guidelines for DR process evaluation beyond those laid out in the OPA EM&V framework and protocols.

⁵ Stephen George and Josh Bode. *Assessing Demand Response Cost-Effectiveness and Load Impacts in Ontario*, prepared by Freeman, Sullivan & Co. for the Ontario Power Authority. November 7, 2008.

How is Demand Response Different from Energy Efficiency?

1. Energy efficiency benefits are typically tied to the installation of specific energy-efficient devices, whereas DR benefits tend to result from behavioral actions associated with a variety of end-use activities. This has the following implications:
 - For energy efficiency, load impacts can often be estimated from engineering calculations that compare the relative efficiencies of two devices based on prescriptive or quasi-prescriptive measure assumptions.
 - For DR, load impacts are best determined from empirical analysis of load data at the individual customer, segment or program level.
2. Energy efficiency benefits typically accrue over the life of a device. For most DR options, the benefits are tied to the continuation of customer participation in a program in response to price signals or direct incentive payments. Even when technology is used to automate DR, consumers can override it, and the benefits may cease if the program incentives are removed.
3. DR resources often have constraints on their frequency and timing (determined by program rules) that must be factored into cost-effectiveness analysis.
4. Classic free-ridership (i.e., paying customers for something they would have done anyway) is a key component of energy-efficiency program cost-effectiveness, but is largely irrelevant for most DR options—people don't "do DR" in the absence of program incentives or time-varying prices.

"Structural beneficiaries" may exist with DR—that is, some consumers may pay lower bills or receive incentives even if they don't shift load, simply due to their inherent pattern of usage. However, structural beneficiaries are not necessarily non-responders, since on the margin, they receive the same benefits from DR as other customers, and may respond similarly.
5. Most DR resources are event driven—that is, benefits accrue only when an event is called. Energy efficiency benefits are continuous—although they may vary over time with the loads they affect, they are not "dispatchable."
6. Many DR resources have insurance or option value—that is, like insurance, much of the value of DR exists even when it is not used.
7. Energy efficiency benefits primarily derive from reductions in energy use, whereas DR benefits are mostly tied to reductions in capacity costs.
8. The magnitude of load impacts, and benefits, associated with DR may vary significantly across the hours of a day, days of the week, months and seasons due to exogenous factors such as weather and customer energy usage behavior. For DR resources that vary with weather, it is often the case that the magnitude of load reduction is greatest when the value of demand reduction is greatest (e.g., on hot summer afternoons).

A.1. LOAD IMPACT ESTIMATION OBJECTIVES

This document contains eight protocols for estimating the load impacts of demand response resource options. These protocols were designed to meet the following primary objectives:

- Establish **minimum requirements** to support resource planning, cost-effectiveness analysis and program design and improvement;
- **Focus on the outputs** that should be provided, rather than on how to obtain them;
- Develop a **common set of outputs** to enable “apples-to-apples” comparison of load impacts across DR resource options, event conditions, and time;
- Be **applicable to a wide range of DR resource options**, to accommodate a changing landscape of policies, programs, and program delivery agents;
- Ensure that the **documentation** of methods and results allow knowledgeable reviewers to judge the quality of the work and the validity of the impact estimates provided; and
- Encourage **recommendations for improvements** to the evaluated DR resources and future load impact evaluations.

A.2. CONTEXT AND USES FOR LOAD IMPACTS

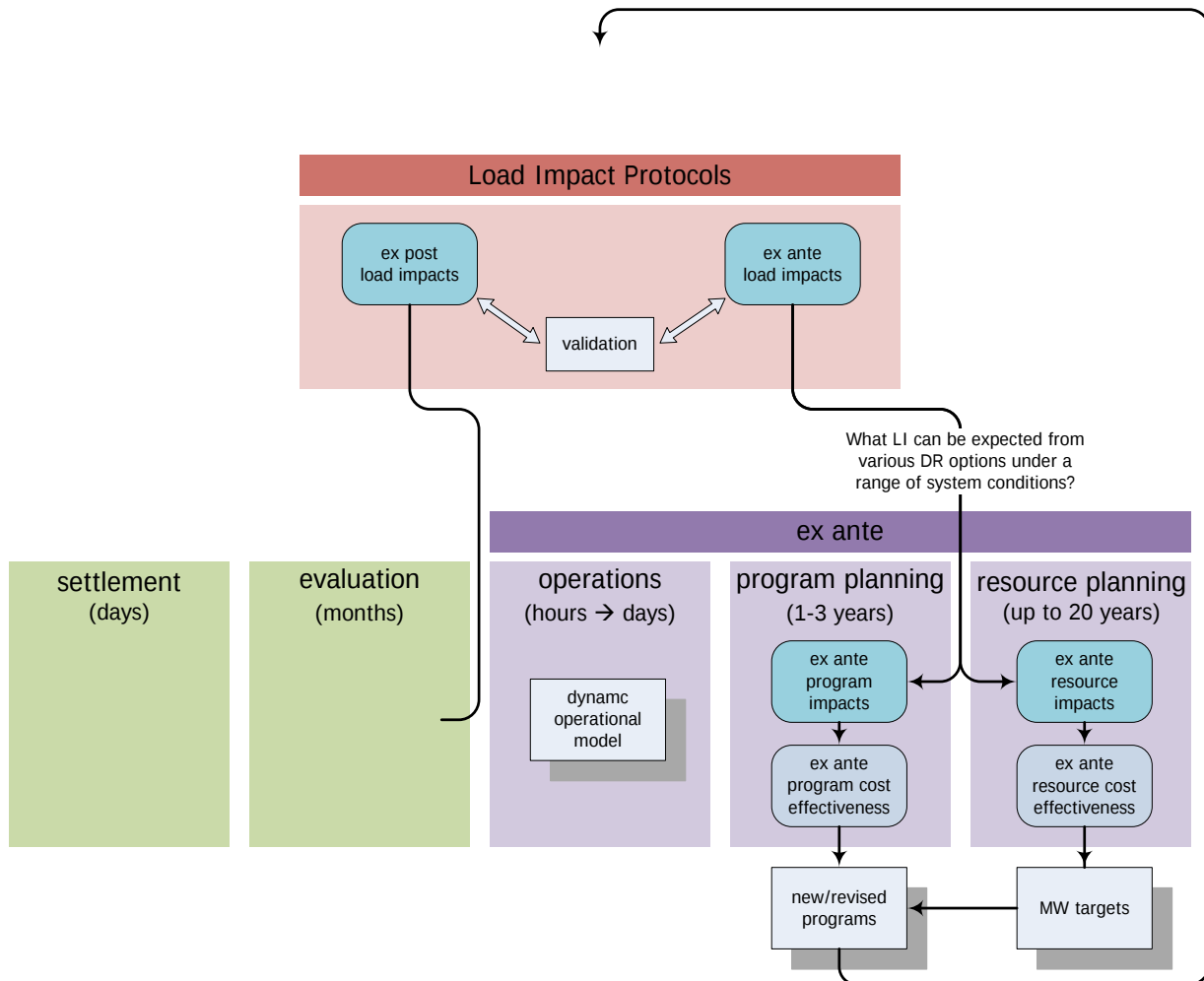
Although load impact estimation is in itself an important tool for determining if a program met its targets, its usefulness extends much further. As is demonstrated in Figure A-1, load impact estimation informs a variety of questions that arise throughout the life cycle of a DR resource—including evaluation, operations, settlement, program planning and resource planning.

The protocols in this document distinguish between **ex post** and **ex ante** load impact estimation (see Figure A-1). This distinction is important.

- **Ex post load impacts are reflective:** They describe the past impact of an existing resource option. In other words, they quantify the demand reduction that occurred during a defined historical period, under the conditions that were in effect during that time. Because ex post performance is tied to past conditions, such as weather, price levels or system conditions that determine the extent to which resources are needed at the time, the impacts may not reflect the full value of the DR resource. As such, it would be inappropriate to use ex post impacts to determine DR program cost-effectiveness. Ex post load impact estimation should be viewed primarily as a means to an end—it is an important step in developing and validating ex ante impact estimates, which are described below.
- **Ex ante load impacts are forward-looking:** They describe the expected load impact under a range of potential conditions of interest, such as extreme weather, high prices or increased program participation. Ex ante load impacts are an important input to DR cost-effectiveness analysis, both for program planning (comparing different DR program designs) and resource planning (comparing archetypal DR options against other conservation and supply resources). For these purposes, ex ante impacts should be based on the extreme conditions for

which the system is designed (e.g., 1-in-10 weather year conditions) in order to capture the full option (or insurance) value of DR. Program planners may also wish to estimate ex ante load impacts for average conditions (or other conditions of interest), particularly for non-event-based DR options such as time-of-use rates, as understanding impacts on a day-to-day basis can be a useful input to calculating lost revenue, identifying structural beneficiaries, or answering other questions that arise in program design and planning.

Figure A-1
Context for DR Load Impact Protocols



Although the focus of these protocols is on ex post evaluation and ex ante estimation for program and resource planning, load impact estimation can also contribute to program operation. Ex ante load impact estimates can form the basis for developing dispatch models to predict, in near-real time, the impact of DR programs as they are activated.

DR program settlement can also be informed by load impact estimation, although these protocols do not provide direct guidance to do so. Nonetheless, similar methods and data

sources can be used to determine how accurately a program's settlement methodology compensates participants for their demand reductions. Going a step further, it may be possible to develop dynamic settlement models based on the same statistical methods (e.g., regression analysis) that are typically used for ex ante impact estimation.

A.3. *HOW TO USE THIS DOCUMENT*

In addition to presenting the load impact protocols described above, this document is intended to provide the necessary background to understand and implement the protocols, as well as provide guidance on related evaluation issues that go above and beyond the minimum requirements.

The intended audience for this document includes:

- evaluation contractors performing DR impact evaluations;
- OPA staff responsible for managing DR program evaluations;
- OPA staff responsible for DR program design and resource planning;
- staff of other organizations that may administer future DR programs/tariffs; and
- other stakeholders with an interest in DR program evaluation.

The remainder of this document is organized as follows:

- **Section B** provides an overview of **issues and methods for estimating DR load impacts**. Evaluation managers and others tasked with understanding the results of DR load impact estimates should find useful background in this section. In addition, this section describes optional elements that evaluators may wish to consider including in specific DR program evaluations.
- The **load impact protocols** are presented in **section C**. Below is a roadmap to the topics covered by the protocols and where they can be found:

Protocol 1—Evaluation Planning	Page 25
Protocol 2—Time Periods	Page 27
Protocol 3—Reporting Format	Page 29
Protocol 4—Day Types and Event Conditions	Page 31
Protocol 5—Portfolio Analysis	Page 33
Protocol 6—Statistical Reporting and Validation	Page 33
Protocol 7—Analysis Based on Sampling	Page 34
Protocol 8—Reporting and Documentation	Page 36

- A **Glossary of Terms** is provided for reference (beginning on **page 40**).
- Finally, additional technical information on the **Criteria for Developing Good Impact Estimates, Sampling Issues, Evaluation Methods and Regression Analysis** is provided in **Appendices 1 through 4**. This should be helpful to

INTRODUCTION

individuals tasked with designing and/or implementing DR load impact estimation studies.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

B. OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

The load impact protocols set forth in Section C are meant to establish minimum requirements for DR load impact evaluations—they focus on what to provide, rather than how to provide it. This is because there is a great degree of variation in the potential types of DR resource options, stakeholder interests, uses for load impact estimates, and other factors, that make it impractical to dictate precise requirements to meet all possible conditions. More importantly, it would be inappropriate to do so, as factors such as the availability of resources, the magnitude of the DR resource (e.g., whether it is a large or small program), and available data, should also be considered when planning a load impact evaluation.

Accurately estimating load impacts associated with DR resources is a challenging exercise. By their very nature, DR resource impacts, and the need for them, are dynamic—load impacts vary with key drivers such as weather, prices and other factors, while system conditions drive the need for them. Customer behaviour is also a significant determinant of DR resource impacts, one that varies across customers and by time of day, day of week and season. The likelihood that a DR resource will be available when it is needed is also influenced by the magnitude and nature of the financial incentives provided (e.g., the magnitude of the price signal, whether non-performance penalties apply, etc.). Finally, different parties and stakeholders have different interests in understanding how DR load impacts vary across individual customers or customer segments, across geographical locations, and across system conditions. All of these factors influence the way in which load impact evaluations should be conducted and what they should provide.

The protocols presented in Section C were designed with sufficient flexibility to achieve the key objectives of minimizing excess evaluation burden and costs while providing valuable input to resource planning and program design. While this leaves room for discretion in designing specific program evaluations, it also means that evaluators must have a solid understanding of key issues, methods and objectives. The goal of this report section is to provide an overview of key issues that should be considered during evaluation planning and that must be addressed when conducting DR impact evaluations. Impact evaluation methods that can be used to address these issues are discussed in Appendices 2 through 4.

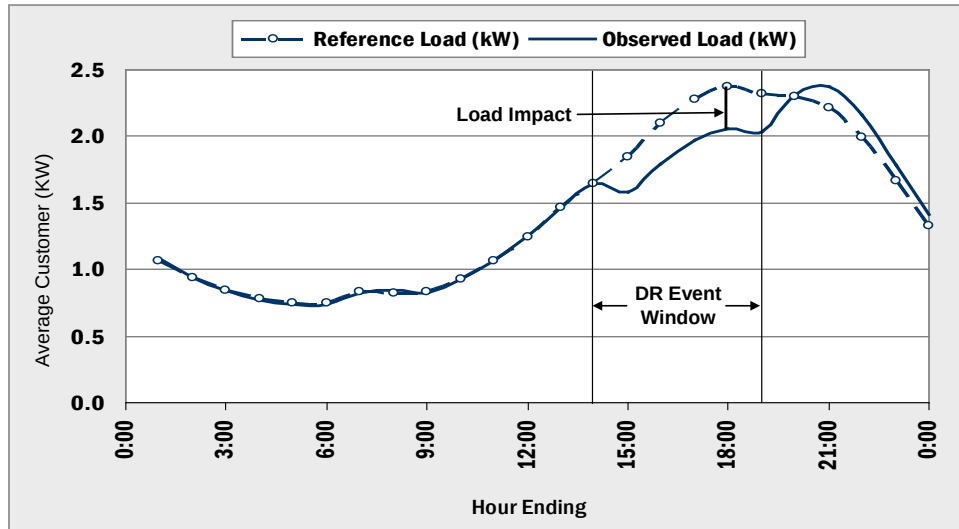
This section begins with a brief discussion of what a “load impact” is and a word of caution on how certain approaches can introduce bias into load impact estimates. This is followed by a discussion of the characteristics of different types of DR resources that influence the type of output required by the protocols. The remainder of the section provides a brief discussion of numerous issues that evaluators and users of impact estimates must address during evaluation planning and when using the results for program design, cost-effectiveness analysis and resource planning.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

B.1. DEFINING LOAD IMPACTS

Load impacts associated with DR resources are defined as ***the difference between a customer's actual (observed) electricity demand, and the amount of electricity the customer would have demanded in the absence of the DR program incentive.*** The latter cannot be observed and must be estimated. This estimate is referred to as the **reference load**. Figure B-1 illustrates the load impacts associated with an event-based DR resource.⁶

Figure B-1
Load Impacts for Demand Response Resources



As seen in Figure B-1, load impacts for an event-based resource can occur not just during the event window (i.e., the period of time between when an event is triggered and when it is stopped), but also during the hours leading up to or following an event window. For example, with critical peak pricing, a residential participant with air conditioning might pre-cool the house before an event window begins, thus increasing load for an hour or two before the event starts relative to normal usage on a non-event day. Similarly, for direct load control, load following the end of an event window is often higher than it otherwise would have been, as air conditioners cycle more frequently once cycling control ceases in order to return the house to its normal temperature setting.

To account for these pre- and post-event impacts, the load impact protocols in Section C require that impact estimates be provided for all hours of an event day (not just during the event window itself), and similarly, for all hours of a typical day in which non-event based (e.g., time-of-use rates or permanent load-shifting) resources are in effect.

⁶ Event-based DR resources are triggered by the occurrence of pre-defined “event” conditions, such as system emergencies or supply resource constraints. Event-based resources are distinguished from “continuous” DR resources, such as time-of-use (TOU) rates or permanent load shifting programs (e.g., OPA’s DR-2), that effectively reduce peak demand all (or most) of the time.

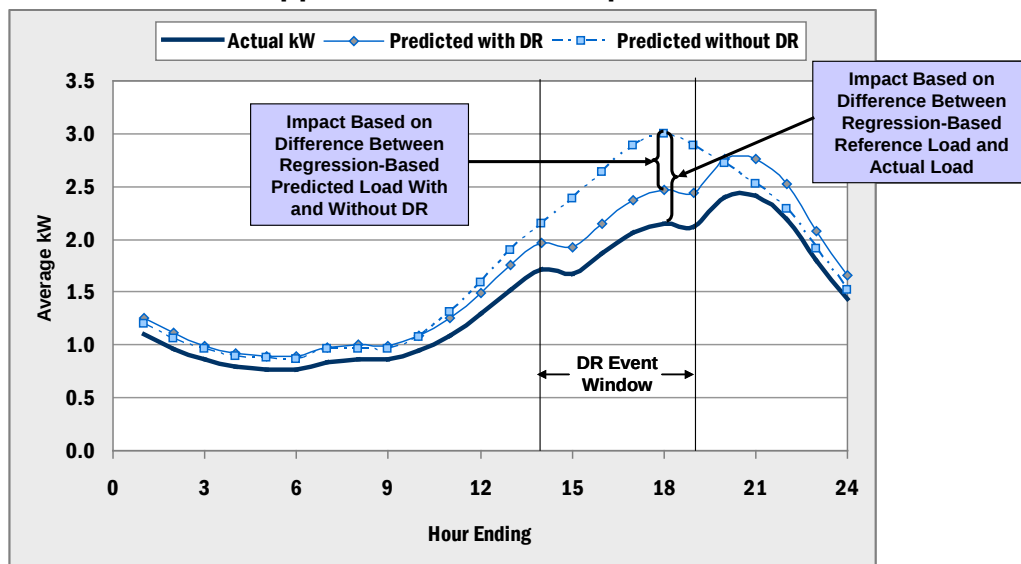
OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

As stated above, load impacts equal the difference between what a customer actually used in the presence of a DR incentive, which can be observed for customers with interval meters, and an estimate of what they would otherwise have used. Given this definition, it would seem logical to estimate load impacts as the difference between observed load and the estimated reference load. However, this approach will not necessarily produce the most accurate estimate of demand response.

A more accurate estimate may result from taking the difference between two estimated values—predicted load without DR in effect and predicted load with DR in effect. This is because the difference between observed load and predicted load is a function not only of the actual load reduction, but also of any error in the predicted reference load. The model used to estimate the reference load could do a very good job of tracking the pattern of usage for a day, and also a good job determining the magnitude of the load drop, but might be biased upward or downward in its prediction of the reference load. Under these circumstances, it would be more accurate to estimate the load impact as the difference between two predicted values—one with DR in effect, and one without DR in effect.

The two different approaches are illustrated with an example in Figure B-2. The regression model over-predicts customer load during the DR event relative to the actual (measured) load in all hours—that is, it is biased upward. As such, estimating load impacts as the difference between the observed load and the reference load predicted by the regression model produces a biased load impact estimate that overstates the actual impacts. Under these circumstances, estimating the impact as the difference between predicted load with and without DR in effect would be less biased. As indicated in Section C, either approach to estimating load impacts may be suitable depending on circumstances and it is up to the evaluator to decide which approach is preferable in each instance and to produce evidence in support of the chosen method.

Figure B-2
Two Approaches to Load Impact Estimation



OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

B.2. *CHARACTERISTICS OF DEMAND RESPONSE RESOURCES*

As DR resource options have proliferated and gained increasingly widespread adoption throughout North America, various parties have attempted to categorize them. Depending on the context, a variety of DR resource characteristics have been adopted for this purpose. For example, as mentioned in the previous section, DR resources may be considered “event-based”—that is, tied to a specific trigger that “calls” the resource when it is needed—or “continuous”—in which response is not tied to a specific trigger but may nevertheless fluctuate with exogenous factors such as weather or prices. Another common categorization is as emergency (reliability) or economic resources—in this case, the triggering conditions (e.g., system emergencies, fluctuations in wholesale market prices) are of interest. DR options may also be categorized as “incentive-based”—in which participants receive payments (typically funded by ratepayers) in return for reducing load—or “price-based”—in which customers respond to time-differentiated prices built into retail electricity rates.

For the purposes of defining analysis and output requirements for load impact estimation, it is useful to categorize DR resources according to the following design characteristics:

- **Frequency of Use:** Some resource options, typically emergency programs like interruptible rates and direct load control, are triggered relatively infrequently. At the other end of the scale are continuous DR resources, such as time-of-use (TOU) and real-time pricing (RTP) rates. Other programs may be called several times a season. For event-based resources, the frequency of use is often tied to program rules that set a maximum number of events per season or other time period.
- **Event Timing and Duration:** For some resource options, such as critical peak pricing, the event window—the event start time and number of hours of duration—is fixed (e.g., noon to 6 p.m.). For other resource options, such as direct load control, both the start time and event duration can be highly variable. Of course, this characteristic is only applicable to event-based DR resources.
- **Number of Participants Called:** For some resource options, it is typical that all program participants are called for every event. For others, flexibility exists to only call program participants in specific locations to provide emergency relief of transmission or distribution system contingencies.

For purposes of these protocols, the above characteristics can be combined to produce the following four DR resource categories:

- **Limited Frequency Resources:** Characterized by very infrequent use, highly variable event timing and duration, and high variability in the number of participants that are called for each event. This option is represented by emergency resources such as interruptible rates and emergency-driven direct load

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

control. OPA's **peaksaver**⁷ program is an example of this type of resource option.

- **Limited Variation Resources:** Typically has the same event window each day and little or no variation in the number of participants that are called for each event. This type of resource is characterized by low to moderate frequency of use. Examples of this resource option include most critical peak pricing tariffs.
- **High Frequency Resources:** Characterized by a very high number of event days and highly variable event timing and duration. The number of customers participating in each event is largely at the discretion of the participants themselves and may vary significantly with market conditions (e.g., more will participate on very high priced days than on other days). OPA's DR-1 program is an example of this resource option.
- **Continuous Use Resources:** This category includes all non-event based resources, such as RTP, TOU and permanent load shifting. OPA's DR-2 program is an example of this option.

The preferred analysis methods for impact evaluation and the outputs of interest vary across the above categories. For low frequency resources, it is not uncommon that no events are called in a given year. As such, any requirement to provide ex post load impact estimates would not be applicable for that year. Furthermore, with very infrequent events, it may be difficult to develop models to predict ex ante load impacts based on actual event data—in these circumstances, it may be necessary to use end-use metering on a sample of participants that are intentionally triggered under event-like conditions in order to develop statistically rigorous ex ante impact estimates.

For limited variation and high frequency resource options, it is much easier to develop sound, statistical estimates of impacts under various conditions. Since events are triggered on some days and not others, empirical estimates can be developed using relatively small participant samples and without the need for pre-enrolment customer load data. Instead, each participant acts as its own control—data from “event-like” days on which events are not called is a good proxy for the reference load on event days. On the other hand, with so many event days in any given year, a requirement to report load impacts on every historical event day would pose a significant burden.

Continuous-use resources present a significant empirical challenge in that one must use a control group, pre-enrolment data, or another approach such as elasticity estimation to determine load impacts, since there are no “event-like” days to use as a reference load proxy. One must also decide which day types are of greatest interest from a reporting standpoint unless impacts are going to be produced for all hours of the year.

⁷ ® Trademark of Toronto Hydro Corporation. Used under license.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

The load impact protocols in Section C, in particular protocol 4, try to strike a balance between the need for comparability across resource options, the need to develop estimates that best support cost-effectiveness analysis, the need to report information that allows reviewers to assess the validity of the impact estimates, and the desire to minimize the analysis and reporting burden by not requiring that estimates be developed for every possible day type or event condition for which each resource might be used.

B.3. *Ex Post Versus Ex Ante Impact Estimation*

As introduced in Section A.2, load impacts can be estimated on an **ex post** or **ex ante** basis. Ex post load impacts are estimates of demand reductions under conditions that actually occurred in the past for the specific set of customers that were called on an event day or were enrolled in a non-event-based program on a selected day. Ex ante load impact estimates represent expected demand reductions under a set of predefined conditions for a specified group of customers.

As indicated in Section A.2, the primary objective of the load impact protocols presented here is to develop estimates that can support program and resource planning, which are inherently forward-looking, or ex ante, exercises. Resource planning seeks to identify the optimal combination of resources that will balance supply and demand at least cost under a specified set of conditions. Program planning involves comparing the cost-effectiveness of different potential resource options, also under a specified set of conditions. While ex post load impact estimates may have some interest to resource planners and program managers, for reasons discussed below, they typically should not be used for cost-effectiveness analysis or for long-term resource planning. Ex post impact estimation is primarily a means to an end—it is essential for determining the relationship between demand response and key explanatory variables, which can then be used to estimate ex ante impacts.

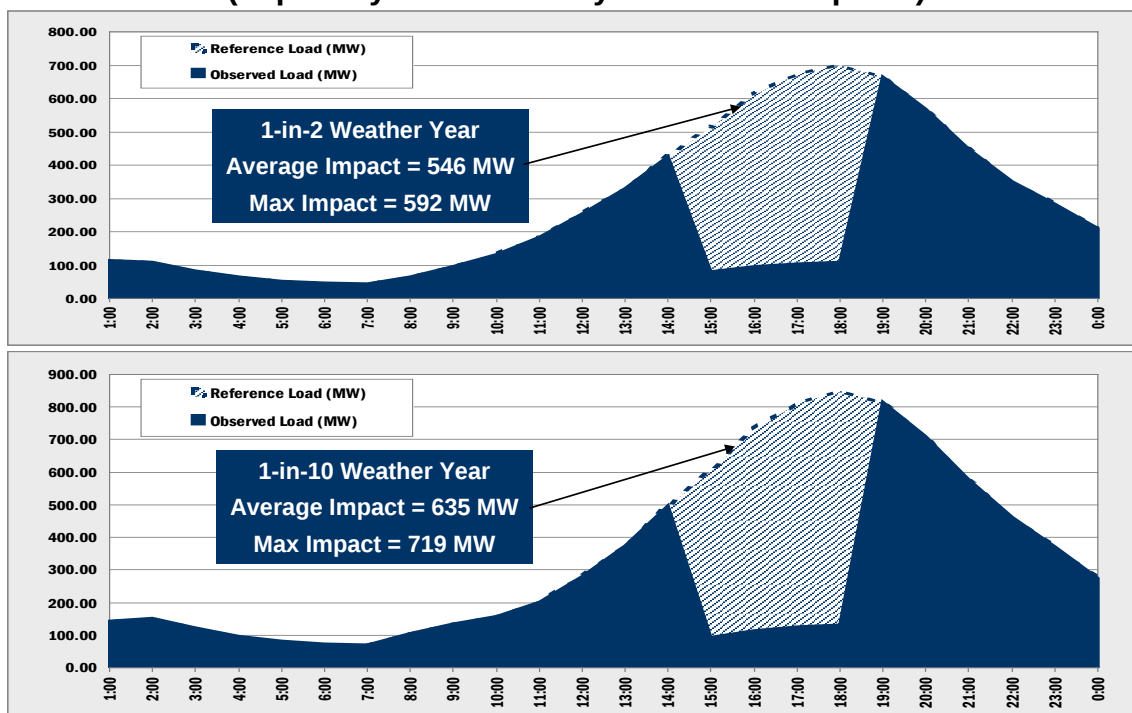
Having said that, it should also be noted that the desire to reflect changing customer characteristics and event conditions in ex ante load impact estimates may impose certain model specification and data requirements on the evaluation process that would not be necessary in order to develop accurate ex post estimates. If the only interest is in knowing what load impacts were during a specific historical time period, simpler estimation methods might not only be suitable, they might produce more accurate or at least more easily understood estimates than would the more complex methods required to produce reasonable ex ante forecasts.

Ex ante load impact estimates for event-based resources should be developed based on the conditions under which the resource is most likely to be called. Quite often, DR resources provide larger load reductions on these high probability days than they would if they were called on other days. For example, in jurisdictions such as Ontario where air conditioning drives system peak demand, an air conditioning cycling program is more likely to be called on hot days when the system is stressed than on cooler days when reserve margins are much larger. Moreover, on hot days, the load reduction potential from air conditioner cycling is significantly greater than on cool days. Similarly, load impacts for a year in which the weather is hotter than normal would be greater on average than for a normal weather year.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

For example, Figure B-3 shows the difference in the aggregate impact across the 20 highest system loads days for Southern California Edison's air conditioner cycling program, based on normal (1-in-2 year) weather conditions and on more extreme (1-in-10 year) weather conditions.⁸ The average impact across the top 20 system load days is more than 16 percent higher under 1-in-10 year weather conditions than under normal weather conditions. The difference during the peak event hour is roughly 21 percent (719 MW compared with 592 MW).

Figure B-3
DR Load Impacts by Weather Year for Southern California Edison's
Air Conditioner Cycling Program
(Top 20 System Load Days for All Participants)



Short of requiring load impacts for every hour of each forecast year over the desired planning horizon, a challenging issue in developing load impact protocols is deciding on the conditions and time periods for which load impact estimates should be provided. A key concept that should guide these decisions is that many DR resources are like options or insurance—they provide value even when they are not used. Emergency resources provide insurance or protection against typically low probability, high cost events such as brownouts and blackouts. Economic resources can provide protection against high wholesale market clearing prices by reducing demand when prices rise and, therefore, allowing the market to equilibrate at lower prices than it otherwise would. Even if a DR

⁸ Stephen S. George, Josh Bode and Josh Schellenberg. *Load Impact Estimates for Southern California Edison's Demand Response Program Portfolio*. September 25, 2008.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

resource is typically operated by calling on just a subset of participants (e.g., air conditioner cycling called in response to local distribution outages), cost-effectiveness analysis should still be based on the full option value of the resource—that is, the magnitude of load reduction that could be achieved if all participants were called simultaneously. In other words, DR resource cost-effectiveness should not be based on the ex post load impacts that were achieved in the past year, or even several years, because the conditions under which those resources were used or the number of participants called may significantly understate the full option value of the resource.

B.4. *CHANGES IN PROGRAM ENROLMENT*

Because ex ante load impacts are forward looking, the issue of program enrolment must be considered. If the enrolment is expected to be more or less static, ex ante load impacts may be estimated for the group of customers currently enrolled in a program, under the conditions of interest (e.g., normal or extreme weather year conditions). If program enrolment is expected to change, either quantitatively (number of participants) or qualitatively (participant size, business activity, end uses, etc.), then these changes must be predicted and the load impacts applied to the expected new participant population to determine overall program impacts and cost-effectiveness.

This presents a significant challenge, as predicting DR program enrolment can be either very difficult, quite subjective or some combination of the two.⁹ Furthermore, since load impacts can vary significantly across customer groups (e.g., average impacts per customer might be quite different for office buildings than for manufacturing plants), determining the average or aggregate load impact for a group of customers that has a significantly different mix of businesses than exists among current enrollees would not only require forecasting the mix of customers in each planning year but also estimating average load impacts for each relevant business type. As such, a key decision that must be made in the planning stages of ex ante load impact estimation is whether the load impacts are required just for the current set of enrollees, or also for a future group of enrollees that might differ significantly from the current mix. If the latter, the evaluation must estimate load impacts at a sufficient level of granularity to support such an analysis—i.e., participant characteristics and other factors that drive variation in load impacts should be identified and load impacts estimated separately for the identified sub-groups. Understanding how impacts vary across customer segments may be important not only as a means of ensuring accuracy as customer populations change, but also as input to developing more effective marketing strategies and tactics and or refining program rules and characteristics.

The primary focus of these protocols with respect to ex ante estimation is on developing estimates for the average enrolled customer that reflect the day types and event conditions needed for program and resource planning. Determining what enrolment will be in future years is considered to be an exogenous exercise. However, the load impact evaluator should be cognizant of how enrolment is expected to change in the future and, as necessary, be prepared to develop models that can reflect the impact of these changes on aggregate impacts.

⁹ To bound the uncertainty in future enrolment, it may be desirable to take a scenario approach.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

B.5. *TIME PERIODS*

When planning an evaluation of DR resources, it is important to consider the granularity of time intervals at which impacts will be reported. The load impact protocols require that impacts for event-based resources be developed for each hour of an event day (in order to capture any adjustments in usage prior to or following an event period). However, some stakeholders may find this level of detail to be unnecessary and may wish only to see estimates for the average impact across an event period. Other stakeholders may wish to see estimates for each half hour or even for each 15-minute interval. The degree of granularity for which load impact estimates are needed can impact both the analysis approach as well as data requirements, and evaluators must balance the availability of data and resources against desired outputs in developing an evaluation plan.

B.6. *DAY TYPES*

As discussed above, DR resources vary significantly with respect to the frequency and duration with which they can be exercised. While some resources are in effect for all 8,760 hours of the year, program rules may limit the availability of others to, for example, only the summer season, or certain hours on weekdays. Some resources are typically only called during system emergencies or on days with very high electricity demand or high prices, while others, such as permanent load shifting, are in effect nearly every day. To reduce the burden of predicting impacts for each hour of the year that a resource may be called, load impacts may be estimated for pre-defined “day types”. The load impact protocols take this approach, dictating the minimum day types for which estimates must be provided, taking into consideration the desire to have common day types across resource options so they can be compared, the desire to recognize the unique characteristics of each resource, and the need to keep the amount of work and evaluation output manageable. In some cases, these minimum output requirements may not meet the interests of all stakeholders. As such, evaluation planners should solicit input from interested parties and decide whether estimates are needed for day types other than those required by the protocols.

Determining the day types for which impacts are reported is particularly challenging when there are correlations between several drivers of demand response and/or it is difficult to predict conditions under which a resource is likely to be triggered because of lack of historical data or likely changes in future conditions. OPA’s DR-1 program is an example of a resource for which it is difficult to develop estimates for day types representing ex ante event conditions. DR-1 is called when prices exceed a certain threshold. Historically, DR-1 has been called very frequently (over 1,000 hours a year), and this high frequency can reduce customer load response relative to what it would be with a lower number of events. Furthermore, high prices are correlated not only with high demand, which in turn, is correlated with hot weather, but also with high natural gas prices, which often occur in winter. For such a program, load impact estimates under normal and extreme weather conditions may have little relevance to what the impacts would be for a normal or extreme program event day and it is difficult even to define what a normal or extreme day is for such a program. Even if data were available to correlate weather, market prices and other factors that drive demand and demand reduction, the correlation between these factors could change in the future as supply conditions become looser or

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

tighter relative to historical conditions or OPA modifies the DR-1 strike price trigger to reduce the number of event hours. This is a particularly perplexing problem for which there is no easy solution.

At a minimum, in such situations, it is imperative that evaluators identify shortcomings of this sort and properly caveat load impacts that meet minimum requirements but that might not capture important correlations or have other limitations due to lack of data or changing conditions. Scenario analysis (e.g., producing impacts under various conditions that reflect all relevant drivers of demand response based on assumed conditions and correlations) is another option to address such shortcomings. Over time, as more historical data become available, the number of resources for which such challenges arise should diminish.

B.7. *EVENT WINDOW*

For programs with variable event timing and duration, deciding on the event window for which ex ante load impacts should be provided is another challenging issue. Even for ex post impact estimates, if there are a large number of events and the timing and duration of events varies significantly, reporting impacts for the “average event” can be somewhat complex. Once again, the protocols describe the event windows for which load impacts must be reported for each program type for both ex post and ex ante estimation, taking into consideration resource planning needs. However, stakeholders may wish to see estimates for other event windows.

B.8. *EXTREME CONDITIONS*

As discussed previously, when examining the cost-effectiveness of DR resources, it is important to estimate the impacts not just under normal conditions, but also under the extreme conditions under which many DR resources are designed to be used and provide their highest value. This means that impact estimation methods must be able to predict DR load impacts under conditions that may not have actually occurred during the historical period from which the data for estimating the impacts is obtained. In these instances, it is important, for example, to pay careful attention to the functional form of the models being developed, making sure to examine non-linearities in the relationship between load impacts and key drivers. For example, the relationship between weather and energy use is typically non-linear—above certain temperatures, air conditioners are already running at full capacity and additional increases in temperature beyond that point do not increase energy use. Similarly, the relationship between price and energy use may be non-linear. For example, across a reasonable range of price changes, the percentage change in energy use given a percentage change in price (e.g., the price elasticity) may be relatively constant. However, above a certain threshold, the same percentage increase in prices may elicit a much lower percentage change in energy use as all largely discretionary energy use has already been eliminated.

The load impact protocols presented in Section C require that estimates be developed based on both normal and extreme weather year conditions. There are various ways to

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

define normal and extreme weather conditions for ex ante load impact estimation.¹⁰ OPA is currently working on developing these definitions.

B.9. UNCERTAINTY

Load impact estimation is inherently uncertain, as impacts are based in part on estimated values. If regression analysis is used to develop ex post impact estimates, uncertainty can be expressed in terms of the standard errors of the impact coefficients, the standard error of the forecast for energy demand, or both. For ex ante impact estimates, not only is there uncertainty in terms of model parameters, but also in the magnitude of the key drivers of demand and demand response, such as weather.

Uncertainty-adjusted load impact estimates can be developed based on standard errors of the model coefficients as reported by any statistical software used for regression analysis, or based on Monte Carlo simulation of model parameters and key drivers. The California Public Utilities Commission's (CPUC) load impact protocols¹¹ require that uncertainty-adjusted impact estimates be provided for the 10th, 30th, 50th, 70th and 90th percentiles for each hour of each required event day. Table B-1 contains an example of what these requirements produce for each DR resource and event day. While this enhances understanding of the uncertainty surrounding the impact estimates, it does add to the burden of load impact estimation and to the proliferation of output provided.

The load impact protocols in Section C do not require that uncertainty adjusted impacts such as those shown in Table B-1 be reported as part of the minimum requirements for program evaluation. However, as some stakeholders may desire such information, the need for developing uncertainty-adjusted impact estimates should be determined as part of evaluation planning.

¹⁰ In recent work in California, proxy 1-in-2 and 1-in-10 weather years were selected by examining weather conditions over the last 30 years and ranking each year based on cooling degree days during the summer. The 1-in-2 and 1-in-10 weather years were chosen as the years representing the 50th and 90th percentiles from this distribution. However, choosing a single year to represent the weather can lead to anomalies when comparing single days across the years. For example, weather on the June system peak day can easily be cooler in the proxy year for 1-in-10 weather conditions than for 1-in-2 weather conditions. An alternative approach would be to use the average weather across several years surrounding the 50th and 90th percentile years. This would not guarantee that there will be no anomalies across the two years, but it should reduce the likelihood that this will be the case. Another alternative would be to construct stylized years that ensure that such anomalies don't exist (e.g., by taking the higher of the June days to represent the 1-in-10 weather year) or by selecting the 50th and 90th percentile years for each month, rather than for the year as a whole.

¹¹ *Load Impact Estimation for Demand Response: Protocols and Regulatory Guidance*, California Public Utilities Commission, March 2008.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

Table B-1
Uncertainty Adjusted Load Impact Estimates for PG&E's SmartRate Program
Average Residential Participant, July 8, 2008 Event Day

Hour Ending	Reference Load (kW)	Observed Load (kW)	Load Impact (kW)	%Load Reduction	Weighted Temp (F)	Uncertainty Adjusted Impact - Percentiles				
						10th	30th	50th	70th	90th
1:00	1.31	1.36	-0.05	-3.63%	85.5	-0.08	-0.06	-0.05	-0.04	-0.02
2:00	1.15	1.21	-0.06	-5.58%	84.5	-0.09	-0.08	-0.06	-0.05	-0.03
3:00	1.03	1.09	-0.06	-5.43%	83.0	-0.09	-0.07	-0.06	-0.04	-0.03
4:00	1.00	1.04	-0.04	-3.51%	82.5	-0.06	-0.05	-0.04	-0.02	-0.01
5:00	0.95	0.95	0.00	-0.33%	81.0	-0.03	-0.02	0.00	0.01	0.03
6:00	0.88	0.91	-0.04	-4.10%	79.0	-0.07	-0.05	-0.04	-0.02	-0.01
7:00	1.03	1.03	0.00	0.17%	81.0	-0.03	-0.01	0.00	0.01	0.03
8:00	1.08	1.13	-0.05	-4.23%	83.5	-0.08	-0.06	-0.05	-0.03	-0.02
9:00	1.13	1.17	-0.03	-2.82%	87.5	-0.06	-0.04	-0.03	-0.02	0.00
10:00	1.25	1.30	-0.05	-3.93%	92.0	-0.08	-0.06	-0.05	-0.04	-0.02
11:00	1.57	1.60	-0.03	-1.92%	97.5	-0.06	-0.04	-0.03	-0.02	0.00
12:00	1.87	1.85	0.01	0.70%	100.0	-0.02	0.00	0.01	0.03	0.04
13:00	2.20	2.17	0.03	1.42%	103.0	0.00	0.02	0.03	0.04	0.06
14:00	2.40	2.27	0.13	5.25%	104.5	0.10	0.11	0.13	0.14	0.16
15:00	2.58	2.08	0.50	19.32%	106.0	0.47	0.49	0.50	0.51	0.53
16:00	2.76	2.25	0.51	18.46%	106.5	0.48	0.50	0.51	0.52	0.54
17:00	3.11	2.48	0.62	20.07%	108.0	0.59	0.61	0.62	0.64	0.65
18:00	3.20	2.53	0.67	20.87%	107.0	0.64	0.65	0.67	0.68	0.70
19:00	3.03	2.51	0.52	17.17%	105.5	0.49	0.51	0.52	0.53	0.55
20:00	2.89	2.89	0.00	0.13%	103.5	-0.03	-0.01	0.00	0.02	0.03
21:00	2.72	2.87	-0.15	-5.55%	100.5	-0.18	-0.16	-0.15	-0.14	-0.12
22:00	2.48	2.68	-0.20	-8.00%	97.5	-0.23	-0.21	-0.20	-0.19	-0.17
23:00	2.14	2.30	-0.17	-7.73%	95.0	-0.19	-0.18	-0.17	-0.15	-0.14
0:00	1.77	1.89	-0.12	-6.79%	92.0	-0.15	-0.13	-0.12	-0.11	-0.09
Daily	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	% Daily Load Reduction	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
						10th	30th	50th	70th	90th
Daily	45.53	43.57	1.96	4.31%	130.9	1.96	1.96	1.96	1.96	1.97

Source: Stephen S. George and Josh Bode. *2008 Ex Post Load Impact Evaluation for Pacific Gas and Electric Company's SmartRate™ Tariff*. Prepared for PG&E. December 30, 2008.

B.10. LOCATION SPECIFIC IMPACTS

For a variety of reasons, both the magnitude and benefits of DR may vary significantly across geographic regions. The magnitude of load reductions may vary geographically due to differences in the underlying population and climate (e.g., air conditioning saturation, a key driver of demand response for residential customers, may vary across different regions of Ontario due to climate and socio-economic factors). Differences in the mix of commercial and industrial customers, and differences in marketing effectiveness and program characteristics across local distribution companies (LDCs), may also influence load impacts geographically.

The benefits, or avoided cost, associated with DR may also differ geographically. The OPA has identified a number of local constrained areas (LCAs) within Ontario; regions

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

with transmission or distribution capacity constraints, with excess generation on one side of the constraint and supply shortages on the other. DR resources have greater value in areas with generation needs than in those with excess generation. Furthermore, if DR can defer the need to invest in new or upgraded supply, the time value of deferring the investment may also be counted as a benefit of DR.

For these and other reasons, it is important that the evaluation planner consider the need to develop load impacts and cost-effectiveness estimates for selected geographic regions, not just for the Province as a whole. The need for geographic-specific impact estimates could significantly affect data requirements, the analysis approach, or both. However, to produce impact estimates for every possible region of interest, it is not necessary to develop load impact estimates based on samples from each geographic region. All that is necessary is to know how many DR program participants are in the region of interest and to combine data on the key drivers of demand response with a model that predicts load impacts based on those key drivers.

For example, in recent work in California, the load impacts associated with an air conditioner cycling program were produced by Local Capacity Areas (transmission-constrained regions) using load impact estimates developed by climate zone, a key driver of demand response. Program enrolment was predicted at the Census Block Group (CBG) level and then each CBG was mapped into a climate zone. The LCA impact estimates were developed by calculating an enrolment-weighted average of the climate zone estimates for each LCA.¹²

B.11. CUSTOMER PRICE ELASTICITIES

Price elasticity is a measure of customer responsiveness to price signals through changes in their demand for the good being priced. Elasticities can be used in impact evaluations in two ways: (1) to evaluate load impacts of DR program participants that are simultaneously exposed to market or other time-varying prices, such as Ontario's large electricity consumers who are Market Participants; and (2) to predict load impacts from participants in continuous use resources such as TOU and RTP rates.

In both instances, elasticities can be used to estimate reference loads that account for price fluctuations. In the first instance, if DR resources are called on high-price days, and price-induced demand fluctuations are not factored into the reference load, DR load impacts could be overstated. In the second instance, where TOU or RTP price incentives are in effect all the time, there are no data for event-like days with which to estimate reference loads. If load data exists for a sufficiently long time period prior to customer enrolment in the DR program, these data could be used to develop reference loads. However, where such information is not available, elasticity estimates can be used to determine what the load shape would be in the absence of the RTP or TOU that is in

¹² Stephen S. George, Mathew G. Mercurio and Josh Bode. *Ex Ante Load Impact Estimates for Pacific Gas and Electric Company's SmartAC™ Program for 2009-2010*. April 1, 2009.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

effect. This approach has been used in recent evaluations of RTP and TOU tariffs in California.¹³

B.12. *FREE RIDERS AND STRUCTURAL BENEFITERS*

With energy efficiency program impact estimation, free riders are defined as those program participants who would have implemented a measure in the absence of the energy efficiency program's influence. Determining what participants would have done in the absence of the program—that is, sorting out the difference between gross impacts and net impacts—is a key element of energy efficiency program evaluation. However, it is not very relevant to impact estimation for most DR resources as few customers would reduce load during DR events in the absence of the stimulus provided by the DR resource. Put another way, customers don't typically "do DR" in the absence of a DR program.

On the other hand, there is an issue that many think of as free-ridership that is relevant to DR impact estimation. This issue concerns customers who, by simply enrolling in a DR program, are better off than if they had not participated due to the nature of their load profile.¹⁴ These customers are often referred to as structural benefiter. An example of a structural benefiter is a customer who volunteers for a critical-peak pricing tariff but who does not have air conditioning or who typically does not use air conditioning during peak periods. If a critical peak price is designed to be revenue neutral relative to current average prices, customers without air conditioning are more likely to be structural benefiter than are those with air conditioning.

Some would argue that enrollment by structural benefiter should be encouraged because it would reduce historical cross-subsidies inherent in average cost pricing. However, others believe that the existence of structural benefiter on a time-varying rate means that incentive payments (in the form of lower electricity bills) will be larger than required to achieve the same level of demand response or, worse, that structural benefiter will not provide any demand response benefits at all. These stakeholders may be interested in estimating the number of structural benefiter participating in a DR resource option.

When deciding whether or not to determine the number of structural benefiter that might be participating in a DR program or tariff, it is important to keep the following in mind:

- First, it is not necessary to estimate the number of structural benefiter in order to obtain unbiased estimates of demand response. For event-based programs, using participants as their own control will automatically produced unbiased impact estimates for the participant population. For non-event based programs, careful attention to selection of an external control group (or use of pre-treatment data, if

¹³ Analysis of Southern California Edison Company's RTP program is presented in Stephen S. George and Josh Bode. *Load Impact Estimates for SCE's Demand Response Programs: Residential and Commercial Summer Discount Plan; Agricultural and Pumping Interruptible Program; Real Time Pricing; Optional Binding Mandatory Curtailment*. April 1, 2009.

¹⁴ This issue arises primarily, if not exclusively, with tariff-based DR options such as time-varying pricing.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

available, to estimate reference loads) will also generate unbiased impact estimates.

- Second, just because a participant's usage pattern might produce a windfall gain from participating in a DR program does not mean that that person will not reduce their energy use during peak periods. Structural benefitters and structural non-benefitters face the same marginal price signal or incentive and, in theory, should respond in the same manner to those economic incentives. Indeed, any attempt to eliminate structural benefitters from participating in DR programs could lead to lower participation rates and lower overall demand response since structural benefitters (assuming they can self-identify as such) are logically more inclined to enroll than other customers.
- Finally, it is important to keep in mind that estimating the number of structural benefitters can require an entirely different approach to impact estimation. Estimating the average or total response, without regard for structural benefitters, can be accomplished using regression methods that involve a single equation estimated from data pooled across customers and over time. To estimate the number of structural benefitters, it would be necessary to estimate individual regression equations for every customer using just the time series data available on each customer. This is a lot of additional work and it is important to assess whether the added information is worth the effort.

B.13. *DISTRIBUTIONAL IMPACTS*

In addition to (or as an alternative to) knowing whether structural benefitters and structural non-benefitters respond differently to DR resource incentives, some stakeholders may want to know how much variation there is in impacts across the full spectrum of enrolled customers. For example, there may be interest in knowing what percent of enrolled customers provide load reductions equal to or greater than selected amounts or to know, for example, that 20 percent of customers provide 80 percent of the total load reduction for a program.

Table B-2 provides an example of the distribution of impacts across customers enrolled in Pacific Gas & Electric Company's SmartRate critical peak pricing program. Roughly 30 percent of customers do not reduce demand at all for the average event day, whereas another 30 percent reduce peak demand by more than 30 percent. This type of information can be very useful for targeted marketing if it is possible to identify the characteristics of high and low responders. However, as with identifying structural benefitters, this can require substantially more analytical work than is needed to develop average impacts based on aggregate data or pooled regressions.

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

Table B-2
Distribution of Customer Response
Pacific Gas & Electric Company's SmartRate Tariff
Summer 2008

SmartDay Date	Share of accounts providing load reductions greater than...					
	0%	10%	20%	30%	40%	50%
7/8/2008	74.7	61.7	49.4	38.5	31.9	26.0
7/8/2008	74.1	62.5	46.8	35.4	28.9	24.0
7/8/2008	76.5	63.6	50.2	38.2	30.1	25.0
8/8/2008	65.3	47.2	34.4	27.9	21.5	16.6
8/8/2008	74.1	55.5	38.1	27.6	22.4	17.5
8/8/2008	75.7	60.9	42.4	31.3	23.7	18.8
9/8/2008	65.7	50.8	38.1	30.5	23.8	18.6
9/8/2008	69.2	52.8	37.8	28.7	21.7	17.1
9/8/2008	70.1	53.1	39.3	28.9	22.0	17.2
Average Event	70.8	54.7	39.8	30.3	23.6	18.7

Source: Stephen George and Josh Bode. *2008 Ex Post Load Impact Evaluation for Pacific Gas and Electric Company's SmartRate™ Tariff*. Prepared for PG&E. December 20, 2008

B.14. PERSISTENCE AND LONG-TERM IMPACTS

Another important difference between DR and energy efficiency is that load impacts for most DR resources do not persist beyond the life of the program. If customers no longer receive DR program signals or incentives, they will usually stop providing demand response.

However, the level or extent of the impact associated with a DR resource may change over time even if incentives continue to be offered. For example, some argue that impacts estimated from the first year or two of a price-based DR program may diminish as customers tire of shifting their usage patterns or realize that the sacrifices made are not commensurate with the payments for load reduction or load shifting. Others argue just the opposite—that impacts will increase over time as customers find new ways of responding to DR incentives or invest in enabling technology to automate and expand DR. To date, there is little empirical evidence to settle this debate, as many DR resource options have not been in place long enough to observe whether load impacts change over time. This is why it is important that OPA and others in Ontario conduct impact evaluations of the same program for a number of years, to develop the data needed to address this important question. It is also important to monitor load impact evaluations being done elsewhere as it may be that programs similar to those of interest in Ontario have been in place for some time in other jurisdictions and can be used as a guide to any potential changes in average load impacts over time.

B.15. UNDERSTANDING WHY, NOT JUST WHAT

These protocols focus on the primary objective of impact estimation, determining the magnitude of impacts associated with a wide variety of DR resources. That is, they focus

OVERVIEW OF LOAD IMPACT ESTIMATION ISSUES AND METHODS

on “what” the impacts have been in the past or are expected to be in the future, not on “why” they are what they are. However, for a variety of reasons, it may also be important to understand why the impacts are what they are. If they are larger than expected or desired, it might be useful to ask, “are we lucky or are we good?” If impacts are less than expected or desired, is it because of marketing ineffectiveness, customer inertia, lack of interest, technology failure, or some other reason?

Some of these questions are more relevant to process evaluation than to impact evaluation. Nevertheless, determining whether or not it is important to know the answers could influence the methodology that will be used for impact estimation and/or place additional requirements on the evaluation process in terms of customer surveys, measurement and verification activities, sampling strategy (e.g., stratification, sample size, etc.) and other activities. For example, if the only question of interest is what the load impacts are for the current group of enrolled participants, this can often be done using just a sample of consumers drawn from the current participant group. However, if, for example, the resulting load impact estimates fall short of what was expected, it might be because of selection issues (e.g., customers who don’t use their air conditioner very much might be more likely to participate in an air conditioning load control program than those who use it more). In this situation, obtaining load data from a random sample of the general consumer population or from a sample of consumers who were offered the option to participate but declined, could help determine whether selection bias, or some other factor, is the cause of the lower than expected load impacts.

LOAD IMPACT ESTIMATION PROTOCOLS

C. LOAD IMPACT ESTIMATION PROTOCOLS

This section contains eight protocols that delineate the minimum requirements for estimating the load impacts associated with DR resources in Ontario. They are designed to afford a high degree of flexibility in the analysis methods, while providing standard outputs that can be used as inputs to resource planning, cost-effectiveness analysis and program planning in the Province.

C.1. *PROTOCOL 1—EVALUATION PLANNING*

Determining how best to meet the minimum requirements in these protocols requires careful consideration of methods, data needs, budget and schedule—that is, it requires planning. As such, the first protocol requires development of a formal evaluation plan.

Protocol 1—Evaluation Planning

Prior to conducting a load impact evaluation for a demand response resource in Ontario, an evaluation plan must be produced. The plan must explain in general terms the approach that will be taken to meeting the minimum requirements of these protocols and contain an initial budget estimate and timeline for the analysis. In addition, the evaluation plan must identify which of the following issues will be addressed as part of the impact estimation process:

- 1. Whether the evaluation activity is intended to produce ex post impact estimates, ex ante estimates, or both and, if ex ante impacts are needed, whether they are needed for current enrollees only, future enrollees or both. If ex ante estimates for future enrollees are needed, the plan should outline the changes that are expected to occur in the characteristics of the DR program and/or in the magnitude or characteristics of the participant population, and contain provisions to estimate load impacts at sufficient granularity to support analysis of ex ante changes in enrolment;***
- 2. Whether it is the intent to explicitly incorporate impact persistence into the analysis and, if so, the types of persistence that will be explicitly addressed (e.g., persistence beyond the funded life of the DR resource; changes in average impacts over time due to changes in customer behaviour; changes in average impacts over time due to technology degradation, etc.);***
- 3. Whether it is the intent to develop impact estimates for geographic sub-regions and, if so, what those regions are;***

LOAD IMPACT ESTIMATION PROTOCOLS

- 4. Whether it is the intent to develop impact estimates for sub-hourly intervals and, if so, what those intervals are;***
- 5. Whether it is the intent to develop estimates of the change in daily, monthly and/or annual energy savings associated with a DR resource;***
- 6. Whether it is the intent to develop impact estimates for specific sub-segments of the participant population and, if so, what those sub-segments are;***
- 7. Whether it is the intent to develop impact estimates for event-based resources for specific days (e.g., the day before and/or day after an event), day types (e.g., hotter or cooler days, high price days, etc.) or series of days (e.g., multiple event days in a row) in addition to the minimum day types delineated in these protocols;***
- 8. Whether uncertainty-adjusted load impact estimates will be produced and, if so, whether just the uncertainty in model parameters will be factored into the analysis or, in addition, whether uncertainty in key drivers will also be incorporated;***
- 9. Whether it is the intent to determine not just what the DR resource impacts are, but to also investigate why the estimates are what they are and, if so, what measurement and verification activities will be used to inform this understanding;***
- 10. Whether free riders and/or structural benefiterers are likely to be present among DR resource participants and, if so, whether it is the intent to estimate the number and/or percent of DR resource participants who are structural benefiterers or free riders and/or to determine whether structural benefiterers respond differently than structural non-benefiterers;***
- 11. Whether it is the intent to determine the distribution of impacts across individual customers;***
- 12. Whether a non-participant control group is needed for impact estimation and, if so, what steps will be taken to ensure that use of such a control group will not introduce bias into the impact estimates;***

LOAD IMPACT ESTIMATION PROTOCOLS

13. The target level of confidence and precision in the impact estimates that is being sought from the evaluation effort if sampling is used.¹⁵

C.2. *PROTOCOL 2—TIME PERIODS*

DR programs are designed to influence the timing of energy use, reducing demand through load reduction or load shifting during peak periods or when resources are otherwise in short supply. The objective of load impact estimation is to accurately determine the magnitude and nature of demand response. Protocol 2 indicates that load impacts should be developed for each hour for selected day types and conditions delineated in subsequent protocols. Requiring impact estimates for each hour of a day ensures that the estimates will reflect not just load reduction during peak periods, but load shifting as well. As discussed in Section B.5 and as delineated in Protocol 1, more or less granularity in time periods may be of interest to some stakeholders or decision-makers.

Protocol 2—Time Periods

Load impact estimates shall be provided for each hour of the day for the day-types and conditions delineated in Protocol 4 according to the format delineated in Protocol 3.

C.3. *PROTOCOL 3—REPORTING FORMAT*

As discussed in Section B.1, load impacts equal the difference in energy use in each hour (or sub-hourly period) with and without the influence of a demand response program or incentive. Impacts can be legitimately estimated either as the difference between an estimated value (the reference load) and a measured value (metered load) or as the difference between two estimated values (load with and without the estimated load impact in effect). As discussed in Section B.1, in some instances, the latter may produce a more accurate load impact estimate than the former.

Protocol 3 describes the format in which load impact estimates must be reported for each day type, customer segment and set of event conditions. The column labelled “load without DR in effect” is always an estimated value, as this is unobservable. The column labelled “load with DR in effect” may be either an estimated value or an actual meter read, at the discretion of the evaluator. In either case, the entries in the “load impact” column should equal the difference between the other two columns.

An additional reporting requirement is to include the temperature in each hour. This is meant to provide reviewers with an easily understood indicator

¹⁵ If sampling is not employed, the precision and confidence levels will largely be dictated by the number of customers in the program, the variation in demand response across customers (both of which are outside the control of the evaluator) and the modeling approach that is employed to develop the estimates.

LOAD IMPACT ESTIMATION PROTOCOLS

of the weather conditions that often are a key driver of load and demand response for the day type and conditions underlying the load impact estimate presented in the table. Such information is quite useful when comparing load impacts across day types, as one can easily see whether the estimates in one table or for one program represent a hotter or cooler day than the estimates in another table or for another program. Including this information in the standard reporting table is not meant to imply that temperature is necessarily a key driver of demand response for every resource or that temperature, rather than some other weather variable, is the most appropriate driver of demand response even for weather sensitive loads. For tables in which impacts are provided for an average day (e.g., typical event day), the average temperature across the event days should be provided.

LOAD IMPACT ESTIMATION PROTOCOLS

Protocol 3—Reporting Format

Table C-1 shall be completed for each event day and set of conditions delineated in Protocol 4 for each customer segment determined to be of interest during evaluation planning.

Table C-1 Required Format to Report Load Impact Estimates for DR Resources Program Name: _____ Day Type: _____ Event Conditions: _____ Customer Segment:: _____ Average or Aggregate Load: _____				
Hour Ending at:	Load w/o DR in Effect	Load With DR in Effect	Load Impact	Temperature ¹⁶
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				

¹⁶ The temperature data presented in the table should be weighted by the number of customers participating in the program or event in each climate zone, if temperatures vary across zones. If the table represents multiple days (e.g., the top 10 system load days), the average temperature should also be weighted by the number of customers participating in each climate zone on each day.

LOAD IMPACT ESTIMATION PROTOCOLS

C.4. *PROTOCOL 4—DAY TYPES AND EVENT CONDITIONS*

Protocol 4 identifies four categories of DR resources based on elements of their program design that affect impact evaluation, and identifies the day types and event conditions for which load impact estimates must be provided for each of these four resource categories. The selection of day types and event conditions to be reported for ex post and ex ante estimates balances the need for comparability across resource options, the need to develop estimates that best support cost-effectiveness analysis, the need to report information that allows reviewers to assess the validity of the impact estimates, and the desire to minimize the analysis and reporting burden by not requiring that estimates be developed for every possible day type or event condition for which each resource might be used.

As indicated in Protocol 4, the first step in determining the day types and event conditions that must be provided for a particular evaluation effort is to examine the characteristics of the resource according to the four dimensions discussed in Section B.2 and to classify each resource into the four categories discussed in Section B.2 and summarized in Table C-2. If a DR resource does not fit neatly into one of these four categories, **it is better to err on the side of providing more information rather than less.** For example, OPA's DR-3 program has characteristics of both a limited variation and a high frequency resource—it can be called between 25 and 50 days per year, but events are of a fixed duration and their timing tends to be fairly consistent. In this instance, it would be best to treat DR-3 as a high frequency resource when developing ex ante load impact estimates, thereby providing more detailed impact estimates.

Protocol 4 describes the reporting requirements for both ex post and ex ante impact estimation. It also includes requirements designed primarily to illustrate the validity of the estimation methodology. Estimates for both the average participant and the sum of all enrolled participants are required for ex post impact estimation as well as for ex ante estimation for current enrollees.

As indicated in Table C-3, ex ante impact estimates are required for normal and extreme weather. The requirement to provide impact estimates based on extreme weather conditions is primarily designed to show the option value of DR resources under the conditions when they have a higher likelihood of being called and provide their highest value. However, as discussed in Section B.6, not all resources are weather sensitive and some resources may be triggered under conditions (e.g., high price days) that are hard to predict due to lack of historical data or inherent complexities. In these instances, the evaluator may wish to report impacts for additional day types that more accurately reflect the conditions under which the resource is likely to be called. At a minimum, the evaluator should document the fact that the weather and event conditions for which load impacts are reported under the protocols may not accurately reflect the typical event conditions under which the resource is likely to be used.

LOAD IMPACT ESTIMATION PROTOCOLS

Protocol 4—Day Types and Event Conditions

The information shown in Table C-1 shall be provided for the day-types and event conditions described in Table C-3 after classifying a DR resource according to the characteristics shown in Table C-2. For ex post estimation and for ex ante estimation based on current enrolees, impact estimates must be provided for both the average participant as well as for all enrolled or notified participants (as indicated in Table C-3).

Table C-2 Characteristics of Resource Options				
DR Resource Characteristics	Limited Frequency Resources	Limited Variation Resources	High Frequency Resources	Continuous Use Resources
Frequency of Use	Very Infrequent	Moderate (5 to 25 days per year)	Very High (25 or more days per year)	Continuous (non-event-based resources)
Variation in Event Timing & Duration	Highly Variable	Little or No Variation	Highly Variable	Not Applicable
Variation in # of Participants Called Across Events	Highly Variable	Little or No Variation	At Discretion of Participant or Market Conditions	None
Examples	Interruptible Rates Emergency Dispatched Load Control (e.g., peaksaver®)	Critical Peak Pricing Economically Dispatched Load Control	DR-1 DR-3	Real Time Pricing Static TOU DR-2

LOAD IMPACT ESTIMATION PROTOCOLS

Table C-3					
Day Types and Event Conditions for Which Load Impacts Shall be Provided					
Purpose	Day Type	Limited Frequency Resources	Limited Variation Resources	High Frequency Resources	Continuous Use Resources
Ex Post Impact Estimation	Each event day	Yes (Based on actual # of participants called)	No	No	No
	Average event day	No	Yes (based on day-weighted average # of program participants)	No	No
	Each of top 5 system load days	Already incorporated in each event day requirement	Yes Based on the actual # of participants called for each event		Yes
Ex Ante Impact Estimation	Monthly system peak day—normal weather year	Yes, for months during which DR resource is available Assume event window is from noon to 6 pm unless maximum event window is less than 6 hours, in which case use maximum event window ending at 6 pm Assume all participants are called for event-based resources			
	Monthly system peak day—extreme weather year	Same as above			
	Average week day for each month—normal weather year	No	No	Yes ¹⁷	Yes
	Average week day for each month—extreme weather year	No	No	Yes ¹⁸	Yes
Validation	Actual & predicted load by temperature	For the estimating sample, provide graph showing predicted and actual average load for each temperature for the average customer			
	Actual & predicted load by hour for average of top 10 system load days	For the estimating sample, provide graph showing predicted and actual average load for each hour of the day for the average of the top 10 system load days for the average customer		Yes	Yes
	Actual & predicted load by hour for average week day for each month	No	No	For the estimating sample, provide graph showing predicted and actual average load for each hour of the day for the average week day for the average customer ¹⁸	

¹⁷ For high frequency resources, average weekday should take into account the fact that the resource is not in effect in all hours on every day and document the assumptions underlying the analysis.

LOAD IMPACT ESTIMATION PROTOCOLS

C.5. *PROTOCOL 5—PORTFOLIO ANALYSIS*

Protocol 5 applies when it is necessary to produce load impact estimates for multiple DR resources that may be called simultaneously. There are two issues that must be addressed when producing impacts for a portfolio of DR resources. One is that a common set of event conditions should be used. This is covered by the ex ante impact requirements in Protocol 4, which dictates that a common set of day types, weather conditions and event windows be used for all resource options for ex ante estimation.

The second issue concerns avoiding double counting when customers are allowed to participate in multiple programs. If two programs for which contemporaneous enrolment is allowed are called simultaneously, ideally, program rules will specify the program under which customers will be paid. For example, if a customer is allowed to participate in both an emergency program and a demand bidding program, the emergency program typically will dominate if both are called simultaneously. If program impact estimates for each individual program were added together, this would overestimate the joint impact of the two programs because, when both are called, only the impacts associated with the dominant program should be counted. Protocol 5 addresses this issue. If no program rules have been established, the evaluator will need to decide which program is likely to take precedence and to report that assumption.

Protocol 5—Portfolio Analysis

Whenever load impacts are produced for a portfolio of DR resources that can be called simultaneously and in which customers are allowed to participate in more than one resource option, the portfolio estimate shall only include load impacts associated with the dominant resource for each individual that is enrolled in more than one resource option.

C.6. *PROTOCOL 6—STATISTICAL REPORTING AND VALIDATION*

Protocol 6 establishes reporting requirements designed to allow a knowledgeable and well trained reviewer to assess the quality and validity of the analysis underlying the impact estimates. Some of the requirements are unique to analysis methods based on regression modeling. Regardless of the method used, Protocol 6 requires that test results be presented showing the extent of bias that may exist in the impact estimates based in part on the analytical tests presented previously in Protocol 4.

Protocol 6—Statistical Reporting and Validation

For regression based methods that estimate models based on data aggregated across customers, the following statistics and information shall be reported:

LOAD IMPACT ESTIMATION PROTOCOLS

- 1. Adjusted R-square or, if R-square is not provided for the estimation procedure, the log-likelihood of the model;**
- 2. Total number of observations used in each regression, the number of cross-sectional units underlying the analysis, and the number of days worth of hourly data that were used;**
- 3. Coefficients for each of the parameters of the model**
- 4. Standard errors for each of the parameter estimates**
- 5. Tests conducted and corrections made, if any, to ensure robust standard errors.**

For regression-based methods that estimate models at the individual customer level, a histogram showing the distribution of R-square values for the individual customer regressions shall be provided, along with an R-square estimate representing the average customer. It is not necessary to report coefficient values or standard errors for each parameter for each individual regression.

For all methods, whether regression-based or not, a summary of the results of all tests that were conducted to determine the extent of bias that exists for the impact estimates presented, including, but not necessarily limited to, the validation tests summarized in Table C-3, must be provided.

C.7. *PROTOCOL 7—ANALYSIS BASED ON SAMPLING*

Depending on the number of customers enrolled in a DR resource and/or the cost of end-use metering, surveys and other activities needed to develop load impact estimates, it may be necessary to work with data from a sample of participants rather than the entire group of enrollees. As discussed at length in Appendix 2, whenever a sample of participants is used for analysis, issues of bias, precision and other concerns come into play. The goal is to reduce or eliminate these problems whenever possible and cost-effective to do so. In all situations, it is important to track and document the characteristics of the samples used relative to the population, and the steps taken to minimize bias and other issues that may arise when sampling occurs. Protocol 7 establishes the minimum requirements that must be followed when samples are employed.

Protocol 7—Analysis Based on Sampling

If sampling is required, evaluators shall use the following procedures to ensure that sampling bias is minimized and that its existence is detected and documented.

LOAD IMPACT ESTIMATION PROTOCOLS

- 1. The population(s) under study must be clearly identified and described – this must be done for both participants and control groups to the extent that they are used;**
- 2. The sample frame(s) (i.e., the list(s) from which samples are drawn) used to identify the population(s) under study must be carefully and accurately described and if the sample frame(s) do not perfectly overlap with the population(s) under study, the evaluator must describe the measures they have taken to adjust the results for the sample frame so that it reflects the characteristics in the population of interest – this would include the use of weighting, matching or regression analysis;**
- 3. The sample design used in the study must be described in detail including the distributions of population and sample points across sampling strata (if any);**
- 4. A digital snapshot of the population and initial sample from the sample frame must be preserved – this involves making a digital copy of the sample frame at the time at which the sample was drawn as well as a clean digital copy of the sample that was drawn including any descriptors needed to determine the sampling cells into which the sampled observations fall;**
- 5. For impact estimates based on experimental studies, the “fate” of all sampled observations must be tracked and documented throughout the data collection process (from initial recruitment to study conclusion) so that it is possible to describe the extent to which the distribution of the sample(s) may depart from the distribution of the population(s) of interest throughout the course of the study; If significant sample attrition is found to exist at any stage of the research process (i.e., recruitment, installation, operation), a study of its impact must be undertaken.**
- 6. If selection bias is suspected, the evaluator must describe it as well as any efforts made to control for it.¹⁸**

¹⁸ The problem of controlling for selection bias has been discussed at great length in the literature on econometrics. The seminal articles on this topic are by James Heckman: “The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models”, in *The Annals of Economic and Social Measurement* 5: 475-492 1976; and Sample selection bias as a specification error” in *Econometrica*, 47: 153-161.

LOAD IMPACT ESTIMATION PROTOCOLS

C.8. *PROTOCOL 8—REPORTING AND DOCUMENTATION*

The final protocol focuses on documentation and reporting. Evaluation reporting has a variety of objectives, including:

- Describing the evaluation objectives and plan;
- Presenting the detailed impact estimates developed as part of the evaluation;
- Comparing these findings with DR resource goals and any impacts that have been previously used to report progress toward goals, and explaining any differences;
- Thoroughly documenting the methodologies used in sufficient detail so that, given access to the same data and information, a trained evaluator would be able to reproduce the reported impact estimates;
- Reporting any deviations from the requirements of these protocols and the reasons why it was not possible to meet them;
- Providing recommendations regarding modifications that would improve the DR program;
- Providing recommendations concerning future evaluation activities.

Evaluation reports should be written for a broad audience, including people who are not familiar with evaluation methods or the field's specialized terminology. Technical information associated with evaluation methodologies, research design, sampling, M&V efforts, regression analysis, bias detection, bias correction and other technical areas must be reported and should not be avoided to ensure readability by a wider audience. While a summary of the methodology, findings and decisions covering these issues should be written for a broader audience, the more technical details relating to these reporting categories must also be provided.

Protocol 8—Reporting and Documentation

DR impact evaluation reports shall include, at a minimum, the following sections:

- 1. Executive Summary - this section should very briefly present an overview of the evaluation findings and any recommendations for changes to the DR resource;***
- 2. Introduction and Purpose of the Study - this section should briefly summarize the DR resource or resources being evaluated and provide an overview of the evaluation objectives and plan,***

LOAD IMPACT ESTIMATION PROTOCOLS

including the research issues that are addressed. It should also provide a summary of the report organization;

3. *Description of DR Resources Covered in the Study - this section should provide a detailed description of the resource option being evaluated in enough detail that readers can understand the DR resource that delivered the estimated impacts. The description should include a history of the DR program, a summary of resource goals (both in terms of enrolment and demand impacts), tables showing reported progress toward goals, projections of future goals and known changes, and any other information deemed necessary for the reader to obtain a thorough understanding of how the DR resource has evolved over time and what changes lie ahead.*
4. *Study Methodology - this section should describe the evaluation approach in enough detail to allow a repetition of the study in a way that would produce identical or similar findings. (See additional content requirements below.)*
5. *Validity Assessment of the Study Findings – this section should include a discussion of the threats to validity and sources of bias and the approaches used to reduce threats, reduce bias and increase the reliability of the findings, and a discussion of confidence levels. (See additional content requirements below.)*
6. *Detailed Study Findings - this section presents the study findings in detail. (See additional content requirements below.)*
7. *Recommendations - this section should contain a detailed discussion of any recommended changes to the DR program as well as recommendations for future evaluation efforts.*

The Study Methodology section shall include the following:

1. *Overview of the evaluation plan;*
2. *Questions addressed in the evaluation;*
3. *Description of the study methodology, including not just the methodology used and the functional specification that produced the impact estimates, but also methodologies considered and rejected and interim analytical results that led to the final model specification. The intent of this section is to provide sufficient detail so that a trained reviewer will be able to assess the quality of the analysis and thoroughly understand the logic behind the methodology and final models that were used to produce the*

LOAD IMPACT ESTIMATION PROTOCOLS

impact estimates, and the statistics required to be reported in Protocol 7.

- 4. How the study meets or exceeds the minimum requirements of these protocols or, if any protocols were not able to be met, an explanation of why they were not and recommendations for what it will take to meet these protocols in future evaluations;***
- 5. How the study addresses the technical issues presented in these Protocols;***
- 6. Sampling methodology and sample descriptions (including frequency distributions for population characteristics from any surveys done in conjunction with the analysis).***

The Validity Assessment section of the report shall focus on the targeted and achieved confidence levels for the key findings presented, the sources of uncertainty in the approaches used and in the key findings presented, and a discussion of how the evaluation was structured and managed to reduce or control for the sources of uncertainty. This section should:

- 1. Discuss and assess all potential threats to validity given the methodology used;***
- 2. Provide the evaluator's opinion of how the types and levels of uncertainty affect the study findings;***
- 3. Include information for estimation of required sample sizes for future evaluations and recommendations on evaluation method improvements to increase reliability, reduce or test for potential bias and increase cost efficiency in the evaluation study(ies);***
- 4. Present the results of the validity tests required in Protocol 4.***

The Detailed Study Findings section shall include the following:

- 1. A thorough discussion of key findings, including insights obtained regarding why the results are what they are;***
- 2. All output requirements and accompanying information shown in Table C-1 for each set of event conditions, day types and customer segments. If the number of data tables is large, the main body of the report should include some exemplary tables and explanatory text with the remaining required tables provided in appendices in electronic format;***

LOAD IMPACT ESTIMATION PROTOCOLS

- 3. *For ex post evaluations of event-based resources, a table summarizing the relevant characteristics associated with each event and the date of each event over the historical evaluation period. At a minimum, the table should include for each event: date, weather conditions (for weather-sensitive loads), start and end times for the event, event duration in hours, number of participants notified, and number of participants enrolled;***
- 4. *For ex ante forecasts, detailed descriptions of the event and day type assumptions underlying the estimates;***
- 5. *For ex ante forecasts, assumptions and projections for all exogenous variables that underlie the estimates;***
- 6. *A comparison and detailed explanation of any significant differences in impact estimates derived from the analysis itself, any impact estimates previously obtained in other studies, and those previously used for reporting of impacts toward resource goals.***

GLOSSARY OF TERMS

Critical-Peak Pricing (CPP)

A retail electricity rate with a basic rate structure similar to a TOU rate, but with the provision to replace the normal peak-period price with a much higher event price under specified trigger conditions (e.g., when system reliability is compromised or supply prices are very high).

Demand response

A change in electricity demand by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity demand at times of high wholesale market prices or when supply resources are limited.¹⁹

Dynamic pricing

A retail electricity rate in which there is some uncertainty in the timing associated with a known price, in the price itself, or both. CPP and RTP tariffs are examples of dynamic prices.

Ex ante load impact estimate

A load impact estimate representing a set of conditions or group of customers, or both, that differ from historical conditions.

Ex post load impact estimate

A load impact estimate representing a set of conditions that actually occurred on a specific date or over some period of time for the customers that were enrolled in the program and called on that date or over that period of time.

Free rider

A participant in a conservation program who would have adopted the measure promoted by the program on his or her own initiative, in the absence of the program's influence.

Evaluation, Measurement & Verification (EM&V)

The undertaking of studies and activities aimed at assessing and reporting the effects of a Conservation program on its participants and/or the market environment. Effectiveness is measured through energy efficiency and cost effectiveness.

Independent Electricity System Operator (IESO)

A regulated government corporation that acts as the settlement agent for the Ontario wholesale spot market and as the system controller. The IESO is

¹⁹ U.S. Department of Energy, *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them*. Report to the United States Congress Pursuant to Section 1252 of the Energy Policy Act of 2005, February 2006, page 6.

GLOSSARY OF TERMS

responsible for maintaining system reliability by forecasting and balancing supply and demand.

Incentives

Financial support designated to encourage program participation and lower additional incremental equipment costs. For energy efficiency programs, the most common form of incentive is a rebate designed to help offset the cost of purchasing a more expensive piece of efficient equipment. For demand response programs, incentives are typically payments for measured and verified demand reductions. For time-varying pricing, customers may face incentives tied to the price differentials inherent to the rate design, but do not receive payments from a program administrator.

Load Impacts

Load impacts associated with DR resources are defined as the difference between a customer's actual (observed) electricity demand, and the amount of electricity the customer would have demanded in the absence of the DR program incentive.

Local Distribution Company (LDC)

An entity that owns and operates low-voltage wires and distributes electricity from the IESO controlled grid to end-use customers in local regions.

Ontario Energy Board (OEB)

An agency responsible for regulating all non-commodity electricity rates, setting electricity prices for low volume and designated customers, and licensing the IESO and all market participants.

Time-varying pricing

Any electricity rate in which prices vary by time of day. Examples include TOU, CPP and RTP rates.

Time-of-use (TOU) rate

A retail electricity rate with fixed unit prices for usage during pre-established blocks of time, usually defined by time-of-day and season.

Real-time pricing (RTP)

A retail electricity rate in which the price for electricity fluctuates on an hourly (or sub-hourly) basis, reflecting changes in the wholesale price of electricity.

Reference load

An estimate of the load that a DR program participant would have used at a specific time if they had not been not participating in a DR program.

Structural benefiter

A participant in a DR program who benefits from participation even if they do not change their behaviour in response to the price or other incentives offered by the program.

APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

A variety of methods can be used to develop load impact estimates. Their suitability depends on the type of DR resource analyzed and the goal of the impact estimates. Some methods that are well suited for *ex post* impact estimation are poorly suited for *ex ante* impact analysis. Regardless of the purpose for which impact estimates are being developed, accuracy and precision are critical criteria. Methodological transparency and the ability to compare “apples to apples” across multiple resource options are also important.

1.1 ACCURACY

The most important characteristic of any impact estimate is accuracy. An accurate estimate is an unbiased estimate. Bias is often confused with precision. Precision is the degree of certainty associated with an estimate. However, it is possible to produce an estimate that is known with a high degree of certainty – that is, one that is precisely estimated – but that is biased relative to the true value of a variable for the population of interest.

Three sources of potential inaccuracy are worth noting. The first is sampling bias.²⁰ Sampling bias can arise if estimates for a particular group of customers are based on a sample of customers that differ in relevant ways from the population at large. For example, suppose that the objective is to estimate what the average impact is likely to be for a DR program for the commercial/industrial customer class as a whole. If the estimate is based on a sample of customers comprised of only commercial office buildings, the impact estimate might be quite precise but it could be biased if, for example, commercial office buildings are either more or less responsive to DR incentives than other segments of the C&I population.

Another source of inaccuracy in load impact estimation can arise from the estimation process itself. As discussed in Section B.1, if load impacts are estimated as the difference between the reference load and the measured load during an event period, any inaccuracy, or bias, in the reference load will affect the load impact estimate. Depending on circumstances, it may be possible to reduce this type of bias by estimating the load impact as the difference between two predicted values, the reference load and the predicted value with DR in effect. This approach would not eliminate any bias that might exist in the DR impact coefficient in a regression model, for example, but it would eliminate any forecast error in the model that is due to model misspecification or to errors in variables that affect overall energy use but that do not affect the change in energy use due to demand response.

²⁰ Appendix 2 provides additional information concerning sampling.

APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

A third source of inaccuracy that can occur when *ex ante* load impact estimates are being developed stems from inaccuracy in predicting the key drivers of demand response. Load impacts will vary with the characteristics of the participant population in a DR program. It is quite common for participant population characteristics to change over time. For example, initial program participants may be comprised primarily of customers that are large or that have large amounts of load that can be called. As the program expands or matures, the typical new participant might be smaller or have a smaller percent of load that can be reduced. Demand response load impact estimates derived by scaling up the average estimates for these early program participants would over state demand response. To the extent possible, it is important for *ex ante* impact estimates to reflect expected changes in participant characteristics and other factors that drive demand response.

As seen in Section C, the load impact protocols require that each program evaluation pay particular attention to developing unbiased impact estimates and that evidence be presented indicating the extent of bias that exists in the estimates and the steps that were taken to ensure that such bias was minimized. Comparisons between predicted and actual energy use under various conditions is the best method for determining the extent of bias. When examining bias in this manner, it is important that such tests be done not just under average conditions, but also under extreme conditions that drive demand and demand response. For example, there is not much value in knowing how well a model predicts for an air conditioner cycling program for a typical summer day, since this DR resource is likely to be called only on very hot summer days. It may also be important to know how well the model predicts under very extreme weather conditions, such as the July 2006 heat wave that affected most of North America for a two-week period.

Figure 1-1 shows a useful graph for assessing model validity. The graph compares the average predicted and actual energy use based on a model that was developed to estimate impacts for Southern California Edison's air conditioner cycling program.²¹ As seen, at temperatures below about 92°F, the model has a slight upward bias whereas, at temperatures above about 97°F, the predictions have a slight downward bias. On average, across summer weekdays, the model predicts quite well, as evidenced in Figure 1-2. However, during the very extreme July heat wave of 2006, which in California represented a 1-in-50 or even 1-in-100 year heat wave (depending on who you ask), the model was downward biased between about 10 am and 6 pm. Working hard to minimize the magnitude of such biases is a key focus of load impact estimation. Even more important is having a clear understanding of any bias that might still exist after all reasonable steps have been taken to minimize it.

²¹ Stephen George, Josh Bode and Josh Schellenberg. *Load Impact Estimates for Southern California Edison's Demand Response Program Portfolio*. Prepared by FSC for SCE. September 25, 2008.

APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

Figure 1-1
Actual versus Predicted Load by Temperature
Southern California Edison Company's Air Conditioner Cycling Program

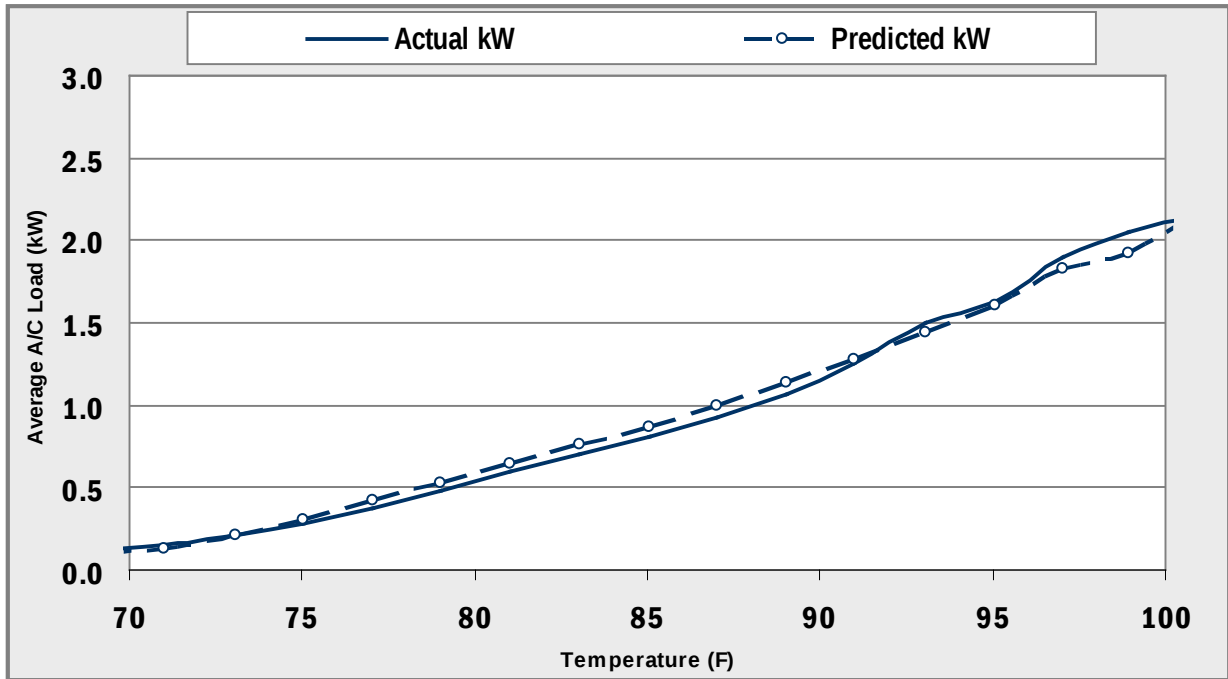
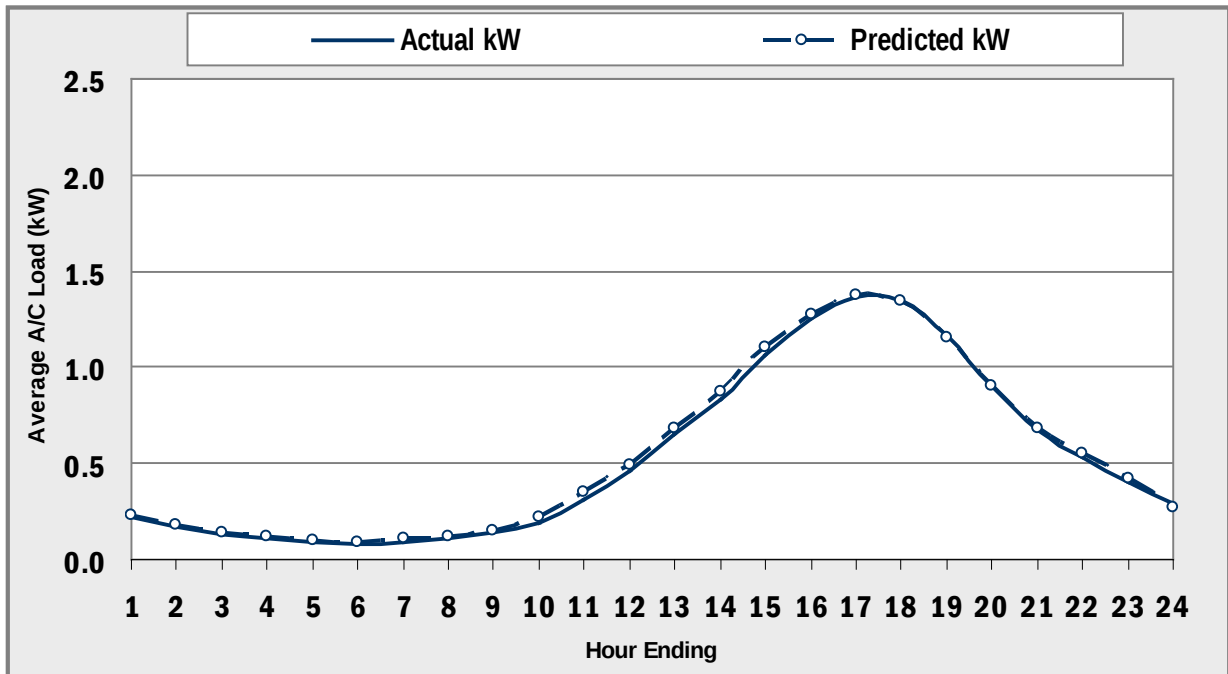
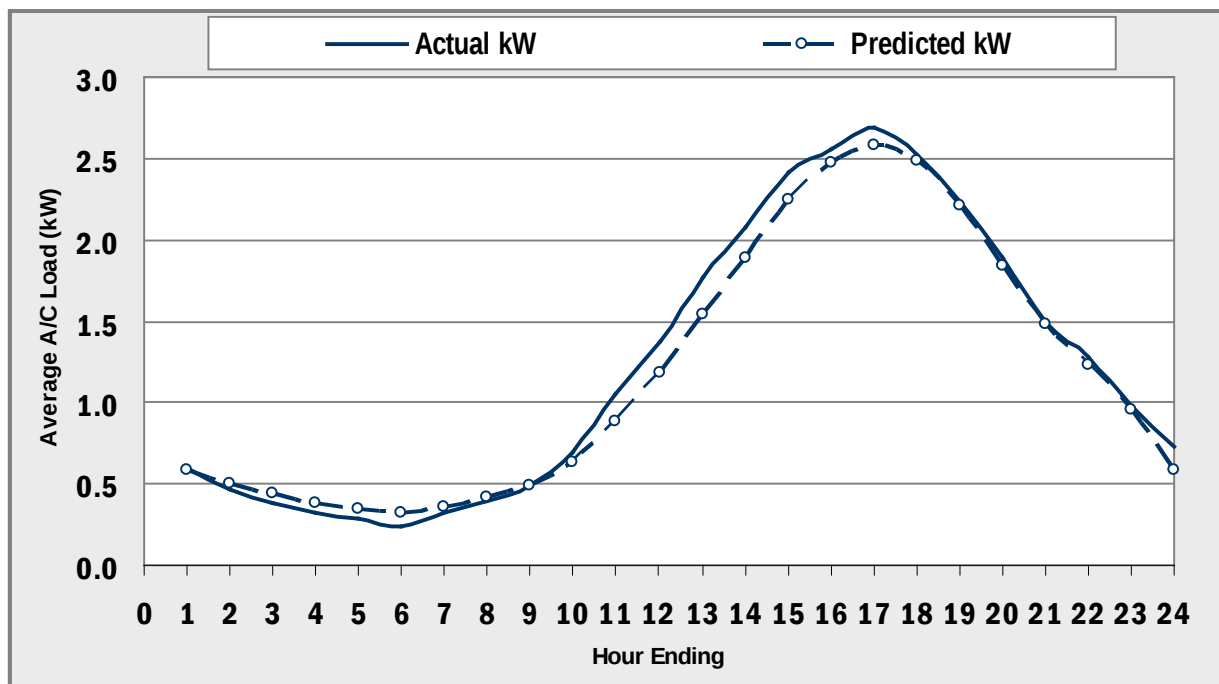


Figure 1-2
Actual versus Predicted Load for a Typical Summer Weekday
Southern California Edison Company's Air Conditioner Cycling Program



APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

Figure 1-3
Actual versus Predicted Load for the 2006 Heat Wave (July 16th through 24th)
Southern California Edison Company's Air Conditioner Cycling Program



1.2 PRECISION

As described above, statistical precision concerns the degree of uncertainty associated with impact estimates. In statistical terms, precision is typically reported in terms of the size of the confidence bands around the predicted impact. There are at least two sources of uncertainty that must be considered when developing DR impact estimates.

One potential source of uncertainty is sampling error.²² A sample is a subset of the population of interest and, as such, typically will not have exactly the same statistical properties as the population as a whole. Consequently, sample estimators such as means, standard deviations, frequency counts, etc. will vary from random sample to random sample. Depending on the number of enrolled customers, it may not be necessary to use a sample to estimate load impacts. However, whenever sampling is used to describe the characteristics of a population, there is some uncertainty about the estimates from the sample that comes from random variation in the sampling process.

Samples can be made to be nearly perfectly precise for all intents and purposes. However, sampling precision is not inherently valuable and it comes at a cost in terms of meter installation, sample maintenance and database management. In

²² This issue is discussed further in Appendix 2.

APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

essence, the reduction in uncertainty associated with sampling error has to be balanced against the increased cost of obtaining more precise estimates in sampling. An important step in designing a DR load impact evaluation is to identify the extent of sampling precision required to support decision making. There are no hard and fast rules concerning how much sampling precision is enough. It depends on how the information is intended to be used. Establishing an appropriate level of sampling precision is best done by consulting with the intended users of the information and asking them to agree to an acceptable sampling error rate.

The extent of sample-to-sample variation in measurements generally depends on the inherent variation in the factor of interest in the population (in this case hourly loads) and the number of observations that are sampled. In general, the more homogeneous the population of interest with respect to the variable of interest, the lower the sample-to-sample variation in measurements that can occur. If every element in a population is the same or nearly the same with respect to the variable of interest, there will be little sample-to-sample variation obtained through random sampling. On the other hand, if the elements in the population are very different from one another with respect to the variable of interest, there will be high sample-to-sample variation obtained through random sampling.

Another source of uncertainty in impact estimation arises from inherent uncertainty in the factors that influence DR impacts. Perhaps the simplest example of this is weather. For such factors, the inherent uncertainty in the impact estimates can be accounted for by estimating impacts based on a probability distribution of, for example, temperature.

1.3 TRANSPARENCY

Another useful criterion for guiding impact estimation methodology selection and analysis is transparency. Transparency affects the ability of decision makers and other stakeholders to understand and agree that the impact estimates are suitable. Certain impact estimation methods are much easier to understand than others. For example, a method that estimates the impact as the difference between the metered load and a reference load calculated as the average load across the prior 10 days, referred to as day-matching or baseline methods, is much easier to understand than a method based on regression analysis using pooled time-series and cross section data with a fixed effects model incorporating weather and price interaction terms. Day matching methods can also be implemented quickly, such as the day after an event, whereas regression methods might take much longer and might logically be done after all events in a season have been called so that variation across events can be incorporated into the impact model. For these and other reasons, day-matching methods are typically used for customer settlement, since the inherent transparency of the approach is useful for obtaining customer buy-in and the speed of application is useful for monthly settlement. On the other hand, regression methods are needed for *ex ante* estimation and the audience for

APPENDIX 1: CRITERIA FOR DEVELOPING GOOD IMPACT ESTIMATES

these estimates is typically sophisticated enough to understand and evaluate the estimates despite their being less transparent.

1.4 COMPARABILITY

One final criterion worth mentioning is comparability. From a planning perspective, it is useful to be able to compare two or more DR program options in terms of their ability to replace supply-side resources under specific event conditions. This does not necessarily mean that the impact estimates for each DR option must be developed using the same methods. Rather, it means that each of the estimates should be based on the same forecast of relevant drivers (e.g., that the same weather be used to predict impacts for two or more weather sensitive programs) and that there be a common set of outputs provided (e.g., that impact estimates be provided for a common set of hours and day types for each program).

APPENDIX 2: SAMPLING OVERVIEW

APPENDIX 2: SAMPLING OVERVIEW²³

Sampling is a useful procedure in estimating DR load impacts because information needed for impact estimation (i.e., interval load measurements) often is not available for the customers who are participating in DR programs and installing interval meters that are often needed to estimate impacts is costly. Even in the future when most customers have interval meters, sampling may be useful as a means to reduce analysis costs when the volume of data available for describing load impacts is large. Despite these obvious advantages, relying on sampling for estimating load impacts increases uncertainty about the accuracy and precision of load impact estimates.

If interval load data is available for the entire population of DR resource participants, evaluators should strongly consider using all available information to estimate load impacts. Analyzing data from the entire population of resource participants eliminates the need for sampling and the attendant concerns about potential sampling bias and sampling precision discussed in this section.

The decision to employ sampling and the numerous technical decisions required in sample design are driven by the broader research issues that are addressed during evaluation planning. These issues were discussed in Section B. Examples include: required sampling precision, statistical confidence; the need for geographical specificity; the need for segmentation by customer types; the temporal resolution of the measurements, etc. As Figure 2-1 illustrates, taking account of these considerations, it is possible to specify an appropriate statistical or econometric estimation model for the study as well as the specific measurements that must be made to drive the estimation process. Working from these decisions, it is then possible to determine whether sampling is appropriate and if so, to identify the most efficient sample design given the available resources. It is also possible as a result of the sampling process to inform stakeholders of the technical constraints associated with the available resources and to therefore make possible adjustments to expectations or resources before the actual study is fielded.

Sampling adds three potential sources of uncertainty about the magnitude of load impact estimates:

- The potential for bias or inaccuracy resulting from the processes used to select and observe load impacts (i.e., sampling bias);

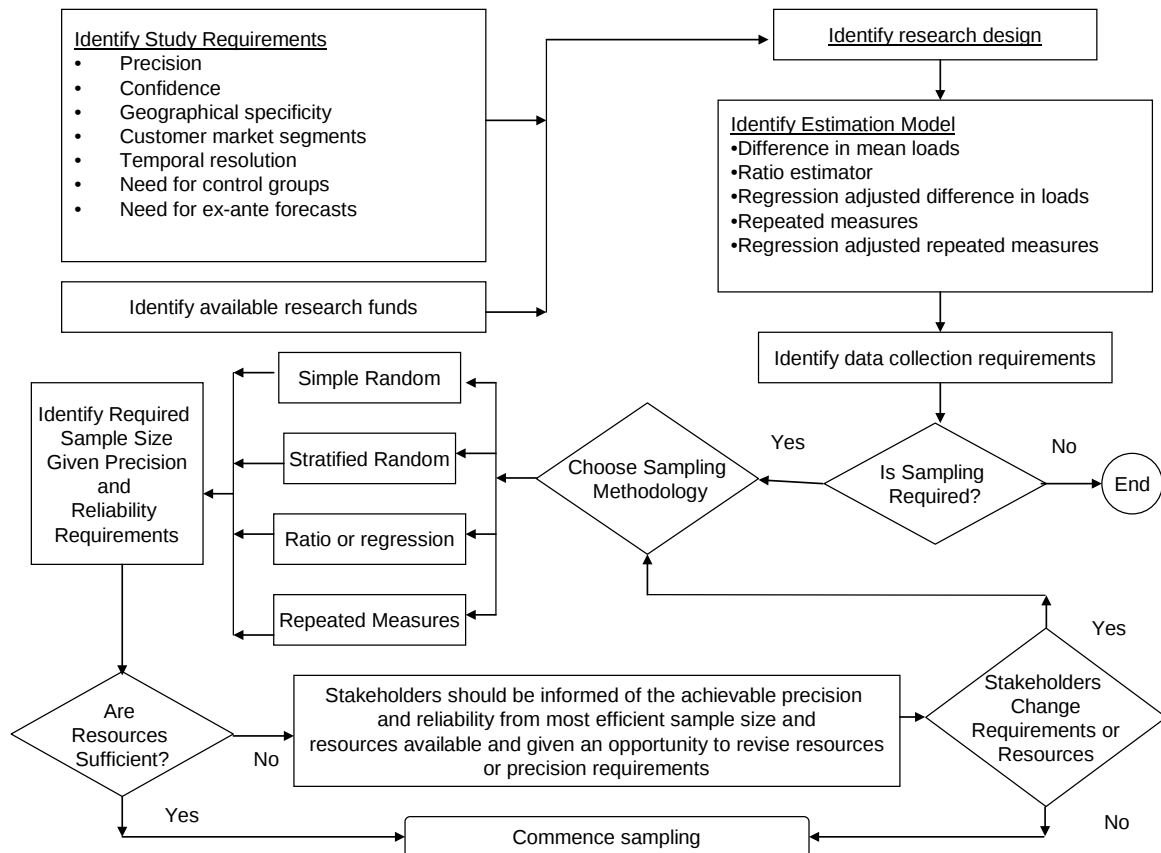
²³ This section is taken largely verbatim from Section 8 of *Load Impact Estimation for Demand Response: Protocols and Regulatory Guidance*, California Public Utilities Commission, March 2008. A draft version of this document was written by Freeman, Sullivan & Co. on behalf of California's three investor owned utilities, PG&E, SCE and SDG&E, and submitted on their behalf to the CPUC. The CPUC made minor edits to the document and adopted it for use by California's utilities to develop load impact estimates for DR programs in CA.

APPENDIX 2: SAMPLING OVERVIEW

- Increased imprecision in the load impact estimates arising from sampling error (i.e., error arising from the inherent sample-to-sample variation that will occur when samples are used to estimate load impacts from the population); and
- Concern about the reliability of load impact estimates obtained from samples (i.e., concern that the results obtained from the sample may accidentally over or understate load impacts).

These issues should be directly addressed whenever sampling is used to estimate load impacts. Recommended approaches and resources for dealing with these issues are discussed below.

**Figure 2-1
Sample Design Process Diagram**



2.1 SAMPLING BIAS

By far the most dangerous source of uncertainty arising from sampling is sampling bias. When sampling bias occurs, what is true of the sample is not necessarily true of the population – no matter how large the sample is.

APPENDIX 2: SAMPLING OVERVIEW

Sampling Bias refers to the accuracy of the estimates obtained from a sample.

To understand sampling bias, it is useful to think of a simple measuring instrument like a ruler or scale. If a scale accurately measures the weight of an object, it is said to be unbiased. Like a household scale, a sample is said to be unbiased if it accurately measures the parameters in a statistical distribution (e.g., the mean, proportion, standard deviation, etc.). The accuracy of a scale or ruler is ensured by calibrating the scale to a known quantity. The accuracy of a sample estimator is ensured by the method used to select the sample (i.e., whether or not observations are sampled randomly.)

There are two important sources of sampling bias:

- Under-coverage bias – a situation in which the sample frame from which the study participants are selected does not represent important elements of the population. (At present, under-coverage bias is not a problem with samples chosen for DR resource impact estimation because the population of participants in DR resources is known); and
- Selection bias – a situation in which elements in the sample are selected in such a way that they are not representative of the population of interest.

The best way to control sampling bias is to eliminate it by sampling observations for study at random from the populations of interest. This practice will ensure that the initial sample is “representative” of the population of interest. Whenever possible, this approach to sampling should be employed. Unfortunately, it is virtually impossible to completely enumerate (i.e., observe all sampled members) a random sample when people are involved; and this opens up the possibility of sampling bias even when random sampling has been undertaken.

There are many ways in which randomly selected observations can be systematically “selected out” of a given study before they can be observed. Examples of potential sources of selection bias include:

- Technical constraints associated with telecommunications, meter installation or other physical constraints that may limit the installation of interval meters to a subset of sampled customers;
- Participants may refuse to supply information that is necessary for impact estimation (i.e., non-response to survey elements that may correspond with load impact measurements); and
- Participants may migrate out of the study while it is in progress.

It is important to keep in mind that the mere fact that some randomly sampled observations are not completely observed (i.e., have been selected out of the sample at some point) does not necessarily mean that the resulting sample has been biased in

APPENDIX 2: SAMPLING OVERVIEW

some significant way. Whether bias is induced by selection depends on whether the selection is somehow related to the magnitude of the impact of the DR resource. This can only be determined by carrying out the work outlined above.

The first and most important step in minimizing selection bias is to dedicate adequate project resources to ensuring that initially selected sample points are observed during the study. Because the cost of data collection varies (sometimes dramatically) from observation to observation, it is sometimes tempting to restrict data collection to observations that are easy to recruit or inexpensive to observe. This temptation should be resisted. The 20% of observations that are the most difficult and expensive to observe may be the most important ones to observe.

2.2 SAMPLING PRECISION

A sample is a subset of the population of interest and as such will not, in general, have exactly the same statistical measurements as the population as a whole. Correspondingly, sample estimators such as means, standard deviations, frequency counts etc. will vary from random sample to random sample. Thus, whenever sampling is used to describe the characteristics of a population, there is some uncertainty about the estimates from the sample that comes from random variation in the sampling process. While we sometimes find it convenient to talk about the results obtained from a sample as though they were “point estimates” of the measures of the population of interest, it is generally inappropriate to interpret the results of sampling without considering the sample-to-sample variation that is likely to have occurred. This is the problem of sampling precision.

Sampling Precision refers to the magnitude of random sampling error present in the parameter estimates obtained from a sample.

Again, it is useful to consider the example of a scale. Some scales (e.g., household scales) can measure the weight of objects to within plus or minus 1/2 lb., while others (like those used in chemistry laboratories) can measure objects to within plus or minus 1 microgram. The range within which an accurate measurement can be taken is the precision of the scale. Likewise, the measurements of the population parameters taken from a sample can be said to be more or less precise—that is, the population parameters can be measured with more or less statistical error depending on a number of considerations such as sample size, stratification and the inherent variability in the parameter of interest. This is what is meant by sampling precision.

The extent of sample-to-sample variation in measurements generally depends on the inherent variation in the factor of interest in the population (in this case hourly loads) and the number of observations that are sampled. In general, the more homogeneous the

APPENDIX 2: SAMPLING OVERVIEW

population of interest is with respect to the variable of interest, the lower the sample-to-sample variation in measurements that can occur. If every element in the population is the same or nearly the same with respect to the variable of interest, then there will be little sample to sample variation obtained through random sampling. On the other hand, if the elements in the population are very different from one another with respect to the variable of interest, there will be high sample to sample variation obtained through random sampling.

It is also true that the larger the sample size, the lower the sample-to-sample variation in measurements. This is because the standard error of the mean (average distance of the sampled mean from the true population mean) decreases with the square root of the sample size. This can be seen in the formula for the standard error of the mean shown in Equation 2-1:

$$\sigma_m^2 = \frac{\sigma^2}{n} \quad (2-1)$$

where σ_m^2 is the standard error of the mean, σ^2 is the variance of the population, and n is the sample size.

Both of the determinants of sampling precision described above can be manipulated by design to establish desired levels of sampling precision.

The standard error or average distance of sampled means from the center of the sampling distribution is a useful measure of sampling precision because it explains how far on average the sample can be expected to stray from the mean of the population given its variance and sample size. However, an even more useful measure of sampling precision can be derived from the standard error of the mean by computing the interval within which the true population estimate is likely to be found. This is called the confidence interval. The confidence interval for a sample estimator is the interval in which the true population value is likely to be found with a certain probability. So, for example, you often see sample estimators described in terms of upper and lower confidence limits expressed in terms of percentages. The confidence interval for a given estimator is obtained by multiplying the standard error of the mean times the area under the sampling distribution for the mean associated with the observation of a given extreme value (i.e., 90%, 95% or 99%). This can be seen in the formula for the confidence interval of the mean shown in Equation 2-2:

$$\bar{x} - z\sigma_m^2 \leq \mu \leq \bar{x} + z\sigma_m^2 \quad (2-2)$$

where $\bar{x}$ is the sample mean, z is the value of the z distribution associated with the selected confidence level, and σ_m^2 is the standard error of the mean.

The confidence interval is a useful statistic because it reflects the upper and lower limits within which the true population value will be found with a given level of certainty. It is particularly useful in operations and resource planning where users will generally want to incorporate the maximum amount of load impact they can confidently expect to occur in

APPENDIX 2: SAMPLING OVERVIEW

their decision making and planning. Whenever load impacts are calculated based on sampling, the upper and lower confidence limits should be reported. The confidence levels or probabilities employed in the calculation should be determined in consultation with the users of the information.

It is important to keep in mind that sampling precision and sampling bias are two very different things. One cannot overcome inaccuracy or bias in load impact measurements induced by inaccurate reference load measurements or sample selection by increasing sampling precision as this will simply result in a more precise estimate of the wrong answer.

2.2.1 Establishing Sampling Precision Levels

Samples can be made to be nearly perfectly precise for all intents and purposes. However, sampling precision is not inherently valuable and it comes at a cost in terms of meter installation, maintenance and database management. In essence, the reduction in uncertainty associated with sampling error has to be balanced against the increased cost of obtaining more precise estimates in sampling.

An important step in designing a DR load impact evaluation is to identify the extent of sampling precision required to support decision making. There are no hard and fast rules concerning how much sampling precision is enough. It depends on how the information is intended to be used. Establishing an appropriate level of sampling precision is best done by consulting with the intended users of the information and asking them to agree to an acceptable sampling error rate.

There are two related issues that must be decided in this conversation – identification of an acceptable level of sampling precision (e.g., plus or minus 5% or 10% or whatever) and identification of the desired reliability of the estimate (e.g., 95% reliable, 90% reliable, etc.). In the end, it is important to agree with intended users about both the precision and reliability of the estimators coming from the sample – since these two issues can be traded off against one another. Once the desired level of sampling precision has been determined, an appropriate sample design can be determined.

APPENDIX 2: SAMPLING OVERVIEW

Confidence Level – refers to the likelihood that parameter estimates obtained from a sample will actually be found within the range of sampling precision calculated from the sample.

It is possible to take a sample, and just by chance to observe a result that is quite different from that of the actual population; and if another sample was taken a completely different result would be found. This can happen just because of sampling error. So, a reasonable question to ask is: “how sure are you that the results obtained in your sample actually describe the situation in the population?”

This question can be answered by calculating the likelihood that the parameter of interest falls within a certain range given the size of the sample and the variation observed in the sample. This likelihood is usually described as a percentage like 90% or 95%. This percentage refers to the percentage of the intervals (between upper and lower limits) that can be expected to contain the true population parameter given the sample size and variation observed in the sample.

2.2.2 Overview of Sampling Methodology

Sampling is a well developed scientific discipline and there are well known textbooks that outline technical approaches to sample design that are appropriate for designing samples to be used in DR load impact estimation. These include classics such as Cochran, Kish and Deming.²⁴ While an in-depth treatment of sample design is well beyond the scope of this document, there are certain sample design options that are more appropriate for DR load impact estimation than others and the remainder of this section discusses issues that favour using some designs over others under certain conditions.

Sample design is a highly technical art that requires training and experience in statistics and survey sampling. If the expected level of investment in metering and data collection is significant for a given resource, it is recommended that evaluators consult with an expert survey statistician in order to develop an efficient sample design for DR resource impact evaluation.

2.2.2.1 Simple Random Sampling

Any discussion of sampling and sample design must begin with a review of simple random sampling because it is the basis of most sampling procedures that are appropriate for DR load impact estimation. However, for reasons that will be discussed below, simple random sampling will seldom be appropriate in studies of DR load impacts.

²⁴ Classic textbooks useful in survey sampling include:

Sampling Techniques: third edition, by William Cochran, John Wiley and Sons. 1977

Survey Sampling, by Leslie Kish, John Wiley and Sons, 1965

Sample Design in Business Research, by William Deming, John Wiley and Sons 1960

APPENDIX 2: SAMPLING OVERVIEW

In simple random sampling, population units are selected for observation with probability $1/N$. That is, all of the elements in the population have an equal chance of being selected for study. Statistical estimators obtained from such simple random samples are unbiased and consistent.

Equation 2-3 identifies the formula for determining the sample size required to obtain a given level of precision under simple random sampling:²⁵

$$n = \frac{z^2 \sigma^2}{r^2 \bar{x}^2} \quad (2-3)$$

where n is the sample size, z is the value in the z distribution associated with alpha (probability of Type II error), σ^2 is the population variance, r^2 is the relative error (error as a percentage of the mean), and $\bar{x}$ is the population mean.

Notice that this formula requires just two types of information; a desired level of sampling error and an estimate of the standard deviation of the variable of interest in the population. In most cases, the standard deviation of the variable of interest in the population is unknown and must be estimated by proxy from the distribution of some variable for which these values are known. It is also possible to substitute an estimate of the coefficient of variation (CV) for the standard deviation in the above equation and solve for sample size. The CV is equal to the ratio of the standard deviation to the mean.

Load research has been underway for many years in the utility industry and in most cases it is possible to identify a reasonable proxy for the standard deviation of an electric load in the population of interest or, in the absence of that, a reasonable estimate of the coefficient of variation. Using the above information, the sample size required to obtain a given level of statistical precision is easy to calculate.

Simple random sampling is easy to do and the results obtained from it can be directly used to estimate population parameters from sample values by multiplying the sample estimates times the sampling fractions (e.g., population weights). So, what's not to like about simple random sampling?

While simple random samples are easy to create and use they have certain limitations in practice. First, because sample elements in simple random samples are selected exactly in proportion to the prevalence of conditions in the population, they may produce relatively small numbers of "interesting" population members that occur relatively rarely. For example, commercial office buildings comprise only a small fraction of all commercial accounts. Too few of these buildings may be selected in a simple random sample of commercial accounts to meaningfully describe the impacts of DR resources on loads in these buildings. To the extent that it is useful to describe the DR load impacts of

²⁵ The actual equation for calculating sample size includes a correction for the size of the population called the finite population correction. This adjustment has been left off of the equation for ease of exposition. In general, its effect on the sample size calculation is *de minimis* when the population of interest is large (e.g., more than a few thousand).

APPENDIX 2: SAMPLING OVERVIEW

important subsets of the population, a simple random sample may not be a practical approach to sampling because the sample size required to select them at random from the population is extremely large.

A second limitation in the usefulness of simple random sampling in DR load impact estimation arises from the fact that customer loads vary widely within populations of DR resources with known customer characteristics (i.e., geographic location, customer type, connected load, etc.). It is not unusual to observe coefficients of variation for energy use and hourly loads ranging from 1 to 4 for these populations. Left unchecked, this variation can lead to greatly inflated requirements for sample size.

These problems are common to most scientific research and many sample design alternatives have evolved to solve them. Consequently, in many applications, more complicated sample designs are often preferred over simple random samples.

2.2.2.2 Stratified Random Sampling

In stratified random sampling, each and every element of the population of interest is pre-sorted into one and only one category for purposes of sampling. Then samples are drawn at random from each category. The sample sizes obtained from each category are generally not proportional to the distribution of the population across the strata, so the sample per se is not representative of the population of interest (i.e., it is biased). This distortion, however, can be used to good effect if properly constructed.

Stratification is very useful in load impact estimation because it allows the researcher to exactly control the distribution of the sample across meaningful categories. Examples of useful stratification variables include: weather zones, usage categories, utility service territories, business types, occupancy patterns and a host of other variables that can have an effect on customer loads. Stratified random samples can be constructed in such a way as to supply known levels of sampling precision within strata and for the population as a whole. In this way they can be used to develop statistically precise estimates of load impacts within weather zones, usage categories and so on. They can also be useful for developing sample designs that are statistically more efficient (i.e., have higher statistical precision at given sample sizes) than simple random samples.

The sample estimators (i.e., means, standard deviations, etc.) for the sampling strata are unbiased estimators of the parameters of interest for the population within each stratum. However, to estimate total population parameters using estimators from stratified random samples, it is necessary to properly weight the estimates obtained from each of the sample strata so that the effects of the measurements from the strata (e.g., mean, standard deviation, proportion, etc.) are proportional to the sizes of the populations in the strata. All statistical estimators obtained through stratified random sampling must be corrected in this manner to produce unbiased total population estimates.

Identification of appropriate sample sizes for stratified random samples is somewhat more complicated than it is in the case of simple random samples. If the purpose of stratification is to obtain designated levels of sampling precision within the strata, then the

APPENDIX 2: SAMPLING OVERVIEW

sample sizes within each stratum are obtained using the formula for simple random sampling – using the estimated standard deviation and desired sampling precision for the stratum. It is not unusual for decision makers to specify that they require a given level of sampling precision for each utility, or by weather zone. In such cases, the sampling precision within the strata will determine the overall sampling precision obtained for the population. The sampling precision for the combined sample (i.e., with all the strata taken together) is obtained by calculating the weighted standard error of the estimate.²⁶ The sampling precision for the entire population should be substantially higher than it is for any of the strata taken alone.

On the other hand, or in addition to the above consideration, stratification can be used to enhance sampling efficiency. In this case, the sample is distributed among the strata in such a way as to minimize the weighted standard error of the total population estimate. Procedures for identifying optimal stratum boundaries and for calculating sample sizes within strata to achieve desired levels of statistical precision in stratified random sampling have been developed by Delanius and Hodges²⁷ and Neyman²⁸ respectively.

Stratified random sampling will almost always be required in assessing DR resource impacts – particularly for resources where it is important to develop reasonably precise measurements within geographic locations or for different customer types. It may also be useful for improving the efficiency of sample designs – though in the case of many resources, the improvements in sampling efficiency obtained from repeated measures designs (discussed below) will overshadow any improvements that may be obtained by pre-stratifying on the basis of customer size.

Whenever stratified random samples are used to estimate DR load impacts, researchers should carefully describe the sample design. Oft-reported measures include;

1. the distribution of the population across sampling strata;
2. the distribution of the sample across sampling strata;
3. any procedures used to identify optimal stratum boundaries used in pre-stratification and the impacts of pre-stratification on sampling efficiency (i.e., if Delanius-Hodges and/or Neyman allocation are used, the researcher should provide a rationale for their choice of the number of strata and stratum boundaries used in the design and their respective impacts on sampling precision);

²⁶ Ibid

²⁷ See “Minimum Variance Stratification” Dalenius T. and Hodges J. L., *Journal of the American Statistical Association*, 1959, 4, pp. 88-101

²⁸ See “On the two different aspects of the representational method: the method of stratified sampling and the method of purposive selection”, Jerzy Neyman, *Journal of the Royal Statistical Society*, 1934, 97, pp 558-625.

APPENDIX 2: SAMPLING OVERVIEW

4. the expected statistical precision for estimators within each strata (including a discussion of any use of proxy measures of the standard deviation used in this calculation); and
5. the expected statistical precision for estimators in the population overall.

2.2.2.3 Sample Designs Using Alternative Estimators

Beyond stratification, there are several other important ways of enhancing the statistical precision of sample estimates. These are used in conjunction with the basic sample designs outlined above. They involve using alternative estimators compared with the conventional approaches discussed above. The conventional sample designs discussed above are focused on identifying sampling procedures that will achieve a certain level of statistical precision in estimating well known parameters of statistical distributions such as the mean and standard deviation. In the case of DR load impacts, these sample designs can be used to achieve a certain level of precision in estimating the average load impact, its standard deviation and confidence intervals.

It is possible and in many cases desirable to create samples designed to measure other parameters in the population that can be used to develop more precise estimates of load impacts than the elementary sample means and standard deviations. Two important alternative estimators that should be considered are *ratio estimators* and *regression estimators*. Under certain circumstances, these estimators can be used to greatly enhance the precision of statistical estimates obtained from sampling and thereby significantly lower the cost of impact evaluation.

Ratio Estimation

Sampling to observe ratio estimators improves efficiency by sampling to observe the relationship in the population between an unknown variable (e.g., the actual load observed during a DR event) and a property that is known for all population members (e.g., the contractual firm service level for subscribers to the resource). To the extent that the actual load observed during the DR event is correlated with the firm service level, the ratio of the two variables will have inherently lower variation than the metric value of the loads involved in the numerator or denominator; and the estimated load impact can be measured with substantially greater precision than the metric loads underlying it. Correspondingly, significantly smaller numbers of sample points are required to observe the ratio of the two variables in the population than would be required to estimate the value of either the numerator or denominator. This is called ratio estimation. Designing samples for ratio estimation follows the same basic logic as for conventional sample designs – except the variable of interest in establishing sampling precision is the ratio, not the metric value of the loads of interest.

The EE protocols devote considerable attention to the technical details of developing samples for ratio estimators and these protocols should be consulted if the use of ratio estimators is being considered in DR load impact estimation. Ratio estimators are very useful in EE resource evaluation because it is relatively easy to conceive of the impact of

APPENDIX 2: SAMPLING OVERVIEW

an EE resource as a ratio of achieved savings to estimated savings for measures that were supposed to have been adopted. DR resources that are excellent candidates for sampling based on ratio estimation are those where participants agree to reduce loads to firm service levels on command and those where participants are demand bidding – both cases where the resource impact is easily defined as a ratio.

Regression Estimation

An extension of the logic of ratio estimation is regression estimation. In regression estimation, sampling efficiency is improved by sampling to observe the relationship in the population between the regression adjusted mean (in this case of hourly load) and variables that influence the value of the regression adjusted mean (e.g., time of day, resource participation, ambient temperature, household size, load in hours prior to the event, etc.). To the extent that hourly loads are correlated with factors that vary systematically in the population, it is possible to define a regression function that will predict those loads more or less precisely.

An interesting property of the regression adjusted mean is that its standard error decreases with $(1-R^2)$. This means that if the R^2 (e.g., the proportion of the variation in the load explained by the regression function) is 0.9, the standard error of the regression adjusted mean is 10% of the standard error of the population mean. Thus, substantial improvement in sampling precision can be obtained if the regression adjusted mean and standard error are estimated instead of the population mean. Of course, the smaller the R^2 for the regression equation, the smaller will be the improvement in sampling precision.

While the potential for improvement in sampling efficiency from regression estimation is tantalizing, researchers have to bear in mind that the extent of improvement in sampling efficiency depends entirely on the predictive power of the regression function that is specified. Practically speaking, this means that the researcher must have some a priori knowledge that the predictors to be included in the regression function actually have substantial predictive power before developing a sample design based on regression estimation. Fortunately, there is ample evidence in prior research concerning customer loads that information about type of customer, time of day, temperature, day of week, and other variables are highly predictive of hourly customer loads.

If the relationships between predictor variables and hourly loads have been studied in prior research, sample sizes for estimating regression functions including variables from the prior research can be calculated directly. This is done by observing the R^2 of the prediction equation (applied to past data) and making a reasonable guess about the incremental increase in R^2 that will result from addition of the effect variable (a new predictor).

Most statistical packages provide algorithms for estimating sample sizes for estimation of effects using multiple regressions. These require making assumptions about R^2 of the model without the effect predictor, the incremental improvement in R^2 that will result from the inclusion of the predictor variable, desired statistical power and alpha (probability of

APPENDIX 2: SAMPLING OVERVIEW

Type II error). For examples of these algorithms see STATA and SPSS software documentation.

In the case where no prior information is available concerning the predictive power of the regression function, sample sizes can be estimated using various rules of thumb involving assumptions about desired statistical power, Type II error (alpha) and the number of predictors in the regression equation. See Tabachnick and Fidell (2001)²⁹ for a discussion of the various rules of thumb that have been applied historically to estimating sample sizes required to estimate regression parameters. Various rules have been suggested. For example, one rule suggests that the minimum sample size for estimating regression coefficients should not be less than 104 plus the number of predictors in the regression equation. Another rule suggests that the sample size should be at least 40 times the number of independent variables in the regression equation. Still another rule says that the minimum sample size should depend both on the effect size that is to be detected and the number of variables in the equation. This rule calculates the minimum sample size as $[8/(\text{effect size})]$ plus the number of independent variables minus 1. All of these rules have some basis in logic and experience, but none can be said to be robust and capable of producing efficient sample size decisions.

Given the uncertainty that may exist about the predictive power of regression models, if circumstances permit, it is advisable to set sample sizes for estimating regression functions using double sampling. In double sampling, an initial sample is drawn that is thought to be sufficient and the parameters in the distribution of interest (in this case regression parameters) are calculated. The initial sample might be drawn according to the first rule of thumb described above which would yield less than 120 observations in most cases. If the initial sample is insufficient to precisely estimate the parameters of interest, sufficient additional samples are then drawn to supplement the first sample.

Regression estimation can be used to good effect in estimating load impacts for most DR resources.

2.2.2.4 Repeated Measures Designs

For event based resources it is possible to employ repeated measures designs. The availability of repeated measures of the outcome variable (i.e., hourly loads) is an interesting complication (and great advantage) in load impact estimation. When multiple events occur over a given period of time (e.g., critical peak days, interruptions, calls for curtailment) each conventionally sampled “point” (i.e., customer) actually produces multiple observations of the resource impacts (hourly loads). In effect, the study design that is being undertaken is a panel in which repeated measurements are taken over some number of time periods.

To talk about this sort of study design, one must distinguish between two kinds of measurements – cross-sectional measurements and time series measurements.

²⁹ *Using Multivariate Statistics (3rd ed.)*, Tabachnick, B. G., & Fidell, L. S. New York: Harper Collins (1996).

APPENDIX 2: SAMPLING OVERVIEW

Repeated measures study designs typically have both kinds of measurements. The cross-sectional measurements are those that vary over customers but not over time – things like location, customer type and income. Time series measurements are those that vary over time within a given member of the cross-section. These are variables like energy use, cooling degree hours, day of week, season and whether a DR event has been called.

Variation in customer loads arises out of variation in factors in the cross-section and out of variation in factors in the time series. For example, in a given hour, one customer in the cross section might use 2 kWh of energy while another might use 4 kWh. Such a difference could be because one of the customers has twice the air conditioner capacity of the other or it might be because one of the customers has a chest freezer in the garage and is charging the battery on their electric car during the time the energy use is observed. The sources of variation among customers that account for these differences are numerous and some are very difficult to measure. From hour to hour for any given customer, the loads also vary as a result of factors that are changing with time – factors such as season, day of week, temperature, occupancy patterns, and whether or not a DR event is called, etc. Some of these are also difficult to measure.

Because observations are being made across the variables in the cross-section and over time, it is possible with repeated measures designs to isolate the effects of cross-sectional and time series variables. In particular, it is possible to observe the main effect of a DR resource in isolation from the cross sectional variation and to observe the interaction between the DR resource and the cross sectional variables of interest. These can be used to produce a very powerful predictive model of the load impacts of event based DR resources.

Repeated measures designs offer several powerful advantages.

- These designs are statistically much more powerful than conventional designs in which a single observation is taken per sampled point. That is, much smaller cross-sectional samples can be used to estimate average load impacts than would otherwise be necessary.
- There is typically no need for a control group in estimating load impacts because load impacts for sampled units (e.g., households, firms, etc.) can be estimated as the difference between loads for “event” days and “non-event” days for each sampled unit. This eliminates the attendant risks of selection bias in comparing volunteers in the DR resource with those who have not volunteered in the general population of interest;
- The potential for estimation bias arising from fixed omitted variables in the estimation equation can be completely eliminated; and
- Variation in load measurements arising from factors in the cross-section can be isolated and accurately described.

APPENDIX 2: SAMPLING OVERVIEW

The conventional sample design techniques discussed under simple random sampling and stratified random sampling provide no basis for selecting an appropriate sample size for this sort of study because they are based on the notion that the sampled observations are independent of one another. The observations within the time series are not.

The sampling precision in a repeated measures design is a function of the size of the cross-section, the number of repeated measurements that occur and the correlation between the measurements. All other things being equal, sampling precision and statistical power increase significantly as the number of measurements increases. For DR resources involving six to ten events per season, sampling precision can be increased very dramatically – making it possible to detect relatively small effects (i.e., load reductions in the range of 5-10%) with only a few hundred observations. A good example of the analysis of repeated measures to observe relatively small load impacts is the SPP.

It is possible to calculate the sample size required to detect effects of a given size with repeated measurements in time given the:

- mean of the variable of interest;
- standard deviation of the variable of interest;
- number of repeated measurements by type (event and non-event);
- the number of groups in the analysis;
- acceptable probability of Type II error (alpha);
- desired power of the statistical test;
- correlation between measurements in the time series (rho);
- type of model used to estimate impact (e.g., Pre/Post, Change, ANOVA or ANCOVA); and
- minimum effect size that is to be detected.

A procedure for making this calculation is available in STATA's sampsi program.³⁰

It is possible to use information from load research samples to estimate the parameters that are required to calculate the sample sizes necessary to undertake a repeated measures study. In general, this will be the minimum sample size required to estimate the load impacts of the DR resource.

³⁰ See Frison and Pocock (1992) "Repeated measures in clinical trials: An analysis using mean summary statistics and its implications for design", in *Statistics in Medicine* 11: 1685-1704 for a technical discussion of the method used to estimate the impacts of repeated measures on sampling precision and sample size.

APPENDIX 2: SAMPLING OVERVIEW

Sample sizes calculated in this way do not include any provision for estimating the effects of the interactions of cross-sectional variables with the treatment effect. Accounting for the effects of the cross-sectional variables on the load impact will in most cases require additional samples. There are two reasons for this. First, the effect size specified in the sample design calculation must be reduced substantially if the effect sizes for the interactions are to be observed because interacting the cross sectional variables with the treatment will, in effect, decompose the treatment effect into smaller pieces (effects). Second, to observe the effects of the cross sectional variables it will be necessary to ensure that these variables have sufficient variation to permit regression type estimation.

If the effects of cross-sectional variables are to be included in repeated measures calculations it is probably more appropriate to employ sample sizes that would be required to estimate cross sectional effects in regression models (i.e., stratified random sampling).

2.2 CONCLUSIONS

Sampling adds uncertainty about the accuracy, precision and reliability of load impact estimates. When interval load data is available for the entire population(s) under study, evaluators should consider using it to avoid these sources of uncertainty. However, there may be instances where using data for the entire population might be impractical and sampling will be the appropriate method for observing DR load impacts. This will be true for mass market resources where interval metered data is not available for all population members. The use of sampling may be desirable even when information is available for a large mass market resource because a more focused effort on a properly designed sample can produce more accurate information than may be available through an attempt to analyze the information for the entire population.

When sampling is used care must be taken to ensure that it is representative of the population of interest and that it is sufficiently precise to meet the needs of the various stakeholders. There are well accepted sampling techniques that should be used whenever sampling is employed. These include: random sampling from the populations of interest and stratifying the random sample to achieve an acceptable level of statistical precision.

In most cases, stratified random sampling will be required for DR resource evaluations because it will be necessary to precisely estimate load impacts for important subsets of the populations under study (e.g., by utility service territories, weather zones and customer types defined in various ways). It may also be necessary to stratify samples by usage or other variables representing customer size in order to achieve acceptable sampling precision within budget limitations. Whenever stratified random samples are used, care must be taken to consider the impacts that sample weighting will have on subsequent analyses and to make sure that sampling weights are appropriately applied when summary measures for the population are calculated.

Efficiency gains arising from regression based estimators and repeated measures designs will generally favour the use of these analysis techniques in DR load impact

APPENDIX 2: SAMPLING OVERVIEW

estimation. Sampling to support the use of these techniques is not straightforward. It is possible in both cases to use either simple random sampling or stratified random sampling to establish appropriate sample sizes for DR load impact evaluations. Sample sizes established using these procedures will be conservative since the effects of the covariates and repeated measures will only serve to make the measurements more precise.

The most robust approach to estimating the sample size required for regression modeling presupposes an understanding of the variation in the customer loads in the population of customers under study; and the relationship between those loads and the factors that are being considered for use as control variables. In some cases, this information is available from prior studies (e.g., SSP) or from load research samples. Whenever such information is available, it should be used to identify an appropriate sample size required to support the analysis. If this information is not available, the sample design should be developed using conventional stratified random sampling techniques (i.e., those that only require information about the population mean and standard deviation within strata).

There are well developed procedures for establishing sample sizes for repeated measures studies used in experiments and clinical trials. An important determinant of the sample size required in a repeated measures design is whether interactions between cross-sectional variables and the effect of the resource have to be estimated. If this is not required, then the sample can be designed using the simple procedures that are appropriate for establishing sample sizes for clinical trials and experiments. On the other hand, if the interactions of the cross-sectional variables are to be described, it is probably more appropriate to employ sample sizes that would be required to estimate cross-sectional effects in regression models. The resulting sample size will be larger than what is possible with a repeated measures design, but will ensure that the cross section is large enough and diverse enough to estimate the cross-sectional effects.

Given the complexity of the analysis procedures used in DR load impact estimation, evaluators are advised to consult with a qualified and experienced survey statistician in developing sample designs to be used in DR load impact estimation. This is particularly true if significant resources will be expended installing meters and surveying customers.

APPENDIX 3: IMPACT EVALUATION METHODS

APPENDIX 3: IMPACT EVALUATION METHODS

As discussed in Section B, DR load impacts can be estimated as the difference between what a customer would have used in the absence of the DR program, referred to as the reference load, and either what they actually used (as determined by metered load) or an estimate of what they used when a DR incentive was in effect. Reference load estimates can be based on observing what participating customers used under conditions similar to event-day conditions but without the DR incentive in place (e.g., a day with similar weather to a DR event day) or they can be based on the load on an event day from a sample of similar customers that don't receive the DR incentive (that is, a suitable control group). Basing estimates on the load for participating customers on non-event days avoids the problem of selection bias that can accompany methods that rely on external control groups for load estimation. On the other hand, an external control group may be necessary to estimate impacts for DR options that are not event-driven (e.g., TOU or RTP pricing, permanent load shifting options such as ice storage) since, in these cases, participant data covering a period when the DR stimulus is not present may not be available.

Below is a very brief summary of several methods that can be used for load impact estimation. No single approach is best for all circumstances and it is not the intent of these protocols or guidelines to dictate the method that should be used for evaluation purposes. Having said that, as long as there is sufficient data and/or event history to allow its use, regression estimation, which is discussed at greater length in Appendix 4, is the leading method for both ex post and ex ante estimation for resource planning and cost-effectiveness analysis. Day-matching, or baseline, methods are not suitable for ex ante impact estimation, but are popular for determining load impacts for customer settlement and may be suitable for ex post impact estimation. While load impact estimation for settlement is not the focus of these protocols, we have included a very brief summary of day-matching methods because, for many stakeholders, these are the only methods with which they are familiar. We also briefly mention several other methods that do not have widespread applicability but that may be suitable for selected DR programs.

3.1 REGRESSION ANALYSIS

Regression methods rely on statistical analysis to develop a mathematical model summarizing the relationship between a variable of interest, known as the dependent variable, and other variables, known as independent or explanatory variables, that influence the dependent variable. When used to determine DR impacts, the dependent variable is typically either energy use³¹ or the change in energy use, and the independent variables can include a variety of factors such as weather, participant characteristics and behaviour and, most importantly, variables representing the influence of the DR resource.

An important factor to keep in mind when using regression analysis is that the goal is to do the best possible job of estimating DR resource impacts, not necessarily to develop

³¹ Some model specifications use ratios of energy use in different time periods as a dependent variable.

APPENDIX 3: IMPACT EVALUATION METHODS

the best model for predicting energy usage. Put another way, impact estimates depend not on the power of the model to predict energy use, but on the accuracy, stability, and precision of the coefficient that represents the change in energy use in response to DR program incentives. A model of energy use as a function of DR resource characteristics and other explanatory variables might have a low R-squared value (a measure of the explanatory power of the model), but very high t-statistics on the DR characteristics variables, meaning that it may explain the impact of the DR resource quite well even if it does not predict overall energy use that well.

When using regression analysis to estimate load impacts for event-based DR resources, the repetitive nature of event-based resources may allow the analysis to be conducted using smaller samples than would be needed for non-event based resources. This repeated measures characteristic typically eliminates the need for external control groups since days on which events are not called that have characteristics similar to event days allow the participant group to act as its own control. This feature also makes it possible to develop customer-specific impact estimates, thus affording the opportunity to examine the distribution of impacts across the participant population.

Regression modeling is the most robust and flexible approach to DR load impact estimation and should be considered the default option for the majority of applications, especially when ex ante impact estimates are the primary objective. While regression modeling requires more skill and experience to implement, and is not as transparent as most day-matching methods, it offers numerous advantages compared with other methods. For example, regression analysis can be used to examine impacts outside the event period and to quantify the influence of event characteristics, heat build up, multi-day events, weather and customer characteristics on demand response.

Several regression approaches and techniques have been used to develop load impact estimates. These include individual customer time series regressions, aggregate time series regressions, panel regressions, and hierarchical linear models. In some instances, data structure issues present specific challenges that require specific techniques (e.g. prevalence of zero values for sub-metered air conditioner data). Experienced evaluators must take into account the advantages and disadvantages of various techniques in light of the challenges presented by data limitations, inherent complexities resulting from the nature of the relationships being modeled and the output requirements of interest. No single model or approach is suitable in all situations.

3.2 DAY-MATCHING (BASELINE) METHODS

With day matching methods, the reference load used to determine load impacts is estimated by calculating the average usage in relevant hours for selected days leading up to an event day. For example, the reference load for a particular event day for the first hour of the event period might be determined by taking the average load in that hour for the top 3 out of the prior 5 load days. This method can only be used to calculate impacts for event-based resources, not for non-event based resources such as static TOU rates or permanent load shifting.

APPENDIX 3: IMPACT EVALUATION METHODS

The first step in developing reference load shapes involves selecting relevant days. Second, an average of the load in each hour for the days that are chosen is computed. If loads vary with weather or other observable factors, a third step that can improve the reference load shape involves making “same day” adjustments to the initial load estimates. These adjustments can be based on differences between load in hours outside the event period on prior days and load during the same hours on the event day or on differences in the value of other variables such as weather on prior days and event days. Below is a list of methods that have been used or tested in the past. This list is intended to be exemplary, not a complete census of all options:

- Previous 3, 5, 7 or 10 business days or weekdays
- Highest 10 out of 11 prior business days
- Highest 5 of the last 10 business days
- Highest 3 out of 10 prior business days with a same-day adjustment based on the two hours prior to the event period³²
- 20 days bracketing the event day
- All relevant days in an entire season.

For commercial and industrial customers, only business days are typically used to calculate the initial reference load. For residential customers, if events only occur on weekdays, weekends would logically be excluded from day selection. When it comes to using day matching, one size definitely does not fit all. What works best will vary by customer type, load shape, whether or not the load is weather sensitive, and other factors.

Day matching methods are easy to understand and often easier to produce and use than regression methods. If the primary question is, “What was the DR impact for yesterday’s event,” day matching can be an intuitively appealing and practical approach. However, day matching is not a suitable approach when the primary focus is on *ex ante* estimation for day types that differ from those that have occurred historically.

3.3 SUB-METERING

Another approach to load impact estimation involves sub-metering. Sub-metering is primarily useful in situations where the load contributing to demand response is relatively easy to isolate without rewiring or other costly procedures. An example is when load response is associated with a single piece of end-use equipment (e.g., an air conditioner, pump or other large motor).

³² This reference method is discussed in a recent LBNL report, *Estimating DR Load Impacts: Evaluation of Baseline Load Models for Commercial Buildings in California*, July 2, 2007.

APPENDIX 3: IMPACT EVALUATION METHODS

If the isolated equipment is always on except when interrupted for an event, sub-metering will provide a very accurate estimate of load impact by simply comparing load just prior to and after the beginning of an event period. If the equipment has a duty cycle, and one that differs across days due to variation in weather or some other variable, there will still be a need to develop a reference load shape or, alternatively, use regression analysis to predict the “but for” load. However, this task will typically be much simpler when the data being used reflects the only relevant load rather than total premise load. Sub-metering may be necessary if there is significant variability in premise load and the DR impact is small relative to total premise load. In these circumstances, day matching and regression analysis are unlikely to generate statistically significant impact estimates, even if the load reduction is reasonably large in absolute terms (but not relative to the total premise load).

3.4 ENGINEERING ANALYSIS

Another method that might be useful in limited situations is engineering analysis. Engineering analysis is much less useful for estimating load impacts for DR programs than it is for EE programs because impacts are driven much more by consumer behavior than by technology implementation. Even some technology enabled DR programs, such as those using programmable communicating thermostats, have a strong behavioral component since consumers can vary the automated set point and/or override the predetermined setting whenever they wish. For very large loads, there may be situations where a program administrator has direct control over the equipment for emergency purposes, thus eliminating any behavioral influence. Under these circumstances, engineering analysis might produce accurate impact estimates, but these loads are likely to be sub-metered so that impacts can be measured directly.

An example where engineering analysis might be useful would be if a program targeted continuously running pumps and the pumps were remotely controlled during DR events. In this case, one could conduct a survey to gather information on the horsepower associated with each pump and use simple engineering calculations to convert that data into estimates of connected load. DR impacts could then be calculated based on the control strategy that was used for each event. However, this somewhat contrived example may have little practical value as these circumstances are rare.

3.5 DUTY-CYCLE ANALYSIS

Another approach to load impact estimation is to combine end-use metering with engineering calculations. This approach was employed in the evaluation of SCE’s air conditioning cycling program for residential customers, and termed the *Duty Cycle Approach*.³³ The approach is designed to take into account the fact that load cycling impacts vary across program participants by temperature, hour of day, size of air conditioner, and the share of time the air conditioner is in operation (the duty cycle). The duty cycle approach is designed to create a reference value for A/C load by collecting data on the total connected load for each enrolled participant, and the share of connected

³³ Quantum Consulting Inc. *The Air Conditioner Cycling Summer Discount Program Evaluation Study*. January 2006. See also George, Bode and Schellenberg, Ibid.

APPENDIX 3: IMPACT EVALUATION METHODS

load utilized by hour of day and temperature bin (for non-event days). The specific load impacts are then calculated by:

- Identifying the average share of connected load utilized during the appropriate temperature and time bins (average duty cycle), and
- Calculating program load impacts by taking into account the average duty cycle, total connected load of each participant, participant cycling selections, and the cycling device failure rate.

Importantly, the approach is able to provide load impact estimates for both ex post and ex ante scenarios as well as information about the uncertainty of those estimates.

3.6 OPERATIONAL EXPERIMENTATION

Still another approach to impact estimation for event-based programs involves the use of what might be called operational experimentation. By operational experimentation, we mean the selective exercise of a program on a sub-sample of participants for the sole or primary purpose of generating data for impact estimation. This is perhaps best understood with an example constructed once again around an air conditioner cycling program.

Given the typically large number of customers participating in load control programs, there are plenty of customers from which a small sample can be drawn for experimental purposes. One could split this sample into two groups using random sampling and either install an interval meter on the whole house or on the air conditioning unit to obtain the data necessary to determine load impacts. With the metering in place, one could experiment with different load control strategies and event windows across a variety of day types to generate a database that would allow you to estimate impacts under various conditions. The control and treatment groups could be alternated to ensure that there is no correlation between customer characteristics and impacts. Given that this approach provides data on both a control and treatment group on event days, a simple comparison of means on event days would provide a valid estimate of average impacts. However, if ex ante estimates are needed, regression analysis would be required. Operational experimentation would be very cost-effective and straightforward if interval meters were already in place (as they ultimately will be in Ontario), and if incentives are largely fixed (that is, if customer payments are not event-specific). This approach could be quite useful for relatively new DR programs or even for long-standing emergency programs that are not triggered very frequently. In these situations, there may not be sufficient data on event days to estimate impacts using other methods.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

Regression analysis is the recommended method for estimating the impact of DR resources in most instances for ex ante estimation, although data limitations or lack of event history could dictate that alternative methods are needed either as a complement to or substitute for regression analysis. Regression methods rely on statistical analysis to develop a mathematical model summarizing the relationship between a variable of interest, known as the dependent variable, and other variables, known as independent or explanatory variables, that influence the dependent variable. Typically, regression models include several variables such as:

- Hourly and day-of-week variables that reflect customers' average load shapes absent curtailments;
- Variables designed to explain variation in load patterns such as weather, electricity prices, and seasonal usage;
- Variables designed to quantify the average and variation in load impacts,

When used to determine DR impacts, the dependent variable is typically either energy use³⁴ or the change in energy use, and the independent variables can include a range of influencing factors such as weather, participant characteristics and, most importantly, variables representing the influence of the DR resource. A very simple regression model that relates energy use to temperature and a variable representing the presence or absence of a DR resource event is depicted in Equation 4-1.

$$E_i = a + bT_i + c(T_i)(D_i) + e_i \quad (4-1)$$

where E_i = energy use in hour i

T_i = the temperature in hour i

D_i = the resource variable, equal to 1 when an event is triggered in hour i , 0 otherwise

e = the regression error term

a = a constant term

b = the change in load given a change in temperature

c = the change in load given a change in temperature when a DR event is triggered.

³⁴ Some model specifications use ratios of energy use in different time periods as a dependent variable.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

Regression modeling can be complicated and it requires strong training in statistics and econometrics. There are many different approaches to regression modeling that vary with respect to the general method used (e.g., classical versus Bayesian), estimation algorithms (e.g., Ordinary Least Squares, Generalized Least Squares, Maximum Likelihood Estimation), functional specification (e.g., conditional demand analysis, change modeling, etc.), the use of control groups (e.g., participants versus non-participants), and the variables that are explicitly included in the model specification. No single approach will be best in all situations. Indeed, the primary objective of regression-based methods for impact estimation is to choose the method that works best for the application at hand, and to justify that choice. There is both an art and science to regression modeling and there is no substitute for a skilled professional when it comes to the successful application of regression-based methods to DR impact estimation.

4.1 OVERVIEW OF REGRESSION ANALYSIS

A very useful overview of regression modeling, including a discussion of the many technical issues that must be considered when developing regression models, is contained in *The California Evaluation Framework*.³⁵ This is a good starting point for readers who want a general understanding of some of the options and challenges associated with regression modeling. However, neither that document nor anything said here is intended to be a “how to guide” for using regression analysis for impact estimation.

An important factor to keep in mind when using regression analysis is that the goal is to do the best possible job estimating DR resource impacts, not necessarily to develop the best model for predicting energy usage. This point is expressed well in *The California Evaluation Report* (p. 115), where it states,

“It is important to recognize that energy savings estimates depend not on the predictive power of the model on energy use, but on the accuracy, stability, and precision of the coefficient that represents energy savings.”

A model of energy use as a function of DR resource characteristics and other explanatory variables might have a low R-squared (a measure of the explanatory power of the model), but a very high t-statistic on the DR characteristics variables, meaning that it may explain the impact of the DR resource quite well even if it does not predict overall energy use that well.

Most of the work that econometricians do is intended to test whether the key assumptions of the estimator employed are valid, and if not, apply the appropriate corrections or alternative estimation methodologies to acquire accurate, stable, and precise load impacts. Errors in applying econometric methods can lead to:

- Biased estimates of load impacts;
- Imprecise estimates of the level of confidence that can be placed on the results;

³⁵ TecMarket Works. *The California Evaluation Framework*, June 2004. pp. 105 – 120.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

- The inability to mathematically find a solution.

For load impacts, both unbiased estimates and correct portrayals of the uncertainty around those estimates are not only desirable, but necessary.

Table 4-1 identifies potential problems in regression modeling that can influence either the accuracy (lack of bias) or the estimated certainty of the load impacts. It is not intended to be an all inclusive list of potential regression pathologies. Rather, it highlights some of those that can be most damaging to estimating DR impacts using regression methods. Some of the statistics required by Protocol 6 are intended to reveal the extent to which many of these issues have been addressed.

Table 4-1
Issues in Regression Analysis

Problems that potentially bias estimates	Problems that lead to incorrect standard errors
1. Omitted Variable: This is a type of specification error. Omitted variables that are related to the dependent variable are picked up in the error term. If correlated with explanatory variables representing the load impacts, they will bias the parameter estimates.	1. Serial-Correlation: Also known as auto-correlation, this occurs when the error term for an observation is correlated with the error term in another observation. This can occur in any study where the order of the observations has some meaning. Although it occurs most frequently with time-series data, it can also be due to spatial factors and clustering (i.e., the error terms of individual customers are correlated).
3. Improper functional form: This occurs when the relationship of an explanatory variable to the dependent variable is incorrectly specified. For example, the function may be treating the variable as linear when, in fact, it is logarithmic. This type of error can lead to incorrect predictions of load impacts.	2. Heteroscedasticity: This occurs when the variance is not constant but is related to a continuous variable. Depending on the model, if unaccounted for, it can lead to incorrect inferences of the uncertainty of the estimates
4. Simultaneity: Otherwise known as endogeneity, this occurs when the dependent variable influences an explanatory variable. This is unlikely to be a problem in modeling load impacts.	3. Irrelevant Variables: When irrelevant variables are introduced into a model, they generally weaken the standard errors of the explanatory variables related to the dependent variable. This leads to overstating the uncertainty associated with the impacts of other explanatory variables.
5. Errors in Variables: Explanatory variables that contain measurement error can create bias if the measurement error is correlated with explanatory variables(s).	
6. Influential data: A data point is considered influential if deleting it changes the parameter estimates. Influential variables are typically outliers with leverage. These are more of an issue with large C&I customers.	

Importantly, a large number of the problems that lead to potential bias are due to model misspecification and the closely related phenomena of correlations between the error

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

terms and the explanatory variables. Despite a large set of diagnostic tools, it is difficult to write down a set of rules that can be used to guide model specification, especially since the best approach for model specification is not a settled question. This is where the art of regression analysis comes into play, making the experience and knowledge base of evaluators and reviewers critical.

Typically, DR load impact analysis involves both a time series and a cross-sectional dimension. This type of data is referred to by a variety of names – including time series cross-sectional, panel, longitudinal, and repeated measures data. With this type of data, evaluators are able to account for a significant share of omitted variables, including those that are unobservable or not recorded, leading to better specified, more robust regression models.

Panel regressions can control for omitted and sometimes unobserved factors that vary across individuals but are fixed over the course of the study (fixed effects – e.g. household size, income, appliance holdings, etc.), and for factors that are fixed for all customers but vary over time (time effects -economic conditions). Regression-like models that can be used to analyze panel data include ANOVA, ANCOVA, and MANOVA. These models are similar in that they allow each individual to act as their own control and account for the effects of the fixed, but unmeasured characteristics of each customer.

However, the ability to control for fixed effects comes at a price. First, panel regressions typically calculate the average effect of DR programs, although variation in the average impact across a limited number of dimensions can be estimated through interactions with impact variables. In contrast, other approaches, such as individual customer time series or hierarchical linear models, can better identify the variation in load response among participants or across different conditions. Second, by controlling for fixed effects, these models cannot incorporate the impact of explanatory variables that are time-invariant (e.g., air conditioning ownership) except through interactions with time-variant variables (e.g. temperature). In other words, a fixed effects model only controls for the variation within individual units; it does not control for the variation across individual units. In many instances, impact evaluations will need to take into account how fixed characteristics such as appliance holdings, household size, etc. affect the load response provided, requiring either:

- The use of interactions;
- A two-stage model, where load impacts for each customer are first estimated using individual regressions (or regressions for customer pools defined by criteria such as industry classification) followed by a second stage that regresses load impacts against customer characteristics;
- Using a random effects model which is able to use fixed characteristics as explanatory variables.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

Two additional topics that are particularly relevant when working with load data are auto-correlation and heteroscedasticity. Having both cross-sectional and time-series dimensions, there are multiple ways in which the errors can be related. Basic panel data methods generally assume:

- No correlation between the error terms of units in the same time period;
- No correlation across units in different time periods;
- No auto-correlations within units over time;
- Constant variances over time within a unit (different variances across units are allowed).

Impact evaluations will most likely have to account for auto-correlation due to the prevalence of a time dimension in load impact data. However, it is important to distinguish between pure and impure auto-correlation. Impure auto-correlation can arise because of a specification error such as an omitted variable or incorrect functional form. Pure auto-correlation is the correlation that is still present when the model is properly specified. This implies that auto-correlation should be viewed as more than a nuisance to be corrected, but as a signal to further explore the potentially larger problem of misspecification. Correcting the standard errors due to auto-correlation is straightforward and there are a number of options for addressing it, including first differencing, Generalized Least Squares, and the use of Maximum Likelihood estimation that does not assume an error matrix with constant diagonals and zero values in the off-diagonals.

Only heteroscedasticity within individual units is problematic in panel data, although when faced with large variations in customer size and impacts, the evaluator should consider transforming the data to a common metric such as the percent change in load. While heteroscedasticity can typically be corrected for using of robust standard errors – also known as Huber-White standard errors and the sandwich standard errors – they do not apply if serial correlation is present³⁶. Because of this, the more labour intensive process of testing for heteroscedasticity, determining the specific form of heteroscedasticity, and applying the appropriate data transformation may often be required to identify and correct for heteroscedasticity within units.

Difficulties in estimating load impacts using regression analysis can also result from variation (or lack thereof) in load. For example, it may be difficult to estimate load impacts if there is a large degree of variation in energy use that can't be explained by variation in observable variables and the DR impact is small relative to the total load. This can occur if data on the independent variables that drive this variation is difficult to obtain, as it could be with industrial customers where variation may be caused by industrial process operations that are hard to measure. If the DR impact is small relative to the normal variation in energy use, and that variation in energy use can not be explained, it will be

³⁶ Page 274-276 of Jeffrey Wooldridge's textbook "Econometric Analysis of Cross-section and Panel Data" provides an excellent discussion on serial correlation and the robust variance matrix estimator.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

very difficult for the regression analysis to isolate changes in energy use due to the DR resource from the unexplained variation in energy use due to other factors.

In contrast to the situation where too much variation creates estimation difficulties is the case where there is too little day-to-day variation in load. For example, with loads that are not at all weather sensitive and, as a result, may not vary much from day-to-day, there may not be much of an advantage in using regression analysis over less complicated and easier to understand methods such as day matching. In these circumstances, regression analysis may be effective for estimating the impact of the DR event, but that impact wouldn't be expected to change from one event to another in response to variation in other observable factors such as weather. As such, one of the primary benefits of regression analysis, the ability to make ex ante estimates for day types or other conditions that differ from the past, is no longer relevant. Given this, if some participants in a DR resource have weather sensitive loads, or loads that vary with other observable variables, while other participants have loads that vary very little, using regression modeling to estimate impacts for the variable segment and day-matching to estimate impacts for the non-variable segment may be the best strategy. In these circumstances, using a regression model to estimate the impacts for both types of customers may distort the impacts associated with the market segment with the variable load.³⁷ It could also distort ex ante estimate if future participation by the two segments is not proportional to that of the ex post group of participants.

Difficulties can also arise from limited variation in event conditions. This is particularly true for resources, such as emergency programs, that are dispatched on a limited basis. In these instances, ex-post load impacts can typically be calculated accurately, but it is often difficult to estimate ex-ante load response due to insufficient variation in historical event conditions that would support extrapolation to the desired ex ante conditions. In such situations, evaluators should explicitly acknowledge the data limitation and, where appropriate, apply alternative techniques to produce ex-ante load impacts. For example, in the case of air conditioner cycling, engineering duty cycle analysis may provide a basis for estimating ex ante load impacts.

4.2 THE ADVANTAGE OF REPEATED MEASURES

One of the interesting and useful characteristics of event based resources that differs from the typical situation with both EE evaluation and the evaluation of non-event based DR resources is the fact that you are typically able to observe the impact of the DR resource multiple times for the same customer. For an energy efficiency resource or for non-event based DR resources, if you have usage data before a customer enrolls in a DR resource option, even if you have daily or hourly usage data, you only have two time periods per customer in which the DR resource variable(s) differs, one before enrolment and one after. If there is no pre-treatment data, you only have one time period for each customer (in which case a suitable control group is needed in order to statistically estimate the impact of the DR resource). However, with event-based resource options,

³⁷ In this instance, separate output tables should be reported for each market segment.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

you get multiple observations for each customer over which the DR incentive either is or is not in effect. For example, if you have twelve days in a year in which a CPP day is called, you have 12 days on which the DR incentive is in effect, and many more days in which it is not.

The repeated measure effect associated with event-based DR resources has several significant advantages for impact evaluation compared with non-event based resources. One concerns sampling efficiency. As discussed in Appendix 2, with repeated measures, you may be able to use much smaller sample sizes to achieve the same level of statistical precision. The reduction in sample size is a function of the expected impact size, the coefficient of variation and the number of repeated measures that occur, but a 10-fold decrease may be possible compared with a simple comparison of means using before-and-after data on participants or side-by-side data with participant and control samples.

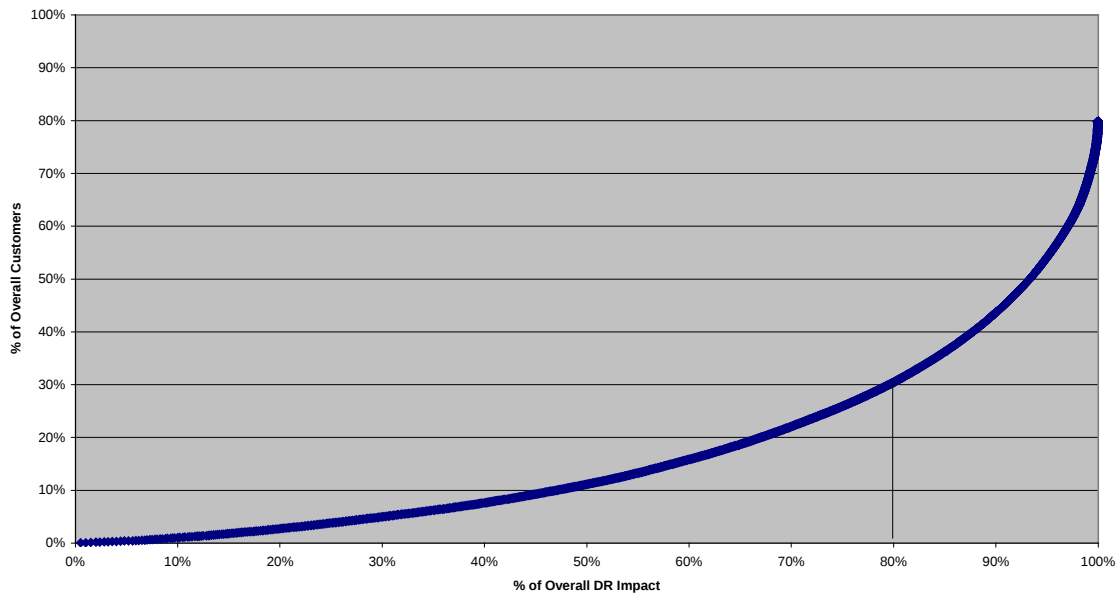
A second advantage of the repeated measure effect associated with event-based resources is that impact estimation typically does not require an external control group.³⁸ The fact that the DR resource incentive is in effect on some days and not on others allows you to estimate the influence of variation in factors that change daily, such as weather, along with the influence of the DR resource. This, in turn, allows you to estimate the impact of the DR resource on any day type that can be characterized in terms of the explanatory variables included in the model without needing a sample of customers who do not participate in the resource. This eliminates any concern about internal validity, as there is no opportunity for differences between control and treatment groups to generate biased estimates. This is a significant advantage as long as your primary interest is in estimating impacts for a set of volunteers behaviourally similar to those who have participated to date.³⁹

A third advantage associated with the repeated measures property of event-based resources is that it allows you to estimate customer-specific regressions. For example, a regression model like the very simple specification shown earlier in Equation 2-1, could be estimated for each individual customer. This would allow you to understand the distribution of impacts across customers, which can be quite useful from a policy perspective, since it allows one to determine if the average impact is more or less typical, or, alternatively, if a relatively small percentage of customers account for the majority of demand response. For example, this type of analysis based on the SPP data produced the distribution of demand response impacts shown in Figure 4-1, indicating that roughly 80 percent of total demand response was provided by roughly 30 percent of participants.

³⁸ There are situations in which an external control group might still be needed. For example, if an event is only called on the hottest days of the year, and the relationship between energy use on those days is different from what it is on other days, the model may not be able to accurately estimate resource impacts on event days. In this instance, it may be necessary to have a control group in order to accurately model the relationship between weather and energy use on the hottest days in order to obtain an unbiased estimate of the impact of the resource on those day types.

³⁹ There may still be some interest in knowing how participants differ from non-participants if there is a need to extrapolate the impact estimates to a population of customers who are unlikely to volunteer (which may differ from those who have not yet volunteered). If so, an external control group may be needed.

Figure 4-1
Percent Demand Response Impact Relative to Percent Population
California's Statewide Pricing Pilot



A final advantage associated with repeated measures for a cross-section of customers is the ability to better specify regression equations and to produce more robust results.⁴⁰ Regressions that have observations over time and across customers can control for omitted variables that vary across customers but are fixed over the study period, known as fixed effects, and for omitted variables that are fixed across customers but vary over time, known as time effects.

4.3 QUANTIFYING THE IMPACT OF EVENT CHARACTERISTICS

One of the primary advantages of regression analysis is the ability to determine the impact of various factors on demand response. One important set of factors is the event characteristics. Notification lead time and the timing and duration of events may influence demand response for resources in which these factors are allowed to vary across events or across customers (e.g., as in cafeteria style resources). The ability to do this is a function of how much these characteristics vary over the estimation time period or across customers. Given sufficient variation, it is relatively straightforward to include interaction terms in the regression model to determine if impacts vary with these event characteristics. For example, it might be possible to define a set of binary variables representing different event periods (e.g., a variable equal to 1 if the event period is less than 3 hours, 0 otherwise). An alternative would be to calculate individual customer impacts for each event day and analyze patterns and drivers of demand response by specifying a second-stage regression that relates load impacts to specific customer and

⁴⁰ Peter Kennedy, in *Guide to Econometrics*, provides an excellent discussion of some of the advantages of having repeated measures across a cross-section of customers in the introduction to Chapter 17.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

event day characteristics. These types of specifications would allow development of ex ante estimates for specific combinations of event conditions that did not occur in the past. This could be quite useful for operational purposes or for longer term resource planning or resource design.

4.4 ESTIMATING IMPACTS FOR HOURS OUTSIDE OF THE EVENT PERIOD

As indicated in Protocol 2, impact estimates for event based resources are required for all hours on an event day. This requirement fulfills the need to understand the extent and nature of load shifting that occurs with some types of DR resources, and to estimate the impact of DR resources on overall energy use. Regression modeling can be used to estimate all of these impact types using a variable representing an event day, as distinct from a variable representing an event window, interacted with variables representing individual hours in a regression analysis that pools all hours in a single regression.

4.5 WEATHER EFFECTS

Accurately reflecting the influence of weather in load modeling and impact estimation is essential, both in order to normalize for day-to-day load variation during impact estimation as well as to develop estimates for day types with weather conditions that differ from those in the past. Incorporating weather into regression modeling is easily done using weather variables and interaction terms as illustrated in the simple model in Equation 4-1 and the example shown in Section 4.10 below.

A related factor is heat build up in buildings caused by multiple hot days in a row. This can also be reflected in a regression model, for example, using a variable representing cooling degree hours on days prior to an event day, or cumulative cooling degree hours leading up to the event period.

4.6 MULTI-DAY EVENTS

Another issue to consider when developing model specifications is variation in impacts across multi-day events. Distinct variables indicating whether an event is the first, second or third day of a multi-day event can be included in a regression specification to determine if impacts vary according to this event feature. Section 4.2 of the *Impact Evaluation of the California Statewide Pricing Pilot*⁴¹ provides an example of this type of specification.

4.7 PARTICIPANT CHARACTERISTICS

The influence of participant characteristics on load impacts can be determined using interaction terms between variables representing customer characteristics, such as air conditioning and/or other equipment ownership, and socio-demographic or firmographic variables such as income, persons per household, business type and others. This capability is essential for predicting how impacts might change as the mix of participant characteristics changes. We mention this here because it is important to consider the need for ex ante estimates when developing a model specification designed to do both ex

⁴¹ Stephen S. George and Ahmad Faruqui. *Impact Evaluation of the California Statewide Pricing Pilot*. Final Report. March 16, 2005. p. 66.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

post and ex ante estimation. It might not be necessary to include socio-demographic variables in the model if only ex post estimates are needed, since fixed or variable-effects specifications can control for variation in energy use across customers without explicitly including such variables in the model. However, if ex ante estimation is needed, it will be necessary to explicitly incorporate variables in the specification that are expected to change in the future.

4.8 GEOGRAPHIC SPECIFICITY

Knowing how impacts vary across regions can be very useful for transmission and distribution planning and for operational dispatch decisions by the IESO, who must balance supply and demand at numerous points on the grid. Understanding the extent to which impact estimates are required for specific locations is an important input to evaluation planning.

There are two basic approaches to developing location-specific impact estimates. One is to obtain large enough samples at each desired location to develop statistically valid and precise impact estimates based on each geographic sub-population. If the number of geographic regions is large, this could be a very costly approach.

An alternative approach is to incorporate variables in a regression model that explain how impacts vary according to weather and population characteristics that vary regionally. Using survey and climate data to develop estimates of the mean values for each explanatory variable by region, such a model can be used to predict what the impacts will be given the local conditions. It may be possible to implement this approach with data on a much smaller sample of customers than the location-specific sampling approach by using stratified sampling methods that ensure sufficient variation in the characteristics of interest to develop the model parameters.

4.9 SUMMARY

Regression modeling is the most robust and flexible approach to DR load impact estimation and should be considered the default option for the majority of applications. While regression modeling requires more skill and experience to implement, and is not as transparent as most day-matching methods, it offers numerous advantages compared with other methods. Regression analysis can be used to examine impacts outside the event period and to quantify the influence of event characteristics, heat build up, multi-day events, weather and customer characteristics on demand response.

The repetitive nature of event-based resources may allow for regression analysis (or other methods) to be implemented using smaller samples than would be needed for non-event based resources. It also eliminates the need for external samples in most situations, and allows customer-specific impact estimates to be developed, thus affording the opportunity to examine the distribution of impacts across the participant population.

Day matching methods can produce reasonably accurate ex post impact estimates and may be preferable for use in customer settlement. However, difficulties in estimating

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

uncertainty adjusted impact estimates and in developing ex ante estimates using day matching are significant shortcomings in many applications.

4.10 REGRESSION ANALYSIS: AN EXAMPLE

This section contains an excerpt from a recent study showing how regression analysis can be used to develop DR load impact estimates. The example is from Pacific Gas and Electric Company's (PG&E) SmartRate^{TM42} tariff. During the summer of 2008, SmartRate was offered to PG&E SmartMeter^{TM43} customers in the Bakersfield and greater Kern County region, a very hot area where maximum temperatures exceed 100°F on many summer days. The tariff was initially offered to customers that were on PG&E's E-1 and E-8 residential tariffs and the A-1 non-residential tariff, which applies to customers with peak demands below 200 kW. Direct mail materials were sent to roughly 135,000 accounts, of which more than 10,000 enrolled. Approximately 7.5 percent of residential customers and 5 percent of small commercial customers that received direct mail materials enrolled in the SmartRate program.

The SmartRate pricing structure is an overlay on top of a customer's otherwise applicable tariff. SmartRate pricing consists of an incremental charge that applies during the peak period on SmartDays and a per kilowatt-hour credit that applies for all other hours from June to September. For residential customers, the additional peak-period charge on SmartDays is 60 ¢/kWh. For non-residential customers, the incremental charge is 75 ¢/kWh. The credit consists of two parts. A credit of roughly 3 ¢/kWh applies to all electricity use other than use during the peak period on SmartDays during the months of June through September. An additional credit of 1 ¢/kWh applies to tier 3 and higher usage for residential customers regardless of time period.

PG&E called nine SmartDays during the summer of 2008. Three days were called in a row in early July (the 8th, 9th and 10th), three in August (27th, 28th and 29th) and three in September (3rd, 4th and 5th). The August and September event periods spanned the Labor Day weekend and covered six out of seven consecutive work days. Although PG&E initially called an event on August 14th, a problem that occurred with the notification process caused PG&E subsequently to cancel the event. Given the confusion over event notification, rather than treat August 14th as a non-event day, data for this date were dropped from the estimation sample.

The 2008 load impacts for the SmartRate tariff were estimated through individual customer time-series regressions. Time series regressions were estimated at the individual customer level rather than for all customers combined for several reasons. Most importantly, PG&E does not typically collect data on a key explanatory variable--the size and type of air conditioning at each household. That being said, by employing individual customer regressions, the presence and use of air conditioning is captured through the temperature variables and their interaction with the hourly binary variables.

⁴² Any use of the term SmartMeter or SmartRate in this document is intended to refer to the trademarked term, whether or not TM is included.

⁴³ SmartMeterTM is a trademark of SmartSynch, Inc. and is used by permission.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

Put differently, the presence of air conditioning or lack thereof is a fixed effect that interacts with weather. By allowing individual customer coefficients to vary, the results are more accurate at the customer level – an important feature when results are desired for various customer segments in addition to the average for all participants. In addition, individual customer regressions can be employed to describe accurately the distribution of customer load reductions as well as the distribution of percent load reductions.

The main regression alternatives, panel regressions and segmented aggregate time series, were not selected due to the unique features of the data and the evolving customer mix and enrolment rates over time. Unlike individual customer regressions, panel regressions can make use of both control groups and pre-enrolment data and can provide very robust average customer impact estimates by controlling for omitted variables.⁴⁴ While panel regression can increase the accuracy of the impact estimates for the average customer, it cannot be employed to describe meaningfully the distribution of impacts among the participant population. Importantly, the lack of data on the type and size of air conditioners at the customer level precluded the use of panel regression. Because air conditioning is a key driver of electricity demand that interacts with weather, its omission in a panel regression would likely lead to inaccurate results. The other alternative, running time series on customer load aggregated by segment, could not adequately control for the evolving customer mix or provide insights into the distribution of impacts among the participant population. Except for the lower amount of effort required, segmented time series did not yield methodological benefits that were not also captured through individual customer regressions.

The analysis of PG&E's SmartRate tariff was based on a proportional random sample of approximately 2,000 customers drawn from the participant population of roughly 10,000.⁴⁵ The dependent variable in each regression is average hourly demand (kW). The explanatory variables can be grouped into three main categories:

- Variables that reflect the average load shape of customers, absent the need for cooling;
- Variables that explain deviations in hourly usage from the average load shape; and
- Variables that estimate the change in energy use during event days and the factors that influence the load reductions.

The explanatory variables include hourly binary variables to capture the inherent variation in usage across hours of the day, day-of-week binary variables to capture variation in usage between week days and weekends and across weekdays, weather variables to capture the influence of temperature on electricity use, and event-day and event-hour variables to estimate the impact of the higher SmartDay prices on energy use during each

⁴⁴ Panel regression can account for omitted variables that are unique to customers but fixed over time (fixed effects) such as household income, and can also account for omitted variables that are common across the participant population but unique to specific time periods. They cannot, however, account for omitted variables that vary both by participant and by time period or for household characteristics (e.g., central air conditioning) that interact with variables that vary over time, such as weather.

⁴⁵ The exact number of participants and sample points vary by date as enrolment grew over time during the 2008 summer months.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

hour of the event period as well as hours leading up to and following the event period. The event variables are interacted with weather, day of week, month, number of consecutive event days, and cumulative number of events throughout the season in order to explain how the impacts vary as a result of changes in those conditions. For customers for whom event notification was unsuccessful or who elected not to be notified, the relevant event days were treated as non-event days in terms of the model specification. However, when the average and aggregate impact estimates were developed from the individual customer regressions, impacts for these non-notified customers were assumed to equal zero.

The model specification was intentionally designed to capture a wide variation of household operating schedules as well as different hourly responses to weather and event conditions. The specification performed well for most customers, although for specific customers, some of the parameters may have been irrelevant.⁴⁶

The regressions were estimated using generalized least squares (GLS) and Huber-White robust standard errors in order to ensure that the confidence bands around the impact variables were not overstated either due to auto-correlation or heteroscedasticity.⁴⁷ The following equation summarizes the model specification. Given the large number of regressions (e.g., 2,000), it was not feasible to customize regressions for each customer. Importantly, the model performed well in the aggregate, as shown below.

$$\begin{aligned}
 KW = & \alpha_0 + \sum_{i=2}^{24} \beta_i \cdot \text{HOUR}_i \cdot \text{NS} \cdot \text{WEEKDAY} + \sum_{i=2}^{24} \gamma_i \cdot \text{HOUR}_i \cdot \text{S} \cdot \text{WEEKDAY} + \sum_{i=2}^{24} \delta_i \cdot \text{HOUR}_i \cdot \text{WEEKEND} \cdot \\
 & + \sum_{j=7}^{10} \varphi_k \cdot \text{MONTH}_k + \sum_{i=1}^{24} \mu_i \cdot \text{HOUR}_i \cdot \text{CDH} + \sum_{i=1}^{24} \eta_i \cdot \text{HOUR}_i \cdot \text{CDH}^2 + \psi \cdot \text{S} \cdot \text{CDH} + \zeta \cdot \text{S} \cdot \text{CDH}^2 + \\
 & \sum_{i=2}^{24} \pi_i \cdot \text{HOUR}_i \cdot \text{EVENTDAY} + \sum_{i=1}^{24} \theta_i \cdot \text{HOUR}_i \cdot \text{EVENTDAY} \cdot \text{CDH} + \sum_{i=1}^{24} \lambda_i \cdot \text{HOUR}_i \cdot \text{EVENTDAY} \cdot \text{CDH}^2 + \\
 & \sum_{j=6}^{10} \varsigma_k \cdot \text{MONTH}_k \cdot \text{EVENT} + \sum_{k=1}^3 \upsilon_k \cdot \text{INAROW}_k \cdot \text{EVENT} + \omega \cdot \text{CUMEVENTS} \cdot \text{EVENT} + \\
 & \sum_{i=2}^7 \xi_k \cdot \text{DOW}_k \cdot \text{EVENT} + \varepsilon
 \end{aligned}$$

Where:

KW = Electricity usage in Hour i for Customer j

NS = No School (period during the summer when school is NOT in session)

S = Period during the summer when school is in session

⁴⁶ Irrelevant parameters can lead to wider standard errors, but do not bias the significant parameters. Given that the amount of observations per regression generally exceeded 2,000, statistical power was not a major concern.

⁴⁷ The GLS method used relied on the Prais-Winsten technique – a form of iterated GLS.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

WEEKDAY = Monday – Friday

WEEKEND = Saturday – Sunday

HOUR_i = Hours of the day, numbered 1-24

MONTH_j = Months of the year, numbered 1-12

CDH_i = Cooling Degree Hour for that hour of the day, defined as Max(0, Temperature(F) -70)

CDH² = CDH squared

EVENTDAY = SmartRate event day (all 24 hours)

EVENT= SmartRate event window (2-7 pm)

INAROW = Number of consecutive events in a row

CUMEVENTS= Cumulative number of events in season

DOW = Day of week

ϵ = the error term

i = Subscript indicating the hour of day (1-24)

j = Subscript indicating the month of the year (1-12)

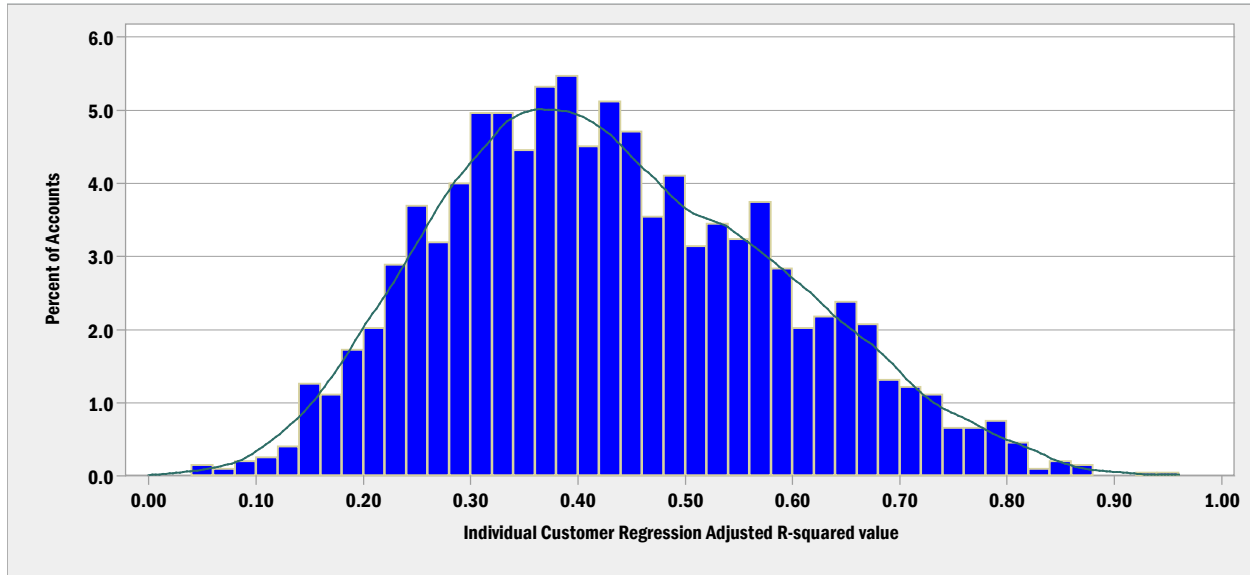
k = Subscript indicating the number of consecutive events in a row

l = Subscript indicating the day of week (1-7)

Although the regressions were developed at the individual customer level, from a policy standpoint, the focus is less on how the regressions perform for individual customers than it is on how the regressions perform for the average participant and for specific customer segments. Overall, individual customers exhibited more variation and less consistent energy use patterns than the aggregate participant population. Likewise, the regressions explained better the variation in electricity consumption and load impacts for the average customer (or average customer within a specific segment) than for individual customers. Put differently, it is more difficult to explain fully how a specific CARE customer behaves on an hourly basis than it is to explain how the average CARE customer behaves on an hourly basis. Because of this, we present measures of the explained variation, as described by the R-squared goodness-of-fit statistic, for the individual regressions and for specific segments as well as for the average customer.

Figure 4-2 shows the distribution of R-squared values from the individual residential customer regressions. As the peak period use, annual consumption, and ratio of summer to non-summer usage increase, the goodness-of-fit from the regressions generally improves. While the individual customer regressions do a reasonably good job of explaining the variation in electricity use, in aggregate, nearly all of the variation in energy use across hours is explained by the model specification.

Figure 4- 2
Distribution of Adjusted R-squared Values from Individual Regressions



When the predicted and actual values are aggregated across the individual results, the model explains 92.2 percent of the variation in energy use. Put another way, only about 8 percent of the variation in energy use over time is explained by variables that are not included in the model. In order to estimate the average customer R-squared values, the regression-predicted and actual electricity usage values were averaged across all customers for each date and hour. This process produced regression predicted and actual values for the average customer, which enabled the calculation of errors for the average customer and the calculation of the R-squared value. The same process was performed to estimate the amount of explained variation for the average customer in specific segments. The R-squared values for the average participant and for the average customer by segment were estimated using the following formula⁴⁸:

$$R^2 = 1 - \frac{\sum_t (\hat{y}_t - y_t)^2}{\sum_t (\hat{y}_t - \bar{y})^2}$$

Where:

y_t is the actual energy use at time t

$\hat{y}_t$ is the regression predicted energy use at time t

⁴⁸ Technically, the R-squared value needs to be adjusted based on the number of parameters and observations from each regression. Given that the number of observations per regression was typically over two thousand, the effects of the adjustment were anticipated to be minimal. As a result, the unadjusted R-squared is presented in order to avoid the complication of tracking the number of observations and parameters from each individual regression.

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

$\bar{y}$ is the actual mean energy use across all time periods.

Table 4-2 summarizes the amount of variation explained by the regression model for the average customer for specific segments. Overall, depending on the specific group assessed, roughly 80 to 93 percent of the variation is explained through the individual regressions.

Table 4-2
Adjusted R-squared Values for the Average Customer by Segment

Annual Consumption	R ²	Ratio of Summer to Non-Summer Usage	R ²	CARE Status	R ²
5,000 or less	0.90	<100%	0.89	Non-CARE	0.91
5,000 to 7,500	0.78	100%-125%	0.89	CARE	0.93
7,500 to 10,000	0.91	125%-150%	0.88		
10,000 to 12,500	0.93	150-175%	0.94		
12,500 to 15,000	0.85	175%-200%	0.93		
15,000 +	0.92	200%+	0.85		

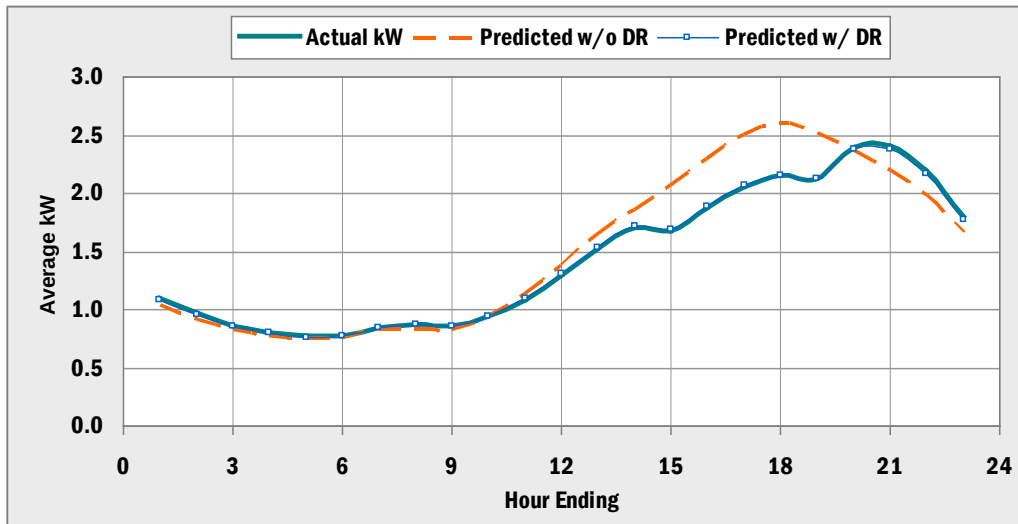
The most important feature of load impact analysis is the ability to predict accurately customer load and load reductions under the extreme conditions for which demand response is designed to provide a reliable resource. Unlike day matching methods, the accuracy of load impact estimates depend more on the accuracy of the regression coefficients representing the load impacts than on how well the regression predicts customer load. Put differently, if properly designed, regressions can accurately estimate load impacts and are more robust to over or under predictions of hourly energy consumption than day-matching methods. For SmartRate, we are not only confident that the load impact parameters are accurate, but the regression predicted values of energy consumption closely mirror and are often nearly indistinguishable from actual energy consumption, further validating the accuracy of the load impact estimates. To assess the accuracy and validity of the model, we compared actual and predicted values during event days by hour and temperature. In addition, given the estimated differences in response between CARE and non-CARE customers discussed in Section 4, we also present the comparisons of actual and regression predicted values for those customer segments.

Figure 4-3 shows the actual average hourly energy use of customers during event days compared to the regression predicted average customer energy use with and without demand response. The close match between predicted values with demand response and actual values reflects the ability of the regressions to predict accurately under event conditions. Figure 4-4 compares the actual and predicted values by temperature, based on data from the nine event days, and illustrates the model's ability to predict accurately customer behaviour under event conditions for a wide range of temperatures. It also illustrates that, in general, load response increases as temperature increases. Figure 4-5

APPENDIX 4: OVERVIEW OF REGRESSION METHODS

compares the actual and regression predicted values for CARE and non-CARE customers during event days. For each figure the relevant comparison of accuracy is between the actual load under event conditions (the solid line) and the regression predicted load under the same conditions (solid line with squares). For most graphs, the two are nearly indistinguishable. In addition, for information purposes, we have included the regression predicted values absent the SmartDay event (the dashed line). All of the comparisons are for time periods with actual interval data reads. Estimates of missing values were not included in the comparisons or in the estimated regressions. As seen, the regressions predict the behaviour of both CARE and non-CARE enrollees in Bakersfield extremely well.

Figure 4 - 3
Average Residential Customer Actual and Predicted Values for the Average SmartDay



APPENDIX 4: OVERVIEW OF REGRESSION METHODS

Figure 4 - 4
Average Residential Customer Actual and Predicted Values by Temperature for Event Days

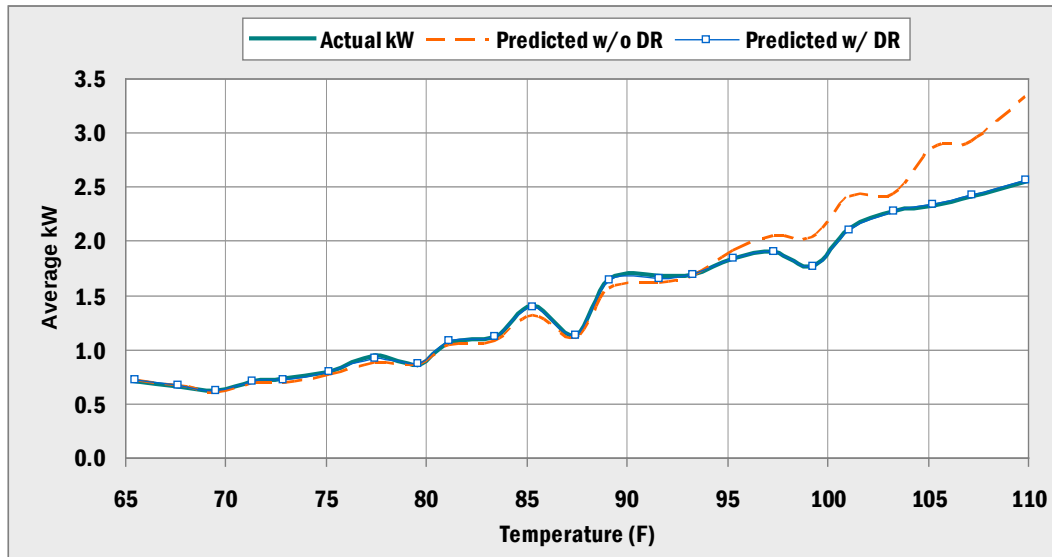
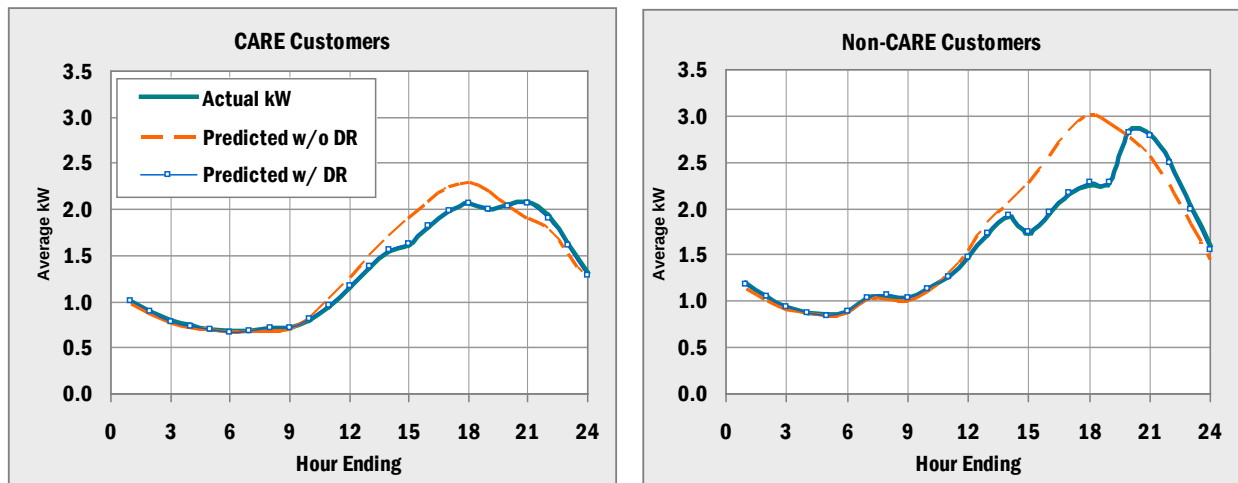


Figure 4 – 5
Comparison of CARE and Non-CARE Participants Actual and Predicted Values for Event Days



Similar comparisons of actual and predicted values were conducted by month, day of week, individual event days, and various other iterations – all of which indicated that the results were not only unbiased for the average day and average customer, but also across multiple customer segments and temporal characteristics.

Ministry of Energy

Office of the Minister

4th Floor, Hearst Block
900 Bay Street
Toronto ON M7A 2E1
Tel.: 416-327-6758
Fax: 416-327-6754

Ministère de l'Énergie

Bureau du ministre

4^e étage, édifice Hearst
900, rue Bay
Toronto ON M7A 2E1
Tél. : 416 327-6758
Téléc. : 416 327-6754



MAR 3 1 2014

Mr. Colin Andersen
Chief Executive Officer
Ontario Power Authority
1600-120 Adelaide Street West
Toronto ON M5H 1T1

Dear Mr. Andersen:

Re: 2015-2020 Conservation First Framework

I write in my capacity as the Minister of Energy in order to exercise the statutory power of ministerial direction I have in respect of the Ontario Power Authority (OPA) under the *Electricity Act, 1988*, as amended (the "Act").

Background

In *Achieving Balance: Ontario's Long-Term Energy Plan* (LTEP 2013), released on December 2, 2013 the Government established a provincial conservation and demand management (CDM) target of 30 terawatt hours (TWh) in 2032. To assist the Government in achieving this target, LTEP 2013 also committed to establishing a new six-year Conservation First Framework beginning in January 2015, replacing the one that is currently winding down. The new Conservation First Framework will enable the achievement of all cost-effective conservation and foster innovation through information sharing and the adoption of new technologies and approaches, including innovative performance management structures to drive greater energy savings.

To remain on track to achieve the LTEP 2013 CDM target, it is forecasted that 7 TWh needs to be achieved between 2015 and the end of 2020 through Distributor CDM programs enabled by the Conservation First Framework. In addition, transmission connected customers will continue to have access to OPA CDM programs.

To this end, I have issued a directive to the Ontario Energy Board (the "Board") (the "CDM Directive"), instructing it to amend the license of each licensed electricity distributor (Distributor) to add a condition that specifies the Distributor shall, between January 1, 2015 and December 31, 2020, make CDM programs available to customers in its licensed service area and shall, as far as is appropriate and reasonable having regard to the composition of the Distributor's customer base, do so in relation to each customer segment in its service area (CDM Requirement). Such Distributor CDM programs are required to achieve reductions in electricity consumption.

Each Distributor will be required to meet its CDM Requirement by:

- i. making a core set of province-wide CDM programs, funded by the OPA, available to customers in its licensed service area (Province-Wide Distributor CDM Programs);
- ii. making local and/or regional CDM programs, funded by the OPA, available to customers in its licensed service area (Local Distributor CDM Programs); or
- iii. a combination of (i) and (ii).

Direction

Therefore, pursuant to my authority under section 25.32 of the Act, I hereby direct the OPA to coordinate, support and fund the delivery of CDM programs through Distributors to achieve a total of 7 TWh of reductions in electricity consumption between January 1, 2015 and December 31, 2020 in accordance with the following guiding principles and requirements.

GUIDING PRINCIPLES

The OPA shall implement this direction according to the following principles:

1. Distributors are the face of electricity conservation to their customers in all sectors.
2. Distributors will be provided with long term, stable funding to provide the certainty they need to implement CDM programs.
3. Customers will be given more CDM program choice along with streamlined oversight and administration.
4. Distributors will have accountability for meeting their assigned CDM targets and will be provided the authority and means for meeting them cost-effectively.
5. Innovation and the adoption of new technologies will be encouraged.

6. While there will be CDM programs available for all residential, commercial and industrial sectors, the value of CDM investments may be higher in some sectors than others.
7. There will be renewed efforts to deepen consumer awareness of CDM and how it relates more broadly to the electricity system.
8. CDM programs for low-income residential customers will be improved.
9. The role of Distributors in the delivery of CDM programs to on-reserve First Nation customers will be enhanced.
10. Distributor CDM programs will result in the full achievement of 7 TWh of electricity savings.
11. Approvals and administrative requirements will be streamlined to provide Distributors flexibility to design, deliver and administer CDM programs to their customers.
12. OPA will provide support to Distributors in the design and delivery of CDM Programs.

REQUIREMENTS

1. GOVERNANCE

- 1.1 The OPA shall manage its relationship with Distributors through new streamlined contracts on a non-competitive basis. The OPA will work with Distributors to put such contracts in place by January 1, 2015.
- 1.2 The OPA shall provide support to Distributors to assist them in submitting their CDM Plans, as outlined in section 3, to the OPA no later than May 1, 2015 for approval. The OPA shall continue to make 2011-2014 OPA contracted Province-Wide CDM Programs available to customers through their Distributor until the Distributor's CDM Plan is approved by the OPA.
- 1.3 The OPA shall provide Distributors with flexibility to design, deliver and administer Province-Wide Distributor CDM Programs and Local Distributor CDM Programs.

- 1.4 The OPA shall establish a budget to achieve 7 TWh of electricity savings over the six-year period, based on current system planning projections. The budget and 7 TWh target will be reviewed as part of the mid-term review, as described in section 6, and revised as needed based on achievable cost-effective conservation and system planning projections at the time.
- 1.5 The OPA shall establish a budget allocation for each Distributor in consideration of the Distributor CDM Target and CDM Plan as outlined in sections 2.2 and 3.
- 1.6 The OPA shall, in consultation with Distributors, develop a cost recovery and performance incentive mechanism for Distributors for making Province-Wide Distributor CDM Programs and/or Local Distributor CDM Programs available to customers in their service areas. For each Province-Wide Distributor CDM Program and Local Distributor CDM Program within the Distributors' CDM Plan, Distributors shall be provided a choice of the following cost recovery mechanisms:
 - i. **Full Cost Recovery:** The Distributor shall be paid the full amount of prudently incurred costs for the administration and implementation of its Province-Wide Distributor CDM Program and/or Local Distributor CDM Program, subject to the Distributor achieving a specified minimum level of its Distributor CDM Target. The OPA shall report back by July 1, 2014 with recommendations on administrative or financial consequences of Distributor underperformance, should it occur. A tiered performance incentive mechanism shall be made available to Distributors with incentives beginning to accrue once a Distributor achieves 100% of the portion of its Distributor CDM Target allocated to the full cost recovery mechanism, in amounts determined by the OPA in consultation with Distributors.; or
 - ii. **Pay for Performance:** The Distributor shall be paid for the administration and implementation of its Province-Wide Distributor CDM Program and/or Local Distributor CDM Program, corresponding to the portion of the Distributor CDM Target allocated to the pay for performance mechanism, based on a pre-specified value for each verified kilowatt hour of electricity savings achieved, in amounts determined by the OPA in consultation with Distributors.
- 1.7 The OPA shall, subject to necessary regulatory amendments, recover payments made under the Province-Wide Distributor CDM Programs and Local Distributor CDM Programs from the Global Adjustment Mechanism up to the budget established under section 1.4.

- 1.8 The OPA shall ensure that its contracts with Distributors include clauses allowing for corrections and changes in each Distributor CDM Target, as outlined in section 2.2, and in Distributor budgets which may be required in accordance with a mid-term review as outlined in section 6.

2. DISTRIBUTOR CDM TARGETS

- 2.1 The OPA, in consultation with Distributors, shall develop an allocation methodology to allocate the full 7 TWh among Distributors. The allocation methodology may take into consideration Distributor CDM potential at a local and/or regional level as identified in the OPA's 2014 energy efficiency achievable potential study, and other factors, as appropriate.
- 2.2 The OPA shall allocate to each Distributor a numeric CDM target ("Distributor CDM Target") to achieve reductions in electricity consumption for all customer segments in the Distributor's licensed service area.
- 2.3 The OPA shall encourage Distributors to aggregate Distributor CDM Targets with neighbouring Distributors to develop 21 regional CDM targets for the period January 1, 2015 to December 31, 2020. The OPA shall encourage Distributors to work cooperatively to develop regional CDM Plans to meet the regional CDM targets.
- 2.4 The OPA shall evaluate Distributor achievement of electricity savings on an annual incremental basis based on the OPA's Evaluation, Measurement and Verification (EM&V) protocols.

3. CDM PLANS AND PROGRAMS

- 3.1 The OPA shall support Distributors in designing a core set of Province-Wide Distributor CDM Programs for the following segments of distribution system connected customers to make available for delivery in Distributors' licensed service areas:
 - i. Residential
 - ii. Low-income
 - iii. Small business
 - iv. Commercial (including multi-family buildings)
 - v. Agricultural
 - vi. Institutional
 - vii. Industrial

3.2 Province-Wide Distributor CDM Programs shall:

- i. Be designed by Distributors, with support from the OPA, through working groups. The membership of the working groups shall consist of OPA and Distributor representatives.
- ii. Balance the value of flexibility for some program customization to meet local and/or regional needs with the value of offering consistent CDM measures to customer segments across all Distributor service areas.

3.3 The OPA shall support Distributors, as required, in designing Local Distributor CDM Programs, including programs for specific industry concentrations or customer segments in a particular licensed service area and/or region that require unique approaches to achieve electricity savings, such as on-reserve First Nation customers.

3.4 The OPA shall require each Distributor to submit a CDM Plan to the OPA for approval.

3.5 The OPA shall establish a streamlined review and approval process for Distributor CDM Plans and proposals for Province-Wide Distributor CDM Programs and Local Distributor CDM Programs. To facilitate this process, the OPA, in consultation with Distributors, shall establish guidelines that include rules relating to the streamlined review and approval of CDM Plans and proposals for Province-Wide Distributor CDM Programs and Local Distributor CDM Programs. In establishing such guidelines, the OPA shall have regard to the following objectives in addition to such other factors as the OPA considers appropriate:

- i. Distributor CDM Plans must provide a description of how the Distributor will achieve its Distributor CDM Target, including but not limited to, a description of the Distributor's year-by-year plan, including milestones for achieving its Distributor CDM Target, a description of Province-Wide Distributor CDM Programs and any Local Distributor CDM Programs, and projected budgets and electricity savings by sector.
- ii. The OPA shall establish a service standard of no more than 60 days for review and approval of Distributor CDM Plans and program. Any request by the OPA for additional information during its review will cause the remaining period for approval to be paused and shall resume at such time as the request is satisfied.

- iii. The OPA shall seek to approve unique Local Distributor CDM Programs that avoid marketplace confusion and ensure the prudent use of funds by avoiding duplication of Province-Wide Distributor CDM Programs. The OPA, in consultation with Distributors, shall establish rules on what constitutes duplication.
- iv. The OPA shall encourage Distributors to incent CDM measures with relatively longer lifespans and energy savings persistence and shall consider the system value of the measures, including reductions at peak times.
- v. The OPA shall ensure there is a positive benefit-cost analysis of each CDM Plan and each Province-Wide CDM Program and Local Distributor CDM Program utilizing the OPA's Total Resource Cost Test and the Program Administrator Cost Test found in the OPA's Cost-Effectiveness Guide, dated October 15, 2010 (OPA Cost-Effectiveness Tests), which may be updated by the OPA from time to time. The OPA will establish hurdle rates to consider the cost of delivering Province-Wide Distributor CDM Programs and Local Distributor CDM Programs against the avoided cost of procuring supply.
- vi. The OPA shall, despite section 3.5 (v), allow Distributors to apply to the OPA for approval of Province-Wide Distributor CDM Programs and Local Distributor CDM Programs where cost effectiveness is not demonstrated if the program is:
 - a) targeted to on-reserve First Nation customers
 - b) designed for educational purposes
 - c) a low-income CDM program
- vii. A Distributor may, despite section 3.5(v), submit a CDM Plan where cost effectiveness is not demonstrated if the Distributor can reasonably demonstrate that it is unable to develop a plan that is cost effective due its size, location, the nature of its customer base or other unusual circumstances. In order to obtain the approval of such a CDM Plan, the Distributor must also demonstrate that:
 - (a) it has made reasonable efforts to determine if a CDM Plan could be delivered cost effectively in its service area by another Distributor; and
 - (b) The CDM Plan will be delivered in as cost effective a manner as is reasonably possible.

- viii. The OPA shall take into consideration the cost and the number of First Nation, educational and low-income CDM programs that a Distributor already has undertaken or plans to undertake when approving these CDM programs. Although there is no requirement that First Nation, educational, or low-income programs be cost effective, Distributors shall be required to provide adequate evidence that the CDM programs will likely result in electricity savings and will be delivered in as cost effective a manner as is reasonably possible.
- ix. The OPA shall allow Distributors to propose changes and modifications to its CDM Plan on an annual basis, or more frequently.
- x. The OPA shall encourage Distributors to maximize administrative and delivery efficiencies by utilizing appropriate program delivery models. Specifically, the OPA and/or Distributors shall provide enhanced co-ordination efforts with regard to:
 - a) Opportunities to target consumers with multiple locations across several licensed service areas (e.g., national accounts) and CDM measures delivered or promoted through provincial or national channels (e.g., retailer in-store rebates or coupons); and
 - b) CDM activities, including, but not limited to, the marketing, procurement and delivery of CDM measures and/or services where these will afford significant administrative cost and/or delivery efficiencies (e.g., call centre, rebate fulfillment and appliance de-commissioning).
- xi. The OPA shall require Distributors, where appropriate, to coordinate and integrate Province-Wide Distributor CDM Programs and Local Distributor CDM Programs with natural gas distributor ("Gas Distributors") conservation programs to achieve efficiencies and convenient integrated programs for electricity and natural gas customers.
- xii. The OPA shall require Distributors, where appropriate, to coordinate and integrate low-income Province-Wide Distributor CDM Programs and Local Distributor CDM Programs with Gas Distributor low-income conservation programs.

4. MARKETING

- 4.1 The OPA shall be responsible for province-wide marketing and mass media buying for Province-Wide Distributor CDM Programs under the saveONenergy brand.
- 4.2 The OPA shall work with Distributors to ensure Province-Wide Distributor CDM Programs and Local Distributor CDM Programs are consistently marketed under the saveONenergy brand, and for local marketing and advertising efforts, co-branded with Distributor logos. The OPA may also work with Distributors to provide them with the advantages of scale (for example, in the purchase of media and the development, production and distribution of marketing material).
- 4.3 The OPA shall make the saveONenergy brand available to the Gas Distributors for marketing of natural gas conservation programs on terms that the OPA may negotiate with the Gas Distributors.

5. REPORTING

- 5.1 The OPA shall continue to produce and publish an annual report on overall progress toward achieving the provincial CDM target of 30 TWh, including contributions to the target achieved through Province-Wide Distributor CDM Programs, Local Distributor CDM Programs, demand response programs, programs for transmission connected customers and product codes and standards. The annual report shall cover the period from January 1 to December 31 of the previous year.

6. MID-TERM REVIEW

- 6.1 The OPA, in consultation with the Ministry of Energy and Distributors, shall no later than June 1, 2018 have completed a formal mid-term review of:
 - i. the 7 TWh target and the overall budget for achieving that target
 - ii. allocation of budgets and Distributor CDM Targets
 - iii. lessons learned on cost recovery and performance incentive mechanisms, and;
 - iv. CDM contribution to regional planning

- 6.2 The OPA shall conduct an achievable potential study for electricity efficiency in Ontario every three-years, with the first study completed by June 1 2016, to inform electricity efficiency planning and programs. The achievable potential study should, as far as is appropriate and reasonable having regard to the respective characteristics of the electricity and natural gas sectors, be coordinated with the natural gas efficiency achievable potential study referred to in the CDM Directive to the Board.

7. DEFINITION OF CDM

- 7.1 The OPA shall consider CDM to be inclusive of activities aimed at reducing electricity consumption and reducing the draw from the electricity grid, such as geothermal heating and cooling, solar heating and small scale (i.e., <10MW) behind the meter customer generation. However, CDM should be considered to exclude those activities and programs related to a Distributor's investment in new infrastructure or replacement of existing infrastructure, any measures a Distributor uses to maximize the efficiency of its new or existing infrastructure, activities promoted through a different program or initiative undertaken by the Government of Ontario or the OPA, such as the OPA Feed-in Tariff (FIT) Program and micro-FIT Program and activities related to the price of electricity or general economic activity.

8. SUPPORT AND FUNDING FOR RESEARCH AND INNOVATION

- 8.1 The OPA Conservation Fund provides financial support to new and innovative electricity conservation initiatives designed to enable Ontario's residents, businesses and institutions to cost-effectively reduce their demand for electricity
- 8.2 The OPA shall continue to provide, through its Conservation Fund, support and funding for new and innovative electricity conservation initiatives, including small scale distribution storage technologies, as a means to assist Distributors and others in their conservation efforts.

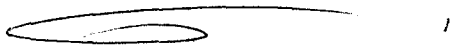
9. PEAKSAVERPLUS PROGRAM

- 9.1 LTEP 2013 committed that Ontario will aim to use demand response to meet 10% of peak demand by 2025, equivalent to approximately 2,400 megawatts under current forecast conditions. To encourage further development of demand response in Ontario, the Independent Electricity System Operator ("IESO") will evolve existing demand response programs in Ontario and introduce new initiatives.

- 9.2 A transition plan is currently being developed to evolve existing programs, potentially including the peaksaverPLUS program, to an IESO administered market. Until such time as the transition plan has been finalized, including plans for the peaksaverPLUS program, the OPA shall continue to make the program available to Distributors to deliver to customers in their licensed service areas.

This direction takes effect on the date it is issued.

Sincerely,



Bob Chiarelli
Minister

- cc. James D. Hinds, Chair, Ontario Power Authority
Rosemarie T. Leclair, Chair and Chief Executive Officer, Ontario Energy Board
Bruce Campbell, President and Chief Executive Officer, Independent Electricity System Operator
Tim O'Neill, Chair, Independent Electricity System Operator
Serge Imbrogno, Deputy Minister, Ministry of Energy
Halyna Perun, Director, Legal Services Branch, Ministries of Energy and Infrastructure

Page Intentionally Blank