ATTACHMENT A

FILED
03-25-08
01:13 PM

Load Impact Estimation for Demand Response: Protocols and Regulatory Guidance

California Public Utilities Commission
Energy Division

March 2008

ACKNOWLEDGEMENTS AND CREDITS

This document benefited from a large number of contributors:

- The Joint IOUs (PG&E, SCE and SDG&E) were responsible for developing the overall framework of this effort by submitting the initial straw proposals for the load impact protocols.
- Steve George, Ph.D., Michael Sullivan, Ph.D., and Josh Bode, MPP, of Freeman, Sullivan & Co. led much of the straw proposal develop as a contractor to the Joint Utilities.
- The Joint Staff of the California Public Utilities Commission and the California Energy Commission with Dorris Lam (CPUC) and David Hungerford (CEC) serving as lead contacts.
- Daniel Violette, Ph.D., and Mary Klos, M.S., of Summit Blue Consulting who served as expert staff to the CPUC and CEC, helped in the development of Joint Staff Guidance Documents, moderated the workshops on the protocols, and reviewed straw proposals submitted by the parties to the proceeding.
- Representatives from the Joint Parties (EnerNOC, Inc., Energy Connect, Comverge, Inc., Ancillary Services Coalition, and California Large Energy Consumers Association).
- Representatives from regulatory agencies such as the Department of Ratepayer Advocates (DRA).
- Parties to the proceeding such as The Utility Reform Network (TURN).

Among specific content credits are:

- The day-matching and regression examples were developed by Freeman, Sullivan & Company, as well as development of much of the discussion on sampling, reporting requirements for the evaluations, and discussions of the evaluation methods.
- Joint Staff recommendations that led to the development of the evaluation planning protocols, the process protocols, and a portfolio protocol to identify positive or negative synergies with other DR and energy efficiency programs.
- Many of these Joint Staff recommendations were developed into protocols by Freeman, Sullivan & Company and by Summit Blue Consulting as advisors to the Joint IOUs and Joint Staff, respectively.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS AND CREDITS 2
1. EXECUTIVE SUMMARY 6
2. BACKGROUND AND OVERVIEW 11
2.1. Background on Load Impact Protocols 11
2.2. Taxonomy of Demand Response Resources 13
2.3. Purpose of this Document 17
2.4. Report Organization 18
3. EVALUATION PLANNING 19
3.1. Planning Protocols (Protocols 1-3) 19
3.2. Additional Requirements to be Assessed in the Evaluation Plan 22
3.2.1. Statistical Precision 23
3.2.2. Ex Post Versus Ex Ante Estimation 24
3.2.3. Impact Persistence 24
3.2.4. Geographic Specificity 25
3.2.5. Sub-Hourly Impact Estimates 25
3.2.6. Customer Segmentation 25
3.2.7. Additional Day Types 26
3.2.8. Understanding Why, Not Just What 27
3.2.9. Free Riders and Structural Benefiters 27
3.2.10. Control Groups 28
3.2.11. Collaboration When Multiple Utilities Have the Same DR Resource Options 29
3.3. Input Data Requirements 29
4. EX POST EVALUATIONS FOR EVENT BASED RESOURCES 33
4.1. Protocols for Ex Post Impact Evaluations – Day-matching, Regression Methods and Other Methods 36
4.1.1. Time Period Protocols (Protocols 4 and 5) 36
4.1.2. Protocols for Addressing Uncertainty (Protocol 6) 37
4.1.3. Output Format Protocols (Protocol 7) 38
4.1.4. Protocols for Impacts by Day Types (Protocol 8) 41
4.1.5. Protocols for Production of Statistical Measures (Protocols 9 and 10) 43
4.2. Guidance and Recommendations for Ex Post Evaluation of Event Based Resources – Day-matching, Regression, and Other Methods 48

4.2.1. Day-matching Methodologies 49
4.2.2. Regression Methodologies 60
4.2.3. Other Methodologies 76
4.2.4. Measurement and Verification Activities 78
5. EX POST EVALUATION FOR NON-EVENT BASED RESOURCES 80
5.1. Protocols for Non-Event Based Resources (Protocols 11-16) 81
5.2. Guidance and Recommendations 84
5.2.1. Regression Analysis 84
5.2.2. Demand Modeling 88
5.2.3. Engineering Analysis 89
5.2.4. Day-matching for Scheduled DR 91
6. EX ANTE ESTIMATION 92
6.1. Protocols for Ex Ante Estimation (Protocols 17-23) 93
6.2. Guidance and Recommendations 98
6.2.1. Ex Ante Scenarios 98
6.2.2. Impact Estimation Methods 101
6.3. Impact Persistence 104
6.4. Uncertainty in Key Drivers of Demand Response 106
6.4.1. Steps for Defining the Uncertainty of Ex Ante Estimates 106
6.4.2. Defining the Uncertainty of Ex Ante Estimates: Example 107
7. ESTIMATING IMPACTS FOR DEMAND RESPONSE PORTFOLIOS (PROTOCOL 24) 111
7.1. Issues in Portfolio Aggregation 114
7.1.1. Errors Resulting from Improper Aggregation of Individual Resource Load Impacts 115
7.1.2. Errors Resulting from Incorrect Assumptions About Underlying Probability Distributions 116
7.1.3. Errors Resulting from a Failure to Capture Correlations across Resources 117
7.2. Steps in Estimating Impacts of DR Portfolios 119
7.2.1. Define Event Day Scenarios 119
7.2.2. Determine Resource Availability 120
7.2.3. Estimate Uncertainty Adjusted Average Impacts per Participant for Each Resource Option 121
7.2.4. Aggregate Impacts across Participants 121
7.2.5. Aggregate Impacts across Resources Options 121
8. SAMPLING 124

8.1. Sampling Bias (Protocol 25) 125
8.2. Sampling Precision 128
8.2.1. Establishing Sampling Precision Levels 130
8.2.2. Overview of Sampling Methodology 131
8.3. Conclusion 139
9. REPORTING PROTOCOLS (PROTOCOL 26) 141
10. PROCESS PROTOCOL (PROTOCOL 27) 147
10.1. Evaluation Planning—Review and Comment Process 147
10.2. Review of Interim and Draft Load Impact Reports 148
10.3. Review of Final Load Impact Reports 148
10.4. Resolution of Disputes 148

1. EXECUTIVE SUMMARY

California's Energy Action Plan (EAP II) emphasizes the need for demand response resources (DR) that result in cost-effective savings and the creation of standardized measurement and evaluation mechanisms to ensure verifiable savings. California Public Utilities Commission (CPUC) Decision D.05-11-009 identified a need to develop measurement and evaluation protocols and cost-effectiveness tests for demand response (DR). On January 25, 2007, the Commission opened a rulemaking proceeding (OIR 07-01-041), with several objectives, including:[1]

- Establishing a comprehensive set of protocols for estimating the load impacts of DR resources;
- Establishing methodologies to determine the cost-effectiveness of DR resources.

In conjunction with this rulemaking, a scoping memo[2] was issued directing the three major investor owned utilities (IOUs) in California, and allowing other parties, to develop and submit a "straw proposal" for load impact protocols for consideration. In order to guide development of the straw proposals, the Energy Division of the CPUC and the Demand Analysis Office of the California Energy Commission (Joint Staff) issued a document on May 24, 2007 entitled *Staff Guidance for Straw Proposals On: Load Impact Estimation from DR and Cost-Effectiveness Methods for DR*. The Staff Guidance document indicated that straw proposals should focus on estimating DR impacts for long-term resource planning.[3]

On July 16, 2007, three straw proposals on Load Impact Estimation were filed by the *Joint IOU*[4], the Joint Parties[5], and Ice Energy, Inc. A workshop to address questions about the straw proposals was held at the Commission on July 19, 2007 and written comments on the straw proposals were submitted to the Commission on July 27th. On August 1, 2007, a workshop was held to discuss areas of agreement and disagreement regarding the straw proposals. Parties worked together to prepare a report, filed by the Joint IOUs on August 22, 2007, describing the area of agreement and disagreement among the parties and a plan incorporating the agreements into a new straw proposal.[6] On September 10, 2007, the Joint IOUs and the Joint Parties each

[1] R07-01-041, p.1.
[2] Assigned Commissioner and Administrative Law Judge's Scoping Memo and Ruling, April 18, 2007
[3] CPUC/CEC. Staff Guidance for Straw Proposals On: Load Impact Estimation from DR and Cost-Effectiveness Methods for DR. May 24, 2007. p.10.
[4] Stephen George, Michael Sullivan and Josh Bode. *Joint IOU Straw Proposal on Load Impact Estimation for Demand Response.* Prepared on behalf of Pacific Gas & Electric Co., Southern California Edison Co., and San Diego Gas & Electric Co. July 16, 2007.
[5] EnerNOC, Inc., Energy Connect, Comverge, Inc., Ancillary Services Coalition, and California Large Energy Consumers Association.
[6] The Joint IOUs filed a motion on August 7th to obtain permission to file a revised proposal incorporating agreements reached at the August 1st workshop and to modify the original schedule to allow for this submission to made and for comments to be provided prior to the Commission's ruling. The presiding administrative law judge granted the Joint IOU request in a ruling on August 13, 2007.

filed their revised straw proposals for DR load impact estimation protocols. The Joint Staff submitted a Recommendation Report on LI estimation on October 12, 2007 in response to the revised straw proposal and the area of agreement/disagreement. Comments[7] on the Joint Staff Recommendation Report on LI Estimation were received on October 24, 2007.

Estimating DR impacts for long-term resource planning is inherently an exercise in ex ante estimation. However, ex ante estimation should, where possible, utilize information from ex post evaluations of existing DR resources. As such, meeting the Commission's requirement to focus on estimating DR impacts for long-term resource planning requires careful attention to ex post evaluation of existing resources. Consequently, the protocols and guidance presented here address both ex post evaluation and ex ante estimation of DR impacts.

> ***The purpose of this document is to establish minimum requirements for load impact estimation for DR resources and to provide guidance concerning issues that must be addressed and methods that can be used to develop load impact estimates for use in long term resource planning. The minimum requirements indicate that uncertainty adjusted, hourly load impact estimates be provided for selected day types and that certain statistics be reported that will allow reviewers to assess the validity of the analysis that underlies the estimates.***

While DR resources differ significantly across many factors, one important characteristic, both in terms of the value of DR as a resource and the methods that can be used to estimate impacts, is whether the resource is tied to a specific event, such as a system emergency or some other trigger. Event based resources can include critical peak pricing, direct load control, and auto DR. Non-event based resources include traditional time-of-use rates, real time pricing and permanent load shifting (e.g., through technology such as ice storage).

These load impact estimation protocols outline what must be done when estimating the impacts of DR activities. They could focus on the output of a study, defining what must be delivered, on how to do the analysis, or both. The protocols presented here focus on what impacts should be estimated, what issues should be considered when selecting an approach, and what to report, not on how to do the job.

The best approach to estimating impacts is a function of many factors— resource type, target market, resource size, available budget, the length of time a resource has been in effect, available data, and the purposes for which the estimates will be used. Dictating the specific methods that must be used for each impact evaluation or ex ante forecast would require an unrealistic level of foresight, not to mention dozens, if not hundreds, of specific requirements. More importantly, it would stifle the flexibility and creativity that is so important to improving the state of the art.

On the other hand, there is much that can be learned from previous work and, depending on the circumstances, there are significant advantages associated with certain approaches to impact estimation compared with others. Furthermore, it is imperative that an evaluator have a good understanding of key issues that must be addressed when conducting the analysis, which vary by

[7] The following parties filed comments on the Staff Report: Comverge, EnerNOC, and Energy Connect (jointly), the IOUs(jointly), CAISO, DRA, TURN, KM, and Wal-Mart.

resource type, user needs, and other factors. As such, in addition to the protocols, this document also provides guidance and recommendations regarding the issues that are relevant in specific situations and effective approaches to addressing them.

While the protocols contained in this report establish minimum requirements for the purpose of long term resource planning, they also recognize that there are other applications for which load impact estimates may be needed and additional requirements that may need addressing. Consequently, the protocols established here require that a plan be provided describing any additional requirements that will also be addressed as part of the evaluation process.

Separate protocols are provided for ex post evaluation of event based resource options, ex post evaluation of non-event based resources and ex ante estimation for all resource options, although the differences across the three categories are relatively minor. In general, the protocols require that:

- An evaluation plan be produced that establishes a budget and schedule for the process, develops a preliminary approach to meeting the minimum requirements established here, and determines what additional requirements will be met in order to address the incremental needs that may arise for long term resource planning or in using load impacts for other applications, such as customer settlement or CAISO operations;

- Impact estimates be provided for each of the 24 hours on various event day types for event based resource options and other day types for non-event based resources;

- Estimates of the change in overall energy use in a season and/or year be provided;

- Uncertainty adjusted impacts be reported for the 10th, 30th 50th, 70th, and 90th percentiles, reflecting the uncertainty associated with the precision of the model parameters and potentially reflecting uncertainty in key drivers of demand response, such as weather;

- Outputs that utilize a common format, as depicted in Table 1-1 for ex post evaluation. A slightly different reporting format is required for ex ante estimation;

- Estimates be provided for each day type indicated in Table 1-2;

- Various statistical measures be provided so that reviewers can assess the accuracy, precision and other relevant characteristics of the impact estimates;

- Ex ante estimates that utilize all relevant information from ex post evaluations whenever possible, even if it means relying on studies from other utilities or jurisdictions;

- Detailed reports be provided that document the evaluation objectives, impact estimates, methodology, and recommendations for future evaluations.

Table 1-1. Reporting Template for Ex Post Impact Estimates

Hour Ending	Estimated Reference Load (kWh/hr)	Observed Load (kWh/hr)	Estimated Load Impact (kWh/hr)	Temp (F)	10th	30th	50th	70th	90th
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
Daily									

Table 1-2. Day Types for which Impact Estimates are to be Provided

	Event Based Resources			Non-Event Based Resources		
Day Types	Event Driven Pricing	Direct Load Control	Callable DR	Non-event Driven Pricing	Scheduled DR	Permanent Load Reductions
Ex Post Day Types						
Each Event Day	X	X	X			
Average Event Day	X	X	X			
Average Weekday Each Month				X	X	X
Monthly System Peak Day				X	X	X
Ex Ante Day Types						
Typical Event Day	X	X	X			
Average Weekday Each Month (1-in-2 and 1-in-10 Weather Year)	X	X	X	X	X	X
Monthly System Peak Day (1-in-2 and 1-in-10 Weather Year)	X	X	X	X	X	X

Finally, these protocols are focused on reporting requirements for resource planning in the future and may not be appropriate or feasible for other applications of demand response load impacts. As a result, the focus of this effort is on estimates of program-wide impacts and projections of these impacts that span a planning horizon. This planning objective is different than much of the research conducted into DR impacts which have had as their objective the estimation of event-based impacts that can be used as a basis for payments to participating customers (termed "settlements" in most of the literature). These settlements often need to be estimated quickly to allow for timely payments to participants, and they may need a level of transparency that can be understood by all the parties. Impact estimates for resource planning can use more complex methods and data spanning longer time frames than would be appropriate if the goal is prompt payments to customers after an event has occurred.

2. BACKGROUND AND OVERVIEW

Demand response resources are an essential element of California's resource strategy, as articulated in the State's Energy Action Plan II (EAP II). EAP II has determined how energy resources should be deployed to meet California's energy needs and ranks DR resources second in the "loading order" after energy efficiency resources. The EAP II emphasizes the need for DR resources that result in cost-effective savings and the creation of standardized measurement and evaluation mechanisms to ensure verifiable savings.[8]

2.1. Background on Load Impact Protocols

California Public Utilities Commission (CPUC) Decision D.05-11-009 identified a need to develop measurement and evaluation protocols and cost-effectiveness tests for demand response. That decision ordered CPUC staff to undertake further research and recommend to the Executive Director whether to open a proceeding to address these issues. Commission staff recommended opening a rulemaking, which the Commission did on January 25, 2007. The objectives of OIR 07-01-041 are to:[9]

- Establish a comprehensive set of protocols for estimating the load impacts of DR resources;
- Establish methodologies to determine the cost-effectiveness of DR resources;
- Set DR goals for 2008 and beyond, and develop rules on goal attainment; and
- Consider modifications to DR resources needed to support the California Independent System Operator's (CAISO) efforts to incorporate DR into market design protocols.

As indicated in the ruling, it is expected that the load impact protocols will not only provide input to determining DR resource cost-effectiveness, but will also assist in resource planning and long-term forecasting.[10]

On April 18, 2007, the Assigned Commissioner and Administrative Law Judge's Scoping Memo and Ruling indicated that the three major investor-owned utilities in California must jointly develop and submit a "straw proposal" for load impact protocols.

On May 3, 2007, the Commission held a workshop on load impact estimation protocols. At the workshop, the joint utilities indicated that there were many potential applications of impact estimates for demand response resources, including:

1. Ex post impact evaluation

[8] R07-01-041, p.2.
[9] R07-01-041, p.1.
[10] Ibid. p.2

2. Monthly reporting of DR results
3. Forecasting of DR impacts for resource adequacy
4. Forecasting of DR impacts for long-term resource planning
5. Forecasting DR impacts for operational dispatch by the CAISO
6. Estimation for customer settlement/reference level methods (e.g., payment of incentives) in conjunction with DR resource deployment.

The joint utilities also indicated that the relevant issues vary substantially across the six applications listed above. Attempting to address all of these issues and methods would be extremely difficult in the short time frame allowed for development of the protocols. The joint utilities asked for guidance and clarification regarding priorities and scope.

On May 24, 2007, the Energy Division of the CPUC and Demand Analysis Office of the CEC issued a document entitled Staff Guidance for Straw Proposals On: Load Impact Estimation from DR and Cost-Effectiveness Methods for DR (hereafter referred to as the Staff Guidance document). The Staff Guidance document indicated the focus of the straw proposals should be on estimating DR impacts for long-term resource planning.[11]

Straw Proposal on Load Impact Estimation for Demand Response was provided to the Commission on July 16, 2007 by the Joint IOUs[12], the Joint Parties[13], and Ice Energy, Inc. A workshop to address questions about the joint IOU straw proposal and straw proposal submissions by other stakeholders was held at the Commission on July 19, 2007 and written comments on the Straw Proposal were submitted to the Commission on July 27th. On August 1, 2007, a workshop was held to discuss areas of agreement and disagreement regarding the Joint IOU straw proposal and proposals submitted by other stakeholders. Parties worked together to prepare a report, filed by the Joint IOUs on August 22, 2007 delineating the areas of agreement and disagreement among the parties, identifying errata and referencing incorporation of the agreements into a revised straw proposal[14]. On September 10, 2007, the Joint IOUs and the Joint Parties each filed their revised straw proposals for DR load impact estimation protocols. The Joint Staff submitted a Recommendation Report on LI estimation on October 12, 2007 in response to the revised straw proposal and the area of agreement/disagreement. Comments[15] on the Joint Staff Recommendation Report on LI Estimation were received on October 24, 2007.

---

[11] CPUC/CEC. Staff Guidance for Straw Proposals On: Load Impact Estimation from DR and Cost-Effectiveness Methods for DR. May 24, 2007, with assistance by Summit Blue Consulting, p.10.
[12] Pacific Gas & Electric Co., Southern California Edison Co., and San Diego Gas & Electric Co.
[13] EnerNOC, Inc., Energy Connect, Comverge, Inc., Ancillary Services Coalition, and California Large Energy Consumers Association.
[14] The Joint IOUs filed a motion on August 7th to obtain permission to file a revised proposal incorporating agreements reached at the August 1st workshop and to modify the original schedule to allow for this submission to made and for comments to be provided prior to the Commission's ruling. The presiding ALJ granted the Joint IOU motion on August 13, 2007.
[15] The following parties filed comments on the Staff Report: Comverge, EnerNOC, and Energy Connect (jointly), the IOUs(jointly), CAISO, DRA, TURN, KM, and Wal-Mart.

Estimating DR impacts for long-term resource planning is inherently an exercise in ex ante estimation. As indicated in subsequent sections, ex ante estimation should, wherever possible, utilize information from ex post evaluations of existing DR resources. Empirical evidence, properly developed, is almost always superior to theory, speculation, market research surveys, engineering modeling or other ways of estimating what impacts might be for a specific DR resource option. As such, meeting the Commission's requirement to focus on estimating DR impacts for long-term resource planning requires careful attention to ex post evaluation of existing resources. Consequently, the protocols and guidance contained in the remainder of this report address both ex post evaluation and ex ante estimation of DR impacts.

2.2. Taxonomy of Demand Response Resources

There is a wide variety of DR resources that are currently in place in California (and elsewhere) and many different ways to categorize them. While DR resources differ significantly across many factors, one important characteristic, both in terms of the value of DR as a resource and the methods that can be used to estimate impacts, is whether the resource is tied to a specific event, such as a system emergency or some other trigger. Event based resources include critical peak pricing, direct load control and autoDR. Non-event based resources include traditional time-of-use rates, real time pricing and permanent load shifting (e.g., through technology such as ice storage).

In addition to whether a resource is event based, there are other characteristics of interest, such as whether a resource uses incentives or prices to drive demand response and whether impacts are primarily technology driven, purely behaviorally driven or some combination of the two. Two groups of DR activities are distinguished by whether or not the resources are event based.

Event based resources include:

- **Event-based Pricing**—This resource category includes prices that customers can respond to based on an event, i.e., a day-ahead or same-day call. This includes many pricing variants such as critical peak pricing or a schedule of prices presented in advance that would allow customers to indicate how much load they will reduce in each hour at the offered price (e.g., demand bidding). The common element is that these prices are tied to called events by the utility, DR administrator, or other operator.

- **Direct Load Control**—This resource category includes options such as air conditioning cycling targeted at mass-market customers as well as options such as auto-DR targeted at large customers. The common thread is that load is controlled at the customer's site for a called event period through a signal sent by an operator.

- **Callable DR**—This resource category is similar to direct load control but, in this case, a notification is sent to the customer who then initiates actions to reduce loads, often by an amount agreed to in a contract. The difference is that load reduction is based on actions taken by the customer rather than based on an operator-controlled signal that shuts off equipment. Interruptible and curtailable tariffs are included in this category.

Non-event based resources include:

- **Non-event based pricing**—This resource category includes TOU, RTP, and related pricing variants that are not based on a called event—that is, they are in place for a season or a year.

- **Scheduled DR**—There are some loads that can be scheduled to be reduced at a regular time period. For example, a group of irrigation customers could be divided into five segments, with each segment agreeing to not irrigate/pump on a different selected weekday.
- **Permanent load reductions and load shifting**—Permanent load reductions are often associated with energy efficiency activities, but there are some technologies such as demand controllers that can result in permanent load reductions or load shifting. Examples of load shifting technologies include ice storage air conditioning, timers and energy management systems.

Tables 2-1 through 2-3 show how the existing portfolio of DR resources for each IOU map into the taxonomy summarized above.

Table 2-1. PG&E Demand Response Resources

Resource	Target Market Segment	Event Based Resources			Non-Event Based Resources		
		Event-based Pricing	Direct Load Control	Callable DR	Non-event Based Pricing	Scheduled DR	Permanent Load Reductions
E-CPP (Voluntary Critical Peak Pricing)	C&I	X					
E-CBP (Capacity Bidding Program)	C&I	X					
E-BEC (Business Energy Coalition)	C&I	X					
E-DBP (Demand Bidding Program)	C&I	X					
TOU (Time-of-Use Pricing)	Residential / C&I				X		
E-RSAC (Residential Smart A/C Program)	Residential		X				
E-CSAC (Commercial Smart A/C Program)	C&I		X				
Aggregator Managed Portfolio	C&I	X					
E-NF (Non-firm Rate Schedule)	C&I			X			
E-BIP (Base Interruptible Program)	C&I			X			
E-SLRP (Scheduled Load Reduction Program)	C&I					X	

Table 2-2. SCE Demand Response Resources

Resource	Target Market Segment	Event Based Resources			Non-Event Based Resources		
		Event-based Pricing	Direct Load Control	Callable DR	Non-event Based Pricing	Scheduled DR	Permanent Load Reductions
CPP (Critical Peak Pricing)	C&I	X					
DBP (Demand Bidding Program)	C&I	X					
CBP (Capacity Bidding Program)	C&I	X					
TOU (Time-of-Use Pricing)	Residential / C&I				X		
RTP (Real-time Pricing)	C&I				X		
SDP (Summer Discount Plan)	Residential / C&I		X				
AP-I (Agricultural and Pumping Interruptible Program)	C&I		X				
Automated DR	C&I		X				
OBMC (Optional Binding Mandatory Curtailment)	C&I			X			
BIP (Base Interruptible Program)	C&I			X			
I-6 Large Power Interruptible Program	C&I			X			
SLRP (Scheduled Load Reduction Program)	C&I					X	
EnerNOC Contract				X			

Table 2-3. SDG&E Demand Response Resources

Resource	Target Market Segment	Event Based Resource			Non-Event Based Resource		
		Event-based Pricing	Direct Load Control	Callable DR	Non-event Based Pricing	Scheduled DR	Permanent Load Reductions
CPP (Critical Peak Pricing)	C&I	X					
CPP-E (Critical Peak Pricing - Emergency)	C&I	X					
DBP (Demand Bidding Program)	C&I	X					
Peak Generation Program	C&I			X			
Summer Saver Program	Residential		X				
Smart Thermostat	Residential		X				
CleanGen Generator Program	C&I		X				
CBP (Capacity Bidding Program)	C&I		X	X			
OBMC (Optional Binding Mandatory Curtailment)	C&I			X			
BIP (Base Interruptible Program)	C&I			X			
Peak Day Credit Program	C&I	X					
SLRP (Scheduled Load Reduction Program)	C&I					X	

2.3. Purpose of this Document

Protocols outline what must be done. They could focus on the output of a study, defining what must be delivered, on how to do the analysis, or both. The protocols provided in this report focus on what impacts should be estimated, what issues should be considered when selecting an approach and what to report, not on how to do the job. The goal is to ensure that the impact estimates provided are useful for planners and operators and that the robustness, precision, and bias (or lack thereof) of the methods employed is transparent.

The best approach to estimating impacts is a function of many factors— resource type, target market, resource size, available budget, the length of time a resource has been in effect, available data, and the purposes for which the estimates will be used. Dictating the specific methods that must be used for each impact evaluation or ex ante forecast would require an unrealistic level of foresight, not to mention dozens if not hundreds of specific requirements. More importantly, it would stifle the flexibility and creativity that is so important to improving the state of the art.

On the other hand, there is much that can be learned from previous work and there are significant advantages associated with certain approaches to impact estimation compared with others. Furthermore, it is imperative that the evaluator have a good understanding of key issues that must be addressed when conducting the analysis, which vary by resource type, user needs, and other factors. As such, in addition to prescribing the deliverables that must be provided with each evaluation, this report also provides guidance and recommendations regarding the issues that are relevant in specific situations and effective approaches to addressing these issues.

> ***The purpose of this document is to establish minimum requirements for load impact estimation for DR resources and to provide guidance concerning issues that must be addressed and methods that can be used to develop load impact estimates for use in long term resource planning. The minimum requirements indicate that uncertainty adjusted, hourly load impact estimates be provided for selected day types and that certain statistics be reported that will allow reviewers to assess the validity of the analysis that underlies the estimates.***

2.4. Report Organization

The remainder of this report is organized as follows. Section 3 provides an overview of evaluation planning and an introduction to some of the issues that must be addressed. It also contains protocols establishing minimum planning requirements. Sections 4, 5, and 6 contain, respectively, protocols associated with ex post evaluation for event based resource options, ex post evaluation for non-event based resources, and ex ante estimation for both event and non-event based resources. These sections also contain detailed discussions of the issues and methods that are relevant to each category of impact estimation. Section 7 discusses issues and challenges associated with the estimation of impacts for portfolios of DR resources and presents the protocol for portfolio information to be included. Section 8 provides an overview of sampling issues and methods. Section 9 contains reporting protocols for LI evaluations for use in planning, and Section 10 describes the protocols for process and review requirements for the load impact evaluations. Appendix A provides a summary of selected studies that provide additional guidance concerning how to approach impact estimation for specific resource options.[16]

[16] The original intent was to include summaries of many more studies in the appendix but there was not sufficient time to complete this work. The studies contained in the appendix are by no means the only examples of exemplary or interesting work in this area.

3. EVALUATION PLANNING

This document contains 27 protocols outlining the minimum requirements for estimation of load impacts for use in long term resource planning. The first three protocols, presented in the following subsection, recognize that good evaluations require careful planning. They also recognize that the minimum requirements established here may not meet all user needs or desires, whether for long term resource planning or for the other potential applications for DR load impact estimates. The remainder of this section discusses the additional requirements that might be met through impact estimation and some of the input data needed to produce impact estimates.

3.1. Planning Protocols (Protocols 1-3)

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. The first three protocols focus on the evaluation planning effort. Protocols 4-27 focus on issues and methods for implementing the evaluation plan. As such, the first load impact estimation protocol requires development of a formal evaluation plan.

Protocol 1:

Prior to conducting a load impact evaluation for a demand response (DR) resource option, an evaluation plan must be produced. The plan must meet the requirements delineated in Protocols 2 and 3. The plan must also include a budget estimate and timeline.[17]

The minimum requirements set forth in Protocols 4-27 indicate that uncertainty adjusted, hourly load impact estimates are to be provided for selected day types and that certain statistics should be reported that will allow reviewers to assess the validity of the analysis that underlies the estimates. Long term resource planners may wish to have additional information that is not covered by these minimum requirements—load impact estimates for additional day types or time periods, for specific customer segments and geographical locations, or for future periods when the characteristics of the DR resource or customer population might differ from what they were in the past. Furthermore, the need for load impact estimates for applications other than long term resource planning may dictate additional requirements. For example, load impact estimation for customer settlement may place a higher priority on methodological simplicity than on robustness and thus require different estimation methods than those used for long term resource planning. Similarly, meeting the operational needs of the CAISO may require greater geographic specificity than is necessary for long term resource planning.

To help ensure that the additional needs of these other stakeholders are considered, Protocol 2 requires that the evaluation plan delineate whether the load impact estimates are intended to be used for purposes other than long term resource planning and, if so, what additional requirements

[17] The final budget and timeline may differ from the planned budget and timeline as a result of the contractor selection process.

are dictated by those applications. Protocol 3 delineates a variety of issues and associated requirements that might be relevant to long term resource planning or to the other applications outlined in Protocol 2. Protocol 3 does not dictate that the load impact estimates meet these additional requirements, only that the evaluation plan indicate whether or not these additional requirements are intended to be addressed by the evaluation and estimation process to which the plan applies.

Protocol 2:

Protocols 4 through 27 establish the minimum requirements for load impact estimation for long term resource planning. There are other potential applications for load impact estimates that may have additional requirements. These include, but are not necessarily limited to:

- ***Forecasting DR resource impacts for resource adequacy;***
- ***Forecasting DR resource impacts for operational dispatch by the CAISO;***
- ***Ex post estimation of DR resource impacts for use in customer settlement; and***
- ***Monthly reporting of progress towards DR resource goals.***

The evaluation plan required by Protocol 1 must delineate whether the proposed DR resource impact methods and estimates are intended to also meet the requirements associated with the above applications or others that might arise and, if so, delineate what those requirements are.

Protocol 3:

The evaluation plan must delineate whether the following issues are to be addressed during the impact estimation process and, if not, why not:

- ***The target level of confidence and precision in the impact estimates that is being sought from the evaluation effort;***
- ***Whether the evaluation activity is focused exclusively on producing ex post impact estimates or will also be used to produce ex ante estimates;***
- ***If ex ante estimates are needed, whether changes are anticipated to occur over the forecast horizon in the characteristics of the DR offer or in the magnitude or characteristics of the participant population;***
- ***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***

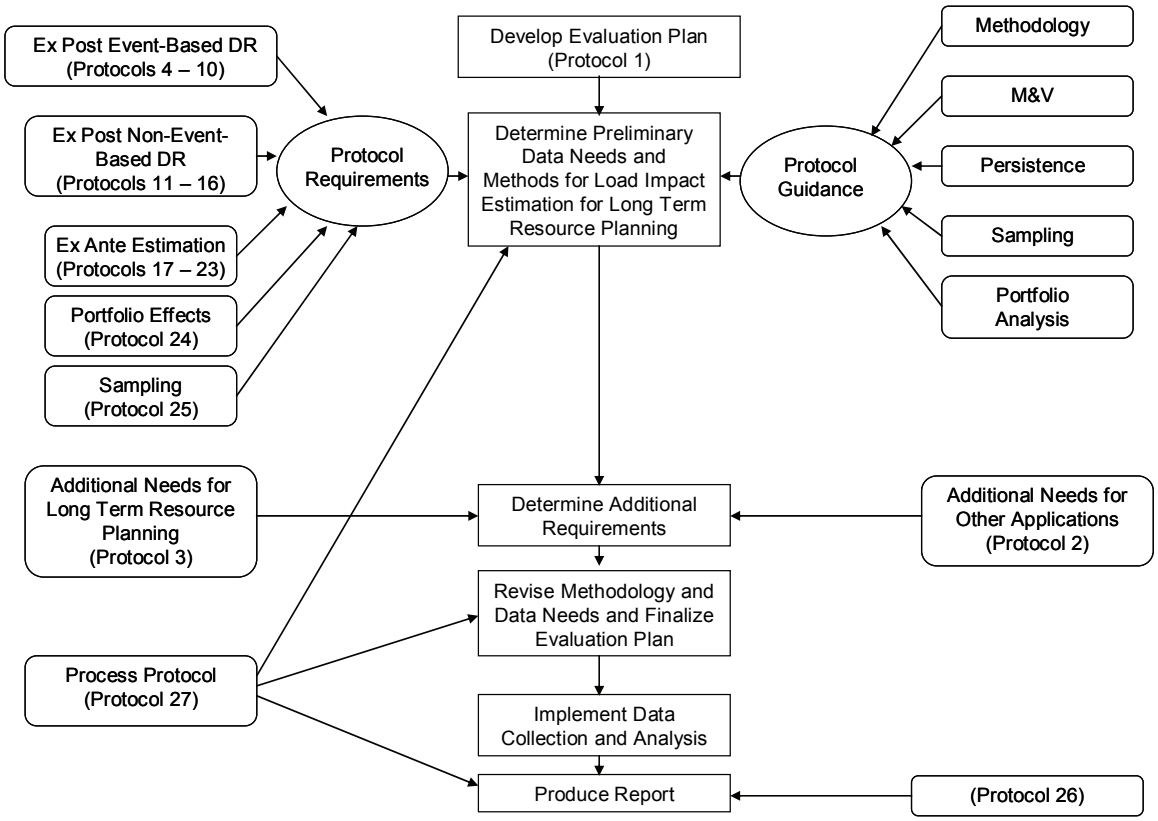
average impacts over time due to changes in customer behavior; changes in average impacts over time due to technology degradation, etc.);

- ***Whether a specified monitoring and verification (M&V) activity is needed to address the above issues, particularly if full evaluations are expected to occur only periodically (e.g., every two or three years);***
- ***Whether it is the intent to develop impact estimates for geographic sub-regions and, if so, what those regions are;***
- ***Whether it is the intent to develop impact estimates for sub-hourly intervals and, if so, what those intervals are;***
- ***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;***
- ***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) or day types (e.g., hotter or cooler days) in addition to the minimum day types delineated in protocols 8, 15 and 22;***
- ***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, the extent to which Measurement and Verification activities will be used to inform this understanding ;***
- ***Whether free riders and/or structural benefiters 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 benefiters or free riders;***
- ***Whether a non-participant control group is appropriate 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; and***
- ***Whether it is the intent to use a common methodology or to pool data across utilities when multiple utilities have implemented the same DR resource option.***

Figure 3-1 depicts a stylized planning process and illustrates how the various protocols and guidance contained in the remainder of this document apply at each step in the process. A preliminary plan can be developed based on the minimum requirements outlined in Protocols 4 through 25. The requirements differ somewhat depending upon the nature of the demand response resource and whether ex ante forecasts are also required. The guidance provided in Sections 4 through 8 can be used to develop a preliminary methodological approach, sampling plan and data development strategy for meeting the minimum requirements. With this initial

plan as a starting point, the evaluator can then determine whether additional requirements are needed to meet the incremental objectives of resource planners or for other applications, such as customer settlement, resource adequacy or CAISO operations. The additional requirements may dictate an alternative methodology, larger samples and/or additional data gathering (e.g., customer surveys). If so, the preliminary plan must be modified prior to implementation.

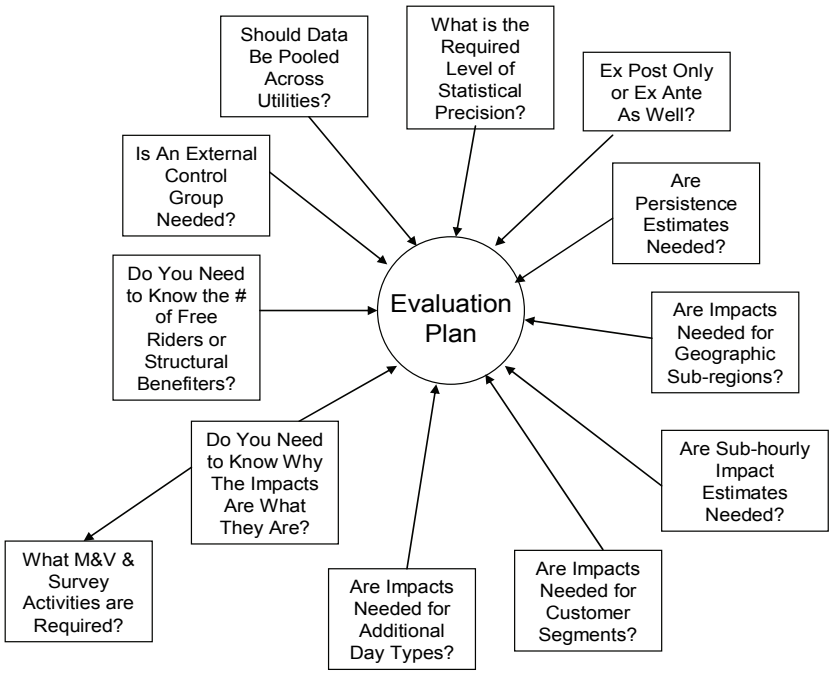
Figure 3-1. Stylized Evaluation Planning Process



3.2. Additional Requirements to be Assessed in the Evaluation Plan

Sub-Section 3.2.1 through 3.2.11 discusses the issues and requirements that must be considered in order to meet the requirements of Protocol 3. Some of these issues are discussed in greater detail in Sections 4 through 8. Figure 3-2 depicts the additional issues and requirements covered under Protocol 3.

Figure 3-2. Additional Requirements Associated With Protocol 3



These additional requirements of the planning process are discussed in sections 3.2.1 through 3.2.11 below.

3.2.1. Statistical Precision

The protocols contained here do not dictate minimum levels of statistical precision and confidence. Several reasons underlie the decision not to establish such minimums. First, and most importantly, the requirements for statistical precision and confidence will vary from resource to resource depending on the needs of the stakeholders who are using the analysis results. In some applications, statistical precision of plus or minus 20% with 80% confidence may be perfectly adequate because other errors in the modeling process (e.g., load forecasts) are known to be at least that large. In other applications, such as estimating/forecasting load impacts for large scale programs that can afford larger sample sizes and employ methods might have higher precision targets. Ultimately, these are considerations that should be dictated by the users of the information after taking into the consideration the costs of the evaluation, and the value of increased accuracy.

Another reason why minimum statistical precision and confidence levels have not been specified is that doing so requires an analysis of benefits and costs associated with increasing sample sizes and this cannot be done in the abstract. The benefits and costs of statistical precision and confidence will vary dramatically from resource to resource depending on a number of factors; the customer segments being sampled, whether interval meters must be installed, the relative size

and importance of the DR resource being evaluated, and the nature of the program impacts being measured.

In short, there are simply too many factors that must be taken into consideration to set minimum levels of precision that would be suitable for all DR resources. On the other hand, setting target levels of precision for a specific evaluation is an important part of the planning process, as it will dictate sampling strategy, influence methodology and be a major determinant of evaluation costs.

3.2.2. Ex Post Versus Ex Ante Estimation

Another important consideration in evaluation planning is whether or not ex ante estimates are needed. There are methodological options that are quite suitable for ex post evaluation but have that have no ability to produce ex ante estimates. Put another way, some methods are suitable for assessing what has happened in the past but can not predict what will happen under future conditions that differ from those in the past. For example, for an event-based resource, comparing loads observed on an event day with reference values based on usage on some set of prior days (referred to as a day-matching methodology) may be quite suitable for ex post evaluation. However, this method is very limited in its ability to predict load impacts that would occur on some future day when weather conditions, seasonal factors or other determinants of load impact may differ from those that occurred during the historical period. Day-matching methods are also not suitable for predicting impacts resulting from changes in customer population characteristics. Ex ante estimation requires methods that correlate impacts with changes in weather and customer characteristics unless loads are not affected by these variables (in which case ex post impacts can be used for ex ante estimation purposes).

Whenever possible, ex ante estimation should be informed by ex post evaluation but ex ante estimation places additional demands on the analysis that aren't necessary if only ex post estimates are needed. Exactly what these additional demands are depends on the extent to which factors are expected to change in the future. For example, it might be that that a set of DR incentives being offered are expected to remain the same over the forecast horizon but changes in the characteristics of the participant population are likely due to planned program expansion or because of a reorientation toward a different target market. In this case, the estimation methodology must incorporate variables that allow for adjustments to the impact estimates which reflect the anticipated changes in participant characteristics. Alternatively, if the participant population is expected to be relatively stable but the incentives (e.g., prices or incentive payments) being offered are expected to change; then, the estimation methodology must incorporate variables that allow predictions to be made for the new prices or incentives. This could require a very different approach to estimation, perhaps one that involves experimentation in order to develop demand models that allow estimates to be made for different price levels.

3.2.3. Impact Persistence

Impact persistence refers to the period of time over which the impacts associated with a DR resource are expected to last. With energy efficiency, impacts for many programs can be expected to last well beyond the life of the program, as EE programs often involve installation of efficient appliances or building shell measures that have long lives. For many DR resources,

impacts can only be expected to occur for as long as the incentives being paid to induce response continue. This is not universally true; however, as some programs may result in upgraded energy management equipment which may continue to provide impacts even after incentives have been discontinued. A permanent load reduction option such as ice storage is another example. Impacts can be expected to persist even if the incentives that led to installation of the measures cease. For other types of resources, such as direct load control of air conditioners, impacts might change over time as load control switches fail and need replacement. For price induced resources, it is possible that demand response will increase over time as participants learn new ways to adjust load or it may decrease over time if consumers decide that the economic savings are not worth the discomfort or inconvenience that are incurred in order to achieve the reductions. Determining the extent to which persistence is an issue and whether or not it is important to predict changes in impacts over time is an important part of the planning process.

3.2.4. Geographic Specificity

Another important consideration is the potential need for geographic specificity. The magnitude of DR impacts will vary by climate zone and participant concentration, and the value of DR varies according to location-specific transmission and distribution constraints and the juxtaposition of load pockets and supply resources. Program planners may want to know the relative magnitude of DR impacts by climate zone and customer characteristics so they can target future marketing efforts. Resource planners may want to know DR impacts for different geographic regions that are dictated by the design of generation, transmission and distribution resources. Both for planning and operational purposes, the CAISO may want to know how DR impacts vary by as many as 30 regions throughout the state. The need to provide impact estimates for various climate zones or other geographic sub-regions will, at a minimum, affect the sampling strategy and could significantly increase sample size. It could also influence methodology, since additional variables may need to be included in the estimation model in order to determine how impacts differ with variation in climate or population characteristics across geographic regions.

3.2.5. Sub-Hourly Impact Estimates

These protocols require that impacts be estimated for each hour of the day for selected day types. For certain types of DR resources and for certain users, estimating impacts for sub-hourly time periods may be necessary. For example, for resources targeted at providing CAISO reliability services, including ancillary services and imbalance energy, sub-hourly impacts may be necessary for settlement and/or operational dispatch.

3.2.6. Customer Segmentation

DR impacts and the optimal methods for estimating them will vary across customer segments. In recent years, large C&I customers have supplied most of the DR resources in California. However, as advanced meters are more widely deployed and dispatchable thermostats become more prevalent, the penetration of demand response among smaller consumers is likely to increase. Issues that affect resource planning vary significantly across these broad customer categories.

For large C&I customers, it is often possible and almost always preferable to use data from all resource participants to estimate load impacts. Most of these customers already have interval meters and the data from these meters is readily obtainable. For these reasons, uncertainty about load impact estimates arising from sampling issues may not be an issue. However, because this customer segment is very heterogeneous, there is the possibility that load impacts from a few very large consumers can dominate the load impacts available at a resource level, thus increasing inherent uncertainty about what the resource will produce on any given day. Large C&I customers also present special challenges in measuring the effects of certain kinds of DR resources. For example, it is often the case that customers above a certain size are required to take service on Time of Use (TOU) rates. When all customers of a given size are required to be on TOU rates, it is virtually impossible to estimate the load impacts of the TOU rate, because there are no customers that can serve as a control group for measuring load shapes that would have occurred in the absence of the rate

With mass market customers, the need for sampling is much more likely, and there are many issues associated with sample design that must be addressed. Unlike load impacts for large C&I customers, load impacts estimated from samples of mass market customers will have some statistical uncertainty. On the other hand, the fact that mass market DR resources may arise from many more customers can also be advantageous in that it provides a robust source of data that can allow for a rich exploration of the underlying causes of demand response. It also can provide more precise estimates of DR impacts that are not subject to wide variation due to the behavioral fluctuations of a few dominant consumers.

Within the broad customer segments discussed above, there may be additional interest in determining whether impacts vary across sub-segments in order to improve resource effectiveness through better target marketing or in order to improve prediction accuracy. It is critical to understand these needs during the planning process, as segmentation could have a significant impact on sample size or may require implementation of a customer survey in order to identify the relevant segments.

3.2.7. Additional Day Types

Still another user-driven consideration is whether there is a need for estimates associated with day-types or days that differ from those required by the protocols outlined below. The output requirements described below are demanding but still try to strike a balance between the diversity of potential user needs and the work required to meet the needs of all potential users. In the ideal world, resource planners would probably prefer impact estimates for all 8,760 hours in a year under an even wider array of weather and event characteristics than those included in these protocols. They might want to know what impacts are likely to be given 1-in-20 weather conditions rather than the 1-in-2 and 1-in-10 weather conditions required by the protocols. The CAISO might want to be able to predict impacts for tomorrow's weather conditions. Some stakeholders may want to know the extent of load shifting to days prior to or following an event day. The evaluator must take these possible needs into consideration when developing an evaluation plan.

3.2.8. 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, the focus is 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 gain an understanding of why the impacts are what they are. If they are larger than what was expected or desired, it might be useful to answer the standard question, "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.

3.2.9. Free Riders and Structural Benefiters

With EE impact estimation, free riders are defined as those customers that would have implemented a measure in the absence of the EE resource stimulus. A significant challenge with EE impact estimation is determining what customers would do in the absence of the resource—that is, sorting out the difference between gross impacts and net impacts. This type of free ridership, which is key to EE impact estimation, is not very relevant to impact estimation for most DR resources as few customers would reduce their load during DR events in the absence of the stimulus provided by the DR resource.

On the other hand, there is another form of free ridership that is relevant to DR impact estimation that stems from the participation of customers who do not use much electricity during DR event periods. This type of free rider is also referred to as a structural benefiter. An example of a structural benefiter is a customer who volunteers for a Critical Peak Pricing (CPP) tariff that does not have air conditioning or typically does not use air conditioning during the critical peak period. Participation by structural benefiters can be viewed as simply reducing historical cross subsidies inherent in average cost pricing. However, some believe that the existence of structural benefiters means that incentive payments will be larger than required to achieve the same level of demand response or, worse, that structural benefiters will not provide any demand response benefits at all. As such, some policy makers may wish to estimate the number of structural benefiters participating in a DR resource option.

When assessing the need to determine the number of structural benefiters that might be participating in a DR program or tariff, it is important to keep a number of things in mind. First and foremost, the methods discussed in sections 4 through 6 are all designed to produce unbiased estimates of demand response. It is not necessary to estimate the number of structural benefiters in order to achieve this goal.

Second, just because a participant's usage pattern might produce a windfall gain from participating in a DR resource program or tariff does not mean that that person will not reduce

their energy use during peak periods. Structural benefiters and non-structural benefiters face the same marginal price signal or incentive and, in theory, should respond in the same manner to those economic incentives. The fact that one group receives a wind fall gain while the other does not does not mean that one group will respond and the other won't. Indeed, any attempt to eliminate structural benefiters could lead to much lower participation in DR programs and tariffs, and much lower overall demand response since structural benefiters are logically more inclined to participate than are non-structural benefiters.

Third, in some instances, it is possible to estimate the magnitude of payment to structural benefiters without having to also estimate the number of structural benefiters. For example, for a peak time rebate option, as long as an unbiased estimate of demand response is obtained for an average customer or for all participating customers, one can estimate the magnitude of payments to structural benefiters by simply using the unbiased demand response impact estimate to calculate the payments associated with demand reductions or load shifting and comparing that value with the amount that was actually paid to participants. The difference will equal the amount of payment to structural benefits based on their preferential usage patterns rather than their change in behavior.

Finally, it is important to keep in mind that estimating the number of structural benefiters can require an entirely different approach to impact estimation than is needed to estimate the average or total demand response. Estimating the average or total response using regression methods can be accomplished using a single equation estimated from data pooled across customers and over time. To estimate the number of structural benefiters, it would be necessary to estimate individual regression equations for every customer using just the longitudinal data available on each customer. While theoretically possible, this approach will not necessarily produce the most efficient or accurate estimate for the group as a whole. Furthermore, doing so will require some minimum number of event days in order to achieve enough statistical precision for individual customers and to avoid concluding that some customers are responding to a price signal when, in fact, they might just be on vacation during several events. In short, there has been very little work done on this issue and the methods that should be used and the circumstances under which they should be applied are largely unproven at this point in time.

3.2.10. Control Groups

The primary goal of impact estimation is to develop an unbiased estimate of the change in energy use resulting from a DR resource. Impacts can be estimated by comparing energy use before and after participation in a DR resource, comparing energy use between participants and non-participants, or both. The primary challenge in impact estimation is ensuring that any observed difference in energy use across time or across groups of customers is attributable to the DR resource, not to some other factor—that is, determining a causal relationship between the resource and the estimated impact.

There are various ways of establishing a causal relationship between the DR resource offer and the estimated impact. One is to compare energy use in the relevant time period for customers before and after they participate in a DR program or, for event-based resources, comparing usage for participating customers on days when DR incentives or control strategies are in place and

days on which they are not. As long as it is possible to control for exogenous factors that influence energy use and that might change over time, relying only on participant samples is typically preferred. Using an external control group for comparison purposes can be costly and can introduce selection bias or other sources of distortion in the impact estimates. When an external control group is needed, it is essential that steps be taken to ensure that the control group is a good match with the participant population in terms of any characteristics that influence energy use or the likelihood of responding to DR incentives. If the control group is not a good match, the impact estimates are likely to be biased.

3.2.11. Collaboration When Multiple Utilities Have the Same DR Resource Options

The final issue that must be considered during evaluation planning arises only when more than one utility has implemented the same DR resource. In this instance, there are a number of advantages to utilities working collaboratively and applying the same methodology to develop the impact estimates. Using the same methodology will help ensure that any differences in impacts across the utilities will be the result of differences in underlying, causal factors such as population characteristics, rather than differences in the analytical approach. Collaboration can also reduce costs and allow for exploration of causal factors that might be difficult to explore for a single utility due to lack of cross-sectional variation. On the other hand, pooling can create challenges as well. For example, two utilities might have very similar dynamic pricing tariffs in place, but operate them independently, possibly dispatching the price signals on different days or over different peak periods on the same days. These operational differences could distort findings based on a pooled sample. Under these circumstances, one might observe impacts that differ across days or time periods and conclude that differences in weather or the timing of an event was the cause when, in fact, the cause of the difference might be due to differences in customer attitudes toward each utility or some other unobservable causal factor.

3.3. Input Data Requirements

An important objective of evaluation planning is determining the type of input data that will be required to produce the desired impact estimates. The type of input data needed is primarily a function of three things:

- The type of impact estimation needed (e.g., ex post estimation for event based resources, ex post estimation for non-event based resources, ex ante estimation);
- The methodology used to produce the estimates; and
- The additional requirements determined as a result of the application of Protocols 2 and 3 (e.g., geographic specificity, customer segmentation, etc.).

Table 3-1 shows how data requirements vary according to the first two factors.[18] This table is not meant to be exhaustive—it is simply meant to illustrate how data needs vary depending upon the application and approach taken and to emphasize the importance of thinking through the input requirements as part of the planning process.

[18] The various methodologies and applications contained in the table are discussed at length in subsequent sections.

Table 3-1. Examples of Variation Input Date Based on Differences in Methodology and Application

Methodology	Ex Post Event Based Resources	Ex Post Non-Event Based Resources	Ex Ante Estimation	
			Participants Similar to the Past	Participants Different from the Past
Day-matching	-Hourly usage for event and reference value days -Customer type[19]	Not Applicable	Not Applicable	Not Applicable
Regression	-Hourly usage for all days -Weather[20]	-Hourly usage for participants -Hourly usage for participants prior to participation and/or for control group -Weather	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios	-Same as prior columns -Survey data on participant characteristics -Projections of participant characteristics
Demand Modeling	-Same as above -Prices	-Same as above -Prices	-Same as prior columns & above row	-Same as prior columns & above row
Engineering	-Detailed information on equipment and/or building characteristics -Weather (for weather-sensitive loads)	-Same as prior column	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios -Projections of participant characteristics
Sub-metering	-Hourly usage for sub-metered loads -Weather for weather sensitive loads	Hourly usage for sub-metered loads for participants prior to participation and/or for control group -Weather for weather sensitive loads	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios -Projections of participant characteristics
Experimentation	-Hourly usage for control & treatment customers -Weather	-Hourly usage for control & treatment customers for pretreatment & treatment periods -Weather	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios	-Same as prior columns -Weather for ex ante day types -Other conditions for ex ante scenarios -Projections of participant characteristics

[19] The best day-matching method may vary across customer segments.
[20] In all cases, weather data must be mapped to the locations of customers in the estimation sample.

Table 3-2 summarizes how input data varies with respect to the additional requirements that may arise from the needs assessments dictated by Protocols 2 and 3. The table entries do not really do justice to the detailed information that may be needed, depending upon the resource options being evaluated and the issues of interest. Data requirements could include:

- Detailed equipment saturation surveys on participant and non-participant populations;
- On-site inspection of technology such as control switches or thermostats to ascertain how many are in working condition;
- Surveys of customer attitudes about energy use and actions taken in response to program or tariff incentives;
- Non-participant surveys to ascertain reasons why customers didn't take advantage of the DR resource option ;
- Surveys of customers who had participated but later dropped out to understand the reasons why they were no longer participating;
- On-site energy audits to support engineering model estimation for impacts;
- Customer bills;
- Zip code data so that customer locations can be mapped to climate zones; and
- Census data or other generally available data to characterize the general population.

In short, the data requirements can be quite demanding and careful thought must be given to determining what data is needed and how best to obtain it.

Table 3-2. Examples of the Variation in Input Data Based on Additional Impact Estimation Requirements

Additional Research Needs	Additional Input Data Requirements
What is the required level of statistical precision?	-*Ceteris paribus*, greater precision requires larger sample sizes.
Are ex ante estimates required and, if so, what is expected to change?	-Incremental data needs will depend on what is expected to change in the future (see Table 3-1)
Are estimates of impact persistence needed?	-Estimating changes in behavioral response over time should be based on multiple years of data for the same participant population. -Estimates of equipment decay could be based on data on projected equipment lifetimes, manufacturer's studies, laboratory studies, etc. -If multiple years of data are not available, examination of impact estimates over time from other utilities that have had similar resources in place for a number of years can be used.
Are impacts needed for geographic sub-regions?	-Data needs vary with methodology. -Could require data on much larger samples of customers (with sampling done at the geographic sub-region level). -Could require survey data on customers to reflect cross-sectional variation in key drivers.
Are estimates needed for sub-hourly time periods?	-Requires sub-hourly measurement of energy use. If existing meters are not capable of this, could require meter replacement for sample of customers.
Are estimates needed for specific customer segments?	-Could require data on much larger samples of customers, segmented by characteristics of interest. -Additional survey data on customer characteristics is needed.
Do you need to know why the impacts are what they are?	-Could add extensively to the data requirements, possibly requiring survey data on customer behavior and/or on-site inspection of equipment.
Do you need to know the number of structural benefiters?	-Could require larger sample sizes and/or additional survey data.
Is an external control group needed?	-Requires usage data on control group. -Survey data needed to ensure control is good match for participant population.
Is a common methodology and joint estimation being done for common resource options across utilities?	-Will likely require smaller samples compared with doing multiple evaluations separately. -May require additional survey data to control for differences across utilities.

4. EX POST EVALUATIONS FOR EVENT BASED RESOURCES

This section contains protocols and guidelines associated with ex post evaluation for event based resource options. There are three broad categories of event-based resources:

- **Event-based Pricing**—This resource category includes prices that customers can respond to based on an event, i.e., a day-ahead or same-day call. This includes many pricing variants such as critical peak pricing (CPP) or a schedule of prices presented in advance that would allow customers to indicate how much load they will reduce in each hour at the offered price (e.g., demand bidding). The common element is that these prices are tied to called events by the utility, DR administrator, or other operator.

- **Direct Load Control**— This resource category includes options such as air conditioning cycling targeted at mass-market customers as well as options such as auto-DR targeted at C&I customers. The common thread is that load is controlled at the customer's site for a called event period through a signal sent by an operator.
- **Callable DR**—This resource category is similar to direct load control but, in this case, a notification is sent to the customer who then initiates actions to reduce loads, often by an amount agreed to in a contract. The difference is that load reduction is based on actions taken by the customer rather than based on an operator-controlled signal that shuts off equipment. Interruptible and curtailable tariffs are included in this category.

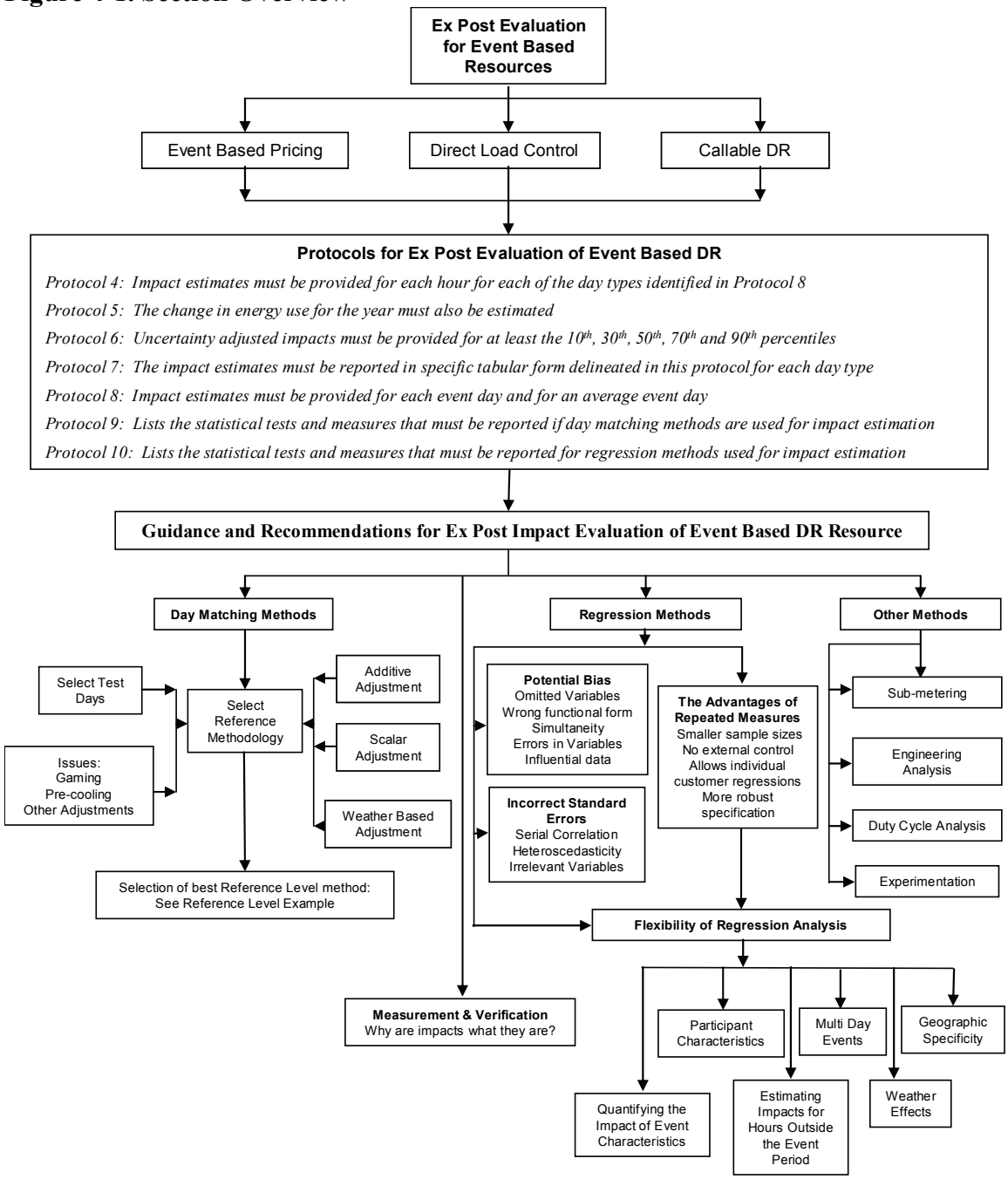
Figure 4-1 provides an overview of the topics covered in this report section. Section 4.1 discusses the seven protocols that outline the minimum requirements for the purpose of conducting ex post impact estimation for event based DR resources. These minimum requirements indicate that uncertainty adjusted, hourly load impact estimates be provided for selected day types and that certain statistics be reported that will allow reviewers to assess the validity of the analysis that underlies the estimates. Section 4.2 contains an overview of many of the issues that will arise when estimating load impacts and provides guidance and recommendations for methodologies that can be used to address these issues.

The three sets of methods for load impact estimation are:

1. Day-matching Methods -- Day-matching is a useful approach for ex post impact estimation and is the primary approach used for customer settlement (i.e., calculating payments to participants) for DR options involving large C&I customers.
2. Regression Methods -- Regression analysis, while more difficult for lay persons to grasp, is more flexible and is generally the preferred method whenever ex ante estimation is also required. As shown in Figure 4-1, while there are technical challenges that must be addressed when using regression analysis, it can incorporate the impact of a wide variety of key drivers of demand response.
3. Other Methods -- Other methods that may be suitable or even preferred in selected situations include sub-metering, engineering analysis, duty cycle analysis and experimentation.

Depending on the circumstances, it may be possible to combine some of these other methods with regression analysis (e.g., estimating models based on sub-metered data or using experimental data). If it is necessary to know not just what the impacts are but also why they are what they are; measurement and verification activities may be required as part of the evaluation process.

Figure 4-1. Section Overview



The three methods shown in figure 4-1 under the box "Guidance and Recommendations for Ex Post Evaluations of DR" are "Day-matching," "Regression Methods," and "Other Methods".

Above this, it states that, regression analysis is more flexible and is generally the preferred method whenever ex ante estimation is also required. Given that statement, why is day-matching methods given so much attention in this document? The reasons are outlined below.

Reasons why day-matching methods is one focus of these protocols:

1. Most of the research on estimating load impacts has involved day-matching methods due to the importance of assessing settlement methods in C&I DR programs. Settlements refers to the method of paying customers for participating in the DR program. This is an important component of DR program design and implementation. While the focus of these protocols is on developing estimates that can be used in resource planning, the extensive literature on day-matching should be explored to determine its potential usefulness. The lack of research on approaches for developing ex ante estimates of DR impacts and the importance of these estimates in developing resource plans is one of the reasons for the stated focus of these protocols, i.e., use in resource planning.

2. Day-matching methods likely will be calculated as part of implementation of most all C&I DR programs since they are used to calculate settlements. This information is essentially produced at no cost to the DR planners developing estimates for forward looking resource plans. As a result, the contribution that these event-day estimates can make to planning should be assessed,[21] and new uses for these estimates might be developed over time. For example:

    a. Day-matching data when available for several years and combined with customer data and event-day data (e.g., weather data) such that influential factors that cause impacts to vary over time and across events can be combined with statistical and regression methods to develop the ex ante estimates needed for planning.

    b. Day-matching methods can be used as a cross check on estimates produced by regression and other methods.

Given the reasons cited above, producing accurate methods of impacts on event days using day-matching methods may provide useful information and enhance approaches for producing ex ante methods needed to develop forecasts of impacts for a relevant planning period.

[21] Some of the reasons why day-matching methods are not viewed as robust as regression approaches include the need to produce estimates in a short time frame for settlements, as a result most day-matching methods are designed to produce estimates within a few days after an event to allow for prompt payments to participants. This limits the amount of data that is used, e.g., regression methods can use an entire season's data and data across multiple events to improve on the accuracy of impact estimates. Forecasting future impacts of DR events is limited with day-matching methods as they usually do not collect data on influential variables that would cause impacts for vary in the future. However, day-matching methods can be combined with regression and other statistical approaches to develop forecasts of impacts if day-matching estimates are available for several years and can be combined specific customer data as well as event-day data such as temperature, and system data.

4.1. Protocols for Ex Post Impact Evaluations – Day-matching, Regression Methods and Other Methods

The protocols discussed in this subsection describe the minimum requirements associated with ex post impact estimation for event based resource options. The protocols outline the time periods and day types for which impact estimates are to be provided, the minimum requirements for addressing the inherent uncertainty in impact estimates, reporting formats, and the statistical measures that provide insight regarding the bias and precision associated with the evaluation and sampling methods. As described in Section 3, additional requirements may be desired in order to meet user needs, including developing estimates for additional day types and time periods, geographic locations, customer segments and other important factors. These protocols are discussed below.

4.1.1. Time Period Protocols (Protocols 4 and 5)

Event-based resources are primarily designed to produce impacts over a relatively short period of time. In addition to impacts that occurred during an event period, spillover impacts such as pre-cooling and snap back cooling might also occur in the hours immediately preceding or following an event period. Some event-based resources might even generate load shifting to a day before or day after an event.

Emergency resources, such as interruptible/curtailable tariffs and direct load control of air conditioning, are typically used only in Stage 1 or Stage 2 emergencies and often for only a few hours in a day. Notification often occurs just a few hours before the resource is triggered or, in the case of load control, with little or no notification at all. The load impacts associated with these resource options often, though not always, are constrained to the event period and perhaps a few hours surrounding the event period. For load control resources, there may be some spillover effects following the end of the event period but there is unlikely to be much impact in the hours leading up to the event unless advance notice of an event is given to participants.[22]

The load impact pattern for price-driven, event-based resources may differ somewhat from that associated with emergency resources in that notification typically occurs sooner, often the day before, and a greater proportion of load reduction during the event period may result from load shifting rather than load reduction. In the residential sector, for example, the dirty laundry doesn't go away during a critical peak period. Some customers will choose to shift their laundry activity to later in the event day, the next day or perhaps even the prior day after receiving notification that the next day will be a high priced day.

Protocols 4 and 5 describe the minimum time periods for which load impact estimates must be provided for event based resources. As discussed in Section 3, additional requirements, such as sub-hourly time periods or other day types, may be necessary to meet the needs of selected users.

[22] This could occur if load control is used in combination with a CPP tariff, for example.

Protocol 4:

The mean change in energy use per hour (kWh/hr) for each hour of the day shall be estimated for each day type and level of aggregation defined in the following Protocol 8. Protocol also calls for the mean change in energy use for the day must also be reported for each day type.

Protocol 5:

The mean change in energy use per year shall be reported for the average across all participants and for the sum of all participants on a DR resource option for each year over which the evaluation is conducted.

4.1.2. Protocols for Addressing Uncertainty (Protocol 6)

One of the most important factors that must be considered when estimating DR impacts is the inherent uncertainty associated with electricity demand and, therefore, DR impacts. Electricity demand/energy use varies from customer-to-customer and within customer from time-to-time based on conditions that vary systematically with weather, time of day, day of week, season and numerous other factors. As such, electricity demand/energy use is a random variable that is inherently uncertain.

In light of the above, it is not sufficient to know the mean or median impact of a DR resource—it is also necessary to know how much reduction in energy use can be expected for a DR event under varying conditions at different confidence levels. For ex post evaluation, uncertainty is largely tied to the accuracy and statistical precision of the impact estimates. For ex ante estimation, uncertainty also results from the inherent uncertainty in key variables such as weather and participant characteristics that influence the magnitude of impacts.

For ex post evaluation, uncertainty can be controlled by selecting appropriate sample sizes, careful attention to sampling strategy, model specification and other means, but it can not be eliminated completely except perhaps in very special situations that almost never occur.[23] Even if data is available for all customers, it is impossible to observe what each customer would have used "but for" the actions they took in response to the DR resource. The "but for" load, referred to as the reference load, must be estimated, and there will be uncertainty in the estimate regardless of what approach is used.[24]

[23] For example, one can imagine a DR resource option that automatically switches off pumps that otherwise are always running and pretty much drawing the same load at all times. In this situation, sub-metering the pumps would provide a highly precise estimate of what the load would have been on the event day if they had not been switched off. However, this is not the typical situation faced by DR impact evaluators.

[24] As discussed in Section 6, with ex ante estimation, uncertainty can also result from the inherent uncertainty associated with key drivers of DR impacts such as weather. If a user wants to know what impacts are likely to occur tomorrow or on a day with a specific weather profile, it is important to recognize that the temperature at 2 pm on the day of interest, for example, is not knowable. It may have a high probability of equaling 92 degrees, say, but it is more realistic to base impact estimates on some distribution of temperatures (preferably derived from historical weather data) with a mean of 92 degrees and a distribution that would indicate, for example, that the temperature has a 90 percent probability of being between 90 and 94 degrees.

Protocol 6 is designed to recognize the inherent uncertainty in impact estimates resulting both from the uncertainty in the estimation methods as well as uncertainty in underlying driving variables when ex ante estimation is required.

Protocol 6:

Estimates shall be provided for the 10th, 30th, 50th, 70th and 90th percentiles of the change in energy use in each hour, day and year, as described in Protocols 4 and 5, for each day-type and level of aggregation described in Protocol 8.

An application of protocol 6 to the production of the information required by the reporting templates (Table 4-1, below) is presented in "Day-Matching Analysis – An Example" on page 54.

4.1.3. Output Format Protocols (Protocol 7)

Impact estimates can be developed using a variety of methodologies. A detailed discussion of the advantages and disadvantages of selected methodologies for event based resource options is provided in Section 4.2. While a variety of methods can be used, two are most common: day-matching and regression analysis.[25]

With day-matching, a reference value representing what a customer would have used on an event day in the absence of the DR resource measure is developed based on electricity use on a set of non-event days that are assumed to have usage patterns similar to what would have occurred on the event days. Impacts are measured as the difference between the reference value and actual loads on the event day.

Regression analysis is an alternative to day-matching. Like day-matching, regression analysis relies on historical information about customer loads, but instead of predicting loads using the averages observed over a given number of previous days, regression analysis focuses on understanding the relationship between loads, or load impacts, during hours of interest and other predictor variables. Examples of predictor variables include temperature, population characteristics, resource effects, and observed loads in the hours preceding the DR event. A detailed discussion of regression analysis is contained in Section 4.2.2.

Regardless of whether day-matching or regression analysis is used, it is possible to report observed load, a reference value and impacts for each event day. For day-matching methods, the impact is calculated as the difference between the reference load and the observed load. For regression methods, the impact estimates can be determined directly from the regression model. These impact estimates can be added to the observed loads in order to estimate a reference value. Protocol 7 indicates the format in which these values should be reported for event based resources. Separate tables should be provided for each day type and, if estimates are developed

[25] Other methods include a comparison of means between control and treatment groups, engineering analysis, sub-metering, etc.

for additional day types, different customer segments or geographic locations, separate tables for each segment, location and day type should be provided.

Protocol 7:

Impact estimates shall be reported in the format depicted in Table 4-1 for all required day types and levels of aggregation, as delineated in Protocol 8.

Table 4-1. Reporting Template for Ex Post Impact Estimates (Separate Tables Shall Be Provided for Each Required Day Type)

Hour Ending	Estimated Reference Load (kWh/hr)	Observed Load (kWh/hr)	Estimated Load Impact (kWh/hr)	Temp (F)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
Daily									

Each variable in Table 4-1 is defined below:

- **Reference Load (Energy Use):** An estimate of the load (average demand) in an hour or total energy use over a period of time that would have occurred "but for" the change in behavior in response to the DR resource offering.

- **Observed Load (Energy Use):** Metered usage in an hour (for load) or over a period of time (for energy).

- **Load (Energy) Impact:** The impact estimate for an hour or over a period of time (e.g., day, season, or year).

- **Temperature:** The average temperature in each hour, measured in degrees Fahrenheit.

- **Uncertainty Adjusted Load (Energy) Impacts:** The estimated load impact value that is likely to be equaled or exceeded X% of the time. For example, if the Uncertainty Adjusted Load Impact at the 10th percentile equals 100 MW, it means that there is a 90 percent probability that the load impact will equal or exceed 100 MW or, alternatively, a 10 percent probability that the impact will be less than 100 MW.

- **Degree Hours:** The difference between temperature in each hour and a base value. For example, if the temperature is 85 degrees in an hour and the base value is 75 degrees, the number of degree hours to base 75 in that hour would equal 10.[26] If the actual temperature is below the base value, the number of degree hours in that hour is set to 0. The number of degree hours in a day is the sum of the degree hours in all hours in the day.

- **Day:** Refers to the day on which an event occurs.

It should be noted that the requirement to report temperature and degree hours in Table 4-1 is designed to allow for easier comparison of impacts across day types, resources and utilities. Inclusion of these variables in the protocols is not intended to dictate that they be used as part of the impact estimation methodology. Other variables, such as relative humidity or some other predictor of weather sensitive load may be more useful than temperature for estimating load impacts. However, a common reporting requirement will facilitate cross-event, cross-resource and cross-utility comparisons.

When reporting temperatures and degree days, it is intended that the temperature be reasonably representative of the population of resource participants associated with the impact estimates. If participation in a resource is concentrated in a very hot climate zone, for example, reporting population-weighted average temperature across an entire utility service territory may not be very useful if a substantial number of customers are located in cooler climate zones. Some sort

[26] Given the significant variation in temperature across a day in many climate zones within California, often rising from the 60s to the 90s or higher between early morning and late afternoon, degree hours may be more informative for comparison purposes across locations than are maximum daily temperature or average temperature. Degree hours are typically better predictors of daily air conditioning load than is average or maximum temperature for a day.

of customer or load-weighted average temperature across weather stations close to participant locations would be much more accurate and useful.

4.1.4. Protocols for Impacts by Day Types (Protocol 8)

DR impacts will vary across event days based on a variety of factors, including variation in usage patterns (often driven by variation in weather), event characteristics (e.g., timing and event duration), event participation, and other factors. In order to understand the influence of these factors on demand response, it is imperative that detailed descriptions of these influencing factors on each event day be provided along with the impact estimates. In addition, for both ex post and ex ante cost-effectiveness analysis, it is useful to have impact estimates for "typical event days". Protocol 8 defines the minimum day types for which impact estimates must be provided and the accompanying information that will aid in interpreting the results.

Among the significant factors that may vary across event days and certainly over time is the number of customers enrolled in a resource, the number who are notified and the number who participate. There is often confusion around these terms so it is useful to define how they are used in the protocols below.

Enrollment is intended to mean the number of customers that have joined a DR program. For any DR programs where a customer needs to take a proactive step in order to enroll, program enrollment equals all customers that have taken that step and are in the program at a given point in time. This can differ significantly from the number of customers who might actually respond during an event or even from the number who are asked to respond for a given event. For any given DR resource, enrollment should be the largest of the three variables.[27]

At a conceptual level, the number of customers notified of an event should equal all those that have actually received the notification. This could differ from the number of notifications sent for various technical reasons (e.g., failure of notification equipment) or because the notification method is not very effective (e.g., it might use a communication channel that doesn't do a good job of reaching its target audience). In most instances, however, there is a pretty high success rate with most notification methods used for DR resources and it is typically much easier to measure the number of notifications sent than it is to measure the number actually received. As such, we define notification as the number of notifications sent out. The number of customers notified may differ from the number of customers enrolled if a resource is geographically targeted and different regions are called on different days or if some other type of dispatch operation is implemented that intentionally does not include all enrolled customers.

[27] There is at least one type of DR resource where enrollment is more difficult to define, namely a peak-time rebate program such as the one outlined by SDG&E in its AMI application. The program concept in that application was that all customers would be eligible to respond to a peak time rebate offering and some subset of the entire customer base would be aware of the offer through promotional schemes. Only customers who were aware would be in a position to respond. Thus, it is difficult to determine whether the number of enrolled customers for such a resource is all customers or just those who are aware and, if the latter, how to measure awareness.

Protocol 8:

The information shown in Table 4-1 shall be provided for each of the following day types and levels of aggregation:

- ***Each day on which an event was called;***
- ***The average event day over the evaluation period;***
- ***For the average across all participants notified on each day on which an event was called;***
- ***For the total of all participants notified on each day on which an event was called; and***
- ***For the average across all participants notified on the average event day over the evaluation period.***

An average event day is calculated as a day-weighted average of all event days.[28] ***The number of event days that apply to each hour may vary for resource options that have variable length event periods.***[29] ***As such, for the average event day, the following information must be provided:***

- ***The number of actual event days included in the calculation for each hour of the average day;***
- ***Average number of customers enrolled in the resource option over the year***[30]***; and***
- ***Average number of customers notified across all event days in the year.***

In addition to the information contained in Table 4-1, the following information must be provided for each event day:

- ***Event start and stop time;***
- ***Notification lead time;***
- ***The number of customers who were enrolled in the resource option on the event day;***
- ***The number of customers who were notified on the event day; and***

[28] Put another way, it is the sum of the impacts in each hour for each event day divided by the number of event days. The reason to think of this as a day-weighted average is because the weights to use when calculating the standard errors are squared.

[29] For example, if there were 10 event days, and the event was triggered from 3 pm to 5 pm on all days and between 5 pm and 6 pm on 5 event days, the average for each hour between 3 pm and 5 pm would be based on all 10 days but the average from 5 pm to 6 pm would be based on the 5 event days on which the event was triggered for that hour.

[30] Since enrollment will change over time, a day-weighted average should be calculated (e.g., if there were 2 event days in the year and there were 100 customers enrolled on the first event day and 200 on the second, the day-weighted average would be 150).

- ***Any other factors that vary across event days that are considered by the evaluator to be important for understanding and interpreting the impacts and why they vary across events.***

4.1.5. Protocols for Production of Statistical Measures (Protocols 9 and 10)

The final protocols that apply to ex post evaluation for event-based resource options concern the calculation and reporting of statistical measures designed to reveal the statistical precision and extent of bias that may be present in the methods used to estimate impacts. The requirements differ between day-matching and regression based methods.

In day-matching, the load impacts of a given resource are measured as the difference between the hourly loads observed on a day of interest (e.g., an event day) and reference values calculated from a set of "matched" days for which similar loads are expected to have occurred. With day-matching methods, calculation of an unbiased reference value is important for the accurate determination of impact estimates. Put differently, the reference values (baseline) must accurately describe not only the load shape, but also the expected demand by hour on event days.

For day-matching methods, it is important to assess bias and overall accuracy that may be present in reference values calculated from day-matching. This is generally accomplished by developing a reference loads estimated for proxy-event days using the day-matching algorithms. Proxy event days are used since it is not possible to observe the loads that that would have occurred if an event had not been called for event days. Proxy event days are selected to be as similar to event days as possible. The actual hourly data on these proxy-event days is compared to the projections from the day-matching algorithm (e.g., the use of 10 days prior to an event). While this is not a direct measure of the accuracy based on actual event days, it is the best information available on the accuracy of a day-matching approach. The accuracy for any day-matching approach is calculated for the selected proxy event days. The basic idea is to assess the accuracy of the day-matching algorithm by observing the errors between projected and actual loads using the proxy event days that are as similar as possible to the actual event days. This method is discussed in more detail below.

Protocol 9 requires evaluators to measure and report the accuracy of the reference values calculated from all day-matching algorithm(s) used to estimate load impacts in given evaluation. There are three steps to this, as follows:

1. Identify a reasonable set of "proxy days" that occurred over a relevant time period. These "proxy days" are days on which the DR resource was not operated and which are as similar as possible to the actual days on which the DR resource was used. As many "proxy days" should be selected as possible, taking care to ensure that these days are indeed similar to the days on which the DR resource was used.
2. Use the day-matching algorithm(s) employed in the study to estimate the loads for each customer on an hourly basis for each proxy day. That is, the evaluator will use the algorithm(s) to estimate the load impacts for each customer and hour for all of the proxy days, just as they would use them to estimate load impacts on the days during which the DR resource is used.

3. Analyze the accuracy of the day-matching algorithm(s) used in the evaluation in terms of the statistics called for in Protocol 9 below.

Protocol 9:

This statistical measures protocol is specific to Day-matching methods. A different protocol (e.g., protocol 10) is appropriate for regression methods. These calculations should be based on a suitable and sufficiently large number of proxy days. From this process, the following statistics should be calculated and reported for day-matching reference value methods:

- ***The number of proxy days used in the calculations below and an explanation of how the proxy days were selected.***
- ***Average error across customers and proxy days for each hour for the entire day. This is calculated as follows:***

$$\overline{e}_{ij\bullet} = \frac{\sum_{i=1}^{n_{cust}} \sum_{j=1}^{n_{days}} \left(L_{ij\bullet} - \hat{L}_{ij\bullet} \right)}{n_{cust} \times n_{days}} \qquad (4\text{-}1)$$

where:

i = ***the cross-sectional unit or customer***

j = ***the event-like day***

$\bullet$ = ***the hour of the day***

$L_{ij\bullet}$ = ***the actual load for the customer on the proxy day of interest for the hour of interest***

$\hat{L}_{ij\bullet}$ = ***the predicted load for the customer on the proxy day of interest for the hour of interest***

n_{cust} = ***the total number of customers in the observation group***

n_{days} = ***the total number of days in the observation group***

- ***Median error across customers and proxy days for each hour for the entire day. The median error is the error corresponding to the exact center of the distribution of errors when all the errors under consideration are arranged in order of magnitude. It is calculated as follows:***

a. ***Calculate the error for each customer and proxy day for the hour of interest:***

b. ***Sort the resulting distribution of*** $n_{cust} \times n_{days}$ ***errors by magnitude for each hour of interest.***

c. ***If the number of errors is odd, the median is the error associated with the*** $\frac{((n_{cust} \times n_{days}) + 1)}{2}$ ***observation.***

d. ***If the number of errors is even, the median is the average of the errors associated with observations*** $\frac{(n_{cust} \times n_{days})}{2}$ ***and*** $\frac{(n_{cust} \times n_{days})}{2} + 1.$

- ***The relative average error for each hour. This is calculated as the ratio of the average error to the average actual load that occurred in the hour:***

$$REL\ e_{ij\bullet} = \frac{\bar{e}_{ij\bullet}}{\sum_{i=1}^{N_{cust}} \sum_{j=1}^{N_{days}} \frac{L_{ij\bullet}}{n_{cust} \times n_{days}}} \qquad (4\text{-}2)$$

where:

$\bar{e}_{ij\bullet}$ = ***the average error across customers and proxy days for the hour of interest***

- ***The relative median error for each hour. This is calculated as follows:***

$$REL\ e^{M}_{ij\bullet} = \frac{e^{M}_{ij\bullet}}{L^{M}_{ij\bullet}} \qquad (4\text{-}3)$$

where:

$e^{M}_{ij\bullet}$ = ***the median error across customers and proxy days for each hour for the entire day, as calculated above***

$L^{M}_{ij\bullet}$ = ***the median load for the customer on the proxy day of interest***

- ***The Coefficient of Alienation***[31]***, which describes the percentage of the variation in actual load for each hour that is not explained by variation in the predicted load. This is calculated as follows:***

$$\sum_{i=1}^{n_{cust}} \sum_{j=1}^{n_{days}} \sum_{k=1}^{n_{hours}} \frac{\left(L_{ijk} - \hat{L}_{ijk}\right)^2}{\left(L_{ijk} - \overline{L}_{jk}\right)^2} \qquad (4\text{-}4)$$

where:

i = ***the cross-sectional unit or customer***

j = ***the event-like day***

k = ***the hour of the day***

L_{ijk} = ***the actual load for the customer on the proxy day of interest for the hour of interest***

$\hat{L}_{ijk}$ = ***the predicted load for the customer on the proxy day of interest for the hour of interest***

$\overline{L}_{jk}$ = ***the average load on the proxy day of interest for the hour of interest***

n_{hours} = ***the total number of hours being observed on the proxy day***

- ***Theil's U, calculated as follows:***

$$\frac{\left[\frac{1}{n_{hours}} \sum_{k=1}^{n_{hours}} \left(L_k - \hat{L}_k\right)^2\right]^{1/2}}{\left[\frac{1}{n_{hours}} \sum_{k=1}^{n_{hours}} \left(L_k\right)^2\right]^{1/2} + \left[\frac{1}{n_{hours}} \sum_{k=1}^{n_{hours}} \left(\hat{L}_k\right)^2\right]^{1/2}} \qquad (4\text{-}5)$$

[31] The Coefficient of Alienation is a measure of the error in a prediction algorithm (of any kind) relative to the variation about the mean of the variable being predicted. It is related to the Coefficient of Determination by the function $k = (1-R^2)$. The Coefficient of Determination is a measure of the goodness of fit of a statistical function to the variation in the dependent variable of interest. Correspondingly, the Coefficient of Alienation is a measure of the "badness of fit" or the amount of variation in the dependent variable that is not accounted for by the prediction function. The R^2 obtained from regression analysis is a special case of the Coefficient of Determination in which the regression function is used to predict the value of the dependent variable. Coefficients of Determination and Alienation can be calculated for virtually any algorithm that makes a prediction of a dependent variable.

where:

n_{hours} = ***the number of periods***

k = ***the period of interest***

L_k = ***the actual observed load for the period of interest***

$\hat{L}_k$ = ***the predicted load for the period of interest***

Theil's U describes the accuracy of a forecasted data series. As U approaches zero, the forecast is judged to be more accurate, and as it approaches one, the forecast does no better than a naïve prediction of the future that assumes no trend. Because U describes the accuracy of a forecast for a particular individual in the population over a given period of time, it is particularly useful for evaluating the performance of day-matching algorithms that do not depend on regression adjustments. To evaluate the goodness of fit over a population of forecasts (i.e., over a group of participants on a given day or series of days) it is necessary to calculate Theil's U for each forecast and then analyze this distribution of errors as indicated by the Theil's U calculations. The characteristics of this distribution, including mean and median, should be described.[32]

For regression methods, a different protocol for statistical measures is appropriate. The regression protocol is designed with two goals in mind:

1. Provide qualified reviewers with sufficient transparency and information so as to enable a thorough assessment of the validity, accuracy, and precision of the results;

2. Provide the information necessary to enable readers to create models that provide the load impacts and the confidence intervals under specific scenarios.

Protocol 10:

For regression based methods, the following statistics and information shall be reported:

- ***Adjusted R-squared or, if R-squared is not provided for the estimation procedure, the log-likelihood of the model;***[33]

[32] For examples of how Theil's U can be applied, see KEMA-XENERGY (Miriam L. Goldberg and G. Kennedy Agnew). *Protocol Development for Demand Response Calculation—Findings and Recommendations.* Prepared for the California Energy Commission, February 2003.

[33] The log-likelihood is a standard output whenever a maximum likelihood method (vs. OLS) is employed. Most statistical packages produce the log-likelihood (or do so by default) when a maximum likelihood estimation is used. Many statistical packages will show the changes to the log-likelihood as the computer goes through the iterative process of finding the best fittings set of parameters. The log-likelihood may be expressed as a pseudo R-square as that may be more familiar to some researchers. The protocols request for the R-square or, if the R-squared is not available, the log-likelihood. The log-likelihood is often used for equations where the dependent variable is a takes on discrete values. This Logit or Tobit type models do not typically produce R-squared values. For example, an A/C cycling evaluation that relies on directly metered A/C units should be, theoretically, analyzed with Tobit regression

- ***Total observations, number of cross-sectional units and number of time periods;***
- ***Coefficients for each of the parameters of the model;***
- ***Standard errors for each of the parameter estimates;***
- ***The variance-covariance matrix for the parameters;***[34]
- ***The tests conducted and the specific corrections conducted, if any, to ensure robust standard errors; and***
- ***How the evaluation assessed the accuracy and stability of the coefficient(s) that represent the load impact.***

4.2. Guidance and Recommendations for Ex Post Evaluation of Event Based Resources – Day-matching, Regression, and Other Methods

Section 4.1 delineated the key requirements associated with estimating ex post impacts for event based DR resources. The protocols describe what must be provided, not how to do the job. This section discusses a variety of issues that should be considered when deciding "how to do the job" and, where appropriate, provides guidance and recommendations concerning how these issues might be addressed.

Two primary methods have typically been used to estimate load impacts for DR resources, day-matching and regression analysis. Day-matching is a useful approach for ex post impact estimation and is the primary approach used for customer settlement for resource options involving large C&I customers. Regression analysis is more flexible and is generally the preferred method whenever ex ante estimation is also required. Other methods that may be suitable or even preferred in selected situations include sub-metering, engineering analysis, duty cycle analysis and experimentation. Depending on the circumstances, it may be possible to combine some of these other methods with regression analysis (e.g., estimating models based on sub-metered data or using experimental data).

because for many hours the A/C unit will have zero usage due to either low temperature or no one at home. In other words, it is a dependent variable (e.g., energy usage) is truncated at a value of zero. The Tobit output will likely not produce an R-squared in which case the log-likelihood is the standard output. Peter Kennedy, *A Guide to Econometrics*, Fifth Edition, MIT Press, 2003 on p. 23-24 and 42-46) discusses maximum likelihood estimation. Another source is Woolridge, *Econometric Analysis of Cross Section and Panel Data,* Chapter 13; and W Green's textbook on *Econometric Analysis,* Chapter 17. SAGE publications has published a booklet titled *Maximum Likelihood Estimation*. Any of the above references will illustrate the use of the log-likelihood of a model. *(Source: Communication with Mr. Josh Bode, Freeman, Sullivan & Co.)*

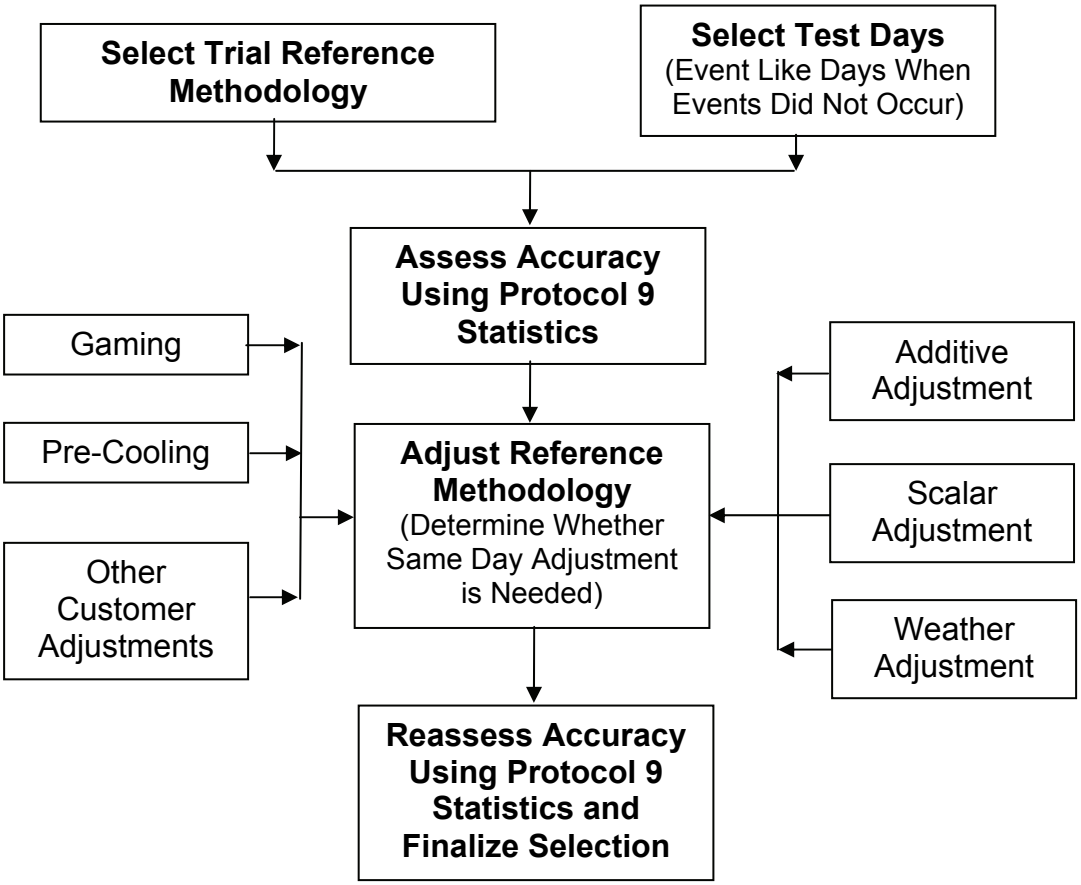
[34] The variance-covariance matrix is needed in order to calculate the correlations between the model parameters for use in determining forecast precision and uncertainty bands.

4.2.1. Day-matching Methodologies

With day-matching, DR impacts are estimated as the difference between a reference value, intended to represent what load would have been had a customer not changed their behavior in response to the DR program or tariff incentive, and actual load on an event day. Developing reference load shapes involves either two or three steps, depending on the nature of the load. The first step involves selecting relevant days and the second involves taking an average of the load in each hour for the days that were chosen. 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 some other variable such as weather on prior days and event days.

As discussed in the previous section, event-like days (e.g., days similar to event days but on which events are not called) should be used to test the accuracy of the reference value based on the various statistics contained in Protocol 9. Figure 4-2 summarizes the process for the best reference value methodology. Additional details are provided below.

Figure 4-2. Reference Load Selection Process



When considering what days to choose for the initial reference load calculation, for C&I customers, only business days are typically used. For residential customers, if events only occur on weekdays, weekends would logically be excluded from day selection as usage on weekends tends to be different on average from weekday usage. When it comes to using day-matching, one size definitely does not fit all. What works best will vary with customer type, load shape, whether or not the load is weather sensitive, and other factors. On the other hand, an objective is to provide some consistency in the impact estimates across resource options to allow for valid comparisons. 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 period;[35]
- 20 days bracketing the event day; and
- All relevant days in an entire season.

"Same-day" adjustment options include:[36]

- **Additive Adjustment:** A constant is added to the provisional reference value for each hour of the curtailment period. For simple additive adjustment, the constant is calculated as the difference between the actual load and the provisional reference value load for some period prior to the curtailment. Ad hoc or judgmental adjustments are also possible.
- **Scalar Adjustment:** The provisional reference value load for each hour of the curtailment period is multiplied by a fixed scalar. The scalar multiplier is calculated as the ratio of the actual load to the provisional reference value load for some period prior to the curtailment.
- **Weather-Based Adjustment:** A model of load as a function of weather is fit to historical load data. The fitted model is used to estimate load (a) for the weather conditions of the days included in the provisional reference value, and (b) for the weather conditions of the curtailment day. The difference or ratio of these two estimates is calculated, and applied to the provisional reference value as an additive or scalar adjustment.

With the additive or scalar adjustment, the two hours prior to an event and the two hours prior to that (e.g., the 3rd and 4th hours prior to the event period) have been tested. There are at least three

[35] 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.
[36] This discussion is based on information in KEMA-XENERGY (Miriam L. Goldberg and G. Kennedy Agnew). *Protocol Development for Demand Response Calculation—Findings and Recommendations.* Prepared for the California Energy Commission, February 2003. p. 2-12. This report uses the term baseline for what we call reference value. Hereafter, we refer to this report as the KEMA/CEC study.

concerns that must be addressed if the two hours prior to an event period are used to adjust an initial reference value for evaluation purposes.

- **Gaming**—if the two hours prior to the event period are also used as part of the reference value for customer settlement, and this is known by the customer, a customer might intentionally increase energy use in the hours leading up to the event period in order to increase their reference value so as to receive a higher payment.
- **Pre-cooling**—a customer might increase cooling in the hours leading up to the event period in order to retain their comfort level longer if, for example, air conditioning is being controlled during the event.
- **Other pre-event adjustments**—a C&I customer might reduce manufacturing or business operations in anticipation of the event period.

If gaming or pre-cooling occurs, impact estimates based on the two hours prior to the event period will be overstated whereas anticipatory behavior by customers, such as canceling production runs or encouraging office workers to work at home, could lead to under estimation of load impacts. These inaccuracies could still arise when earlier hours in the day are used rather than the two hours prior to the event period, but the bias may be smaller. On the other hand, for weather sensitive loads, using the earlier hours in a day may not be as accurate if temperatures increase significantly as the day progresses.

A variety of research has been done to compare the accuracy and other attributes of various day-matching methods. A useful study was completed in 2003 for the California Energy Commission and should be reviewed if a day-matching approach is being considered. The KEMA/CEC study examined the relative accuracy, simplicity and other factors associated with a number of day-matching methods using data on 646 large C&I customers from utilities scattered throughout the country.

The KEMA/CEC analysis concluded that the reference value calculation method that worked best for a range of load types consists of taking a simple average of the last 10 days of demand data, by hour of the day, and then shifting the resulting profile up or down so that it matches the average observed load for the period 1 to 2 hours prior to curtailment. This method worked well for both weather-sensitive and non-weather sensitive accounts, with both low and high variability, for summer and non-summer curtailments.

The KEMA/CEC study went on to report that, if the default method is problematic either because of the potential for customer gaming or because of a need to curtail more promptly, the next best alternative depends on the weather sensitivity and energy use variability of the account. The default reference value and alternatives that performed reasonably well for different types of accounts and seasons are shown in Table 4-2.

Table 4-2. Findings from the KEMA/CEC Study

Methods for Estimating a Baseline Energy Use Profile
By Customer Account Type and Event-Specific Baseline

Season:	Summer				Nonsummer			
Weather Sensitivity:	Weather-Sensitive		Non-Weather-Sensitive		Weather-Sensitive		Non-Weather-Sensitive	
Variability:	Low	High	Low	High	Low	High	Low	High
DEFAULT Baseline Profile : 10 day average baseline load profile (BLP) plus additive adjustment based on actual energy use for 1-2 hours prior to demand response signal.	X	X	X	X	X	X	X	X
ALTERNATIVES:								
10 day average BLP plus additive adjustment based on actual load for three to four hours prior to demand response signal.			X	X			X	X
Weather model to construct BLP based on actual average outside temps when DR signal sent; diagnostics model used to determine if building loads are suffficiently weather sensitive to use this approach.	X	X	X					
Use only the highest five of the last 10 days to construct BLP, with scalar adjustment based on a Temperature-Humidity Index load model.	X	X				X		
Use only the highest five of the last 10 days to construct BLP, without any adjustment for control day.					X			
Use the highest 10 of the last 11 days to construct BLP without adjustment for control day conditions.			X	X			X	X

Baseline load proflie- The kw demand levels for each hour of a 24 hour day averaged over the previous 10 non curtailment days

Additive adjustment: The adjusted BLP is calculated by adding a fixed amount A to the 10 day average BLP for each hour.

Demand response signal: Price or emergency signal sent to request reduced electric loads for a fixed time period

Scalar adjustment: The adjusted BLP is calculated by multiplying the 10 day average DLP for each hour by a fixed amount S.

Weather model: Hourly load data from non-curtailed days is fit by a regression model using weather and calendar variables. The BLP is the fitted model applied to the observed conditions during and after a DR signal.

Analysis done by San Diego Gas & Electric (SDG&E) in support of its advanced metering application also found that same-day adjustment improves reference value calculations. SDG&E used data on roughly 340 residential customers from the year 2004 to examine the relative accuracy and bias associated with more than two dozen reference value methodologies. Methodologies using 3, 5 and 7 prior days, with and without various forms of adjustment, were examined. Average error and the sum-of-squared errors (SSE) were calculated for each method. Average error was much closer to 0 for the methods using same-day adjustment, and the SSE was also among the lowest for these methods.[37]

Day-matching methods are easy to understand and often easier to produce and use than regression methods. With same-day adjustment, day-matching methods exist that have very small average errors and that are reasonably precise. If the primary question is—"What was the DR impact for some set of historical event days, or for individual event days?"—day-matching can be an intuitively appealing and practical approach.

However, there are certain challenges with day-matching methods even when ex post estimation is the primary focus. One problem arises when there is significant variation in customer loads across days. When this occurs, a reference value based on average usage over even a large number of days still may not be a good proxy for what the load would have been on an event day in the absence of the event. If there is less variation in the loads that are contributing to the DR

[37] SDG&E AMI Proceeding (A.05-03-015). DRA Exhibit 109.

impact than there is in the total customer load, it may be possible to use day-matching analysis with sub-metered data for these partial loads.

One practical problem with day-matching is that there are no established approaches for calculating the statistical uncertainty associated with ex post load impact estimates (e.g., the estimates in the right hand columns of Table 4-1). The proxy event-day approach outlined in Protocol 9 shows which day-matching methods have the best fit, but these statistics, by themselves, do not estimate the uncertainty associated with a day-matching method. There is a need to explore new methods for assessing uncertainty in day-matching methods.

The problem of estimating uncertainty adjusted load impacts using day-matching is more complicated when load impacts from more than one customer are aggregated to achieve an overall resource load impact estimate. This is because the uncertainties associated with multiple load impact estimates must be combined.

Calculating average and aggregate resource impacts is relatively straightforward. The sum of the load impacts for the program can be obtained by summing over the load impacts observed for each of the participants, and the average load impact can be similarly obtained by dividing by the number of participants. However, procedures for estimating the uncertainty in these load impacts are neither well developed nor extensively tested.[38]

---

[38] There are several ways to approach this calculation. Three are outlined below:

1. This approach involves estimating the standard error of the aggregate estimates by calculating the between and within variances for the participants for each hour. The uncertainty in the aggregate load impact estimate has two components – one arising from variation of the participant means around the mean for all participants and the other arising from variation in the loads used to estimate the reference load for each hour in question. This approach calculates the uncertainty in the aggregate load impact estimates is to calculate the standard error of the estimate by combining these two known variance components. This is essentially the standard error of the aggregate load impact estimate, which in turn can be used to identify the upper and lower limits of the calculation.
2. Alternatively, it is possible to describe the uncertainty in the aggregate load impact estimate using Monte Carlo simulation to sample repeatedly from the population of participants using the range of uncertainty observed for each of the participants.
3. The third approach may be the most straight-forward method. This approach takes the statistical data from Protocol 9 which are used to select the day-matching method that is most accurate, taking into account bias in the method. Once the day-matching method is selected, it is possible to calculate the standard deviation and variance by comparing the estimated loads with the actual loads on event-like days (see the three steps, page 43-44). For each hour of the event, the estimated variance is calculated by taking the sum of the differences between each estimate of the estimated load from the selected day matching method and the mean value for all the estimated values divided by n-1. The value "n" is the size of the sample which will be determined by the number of event-like days. In this case, "n" should be the number of event-like days. The standard deviation is simply the square root of the variance. These equations are available in all statistics text books. The problem with this method is that the estimated variance for a set of actual event days is assumed to the same as the variance calculated for the event-like days used in protocol 9. While not an exact variance calculation using the actual data from the event days, it may be the best information available on the likely variance for a day-matching method for the actual event days.

The estimation of the variance in the estimates of hourly loads on event days will benefit from additional thought and research. It is hoped that the evaluation planning phase will bring out approaches that best address the uncertainty in the day-matching methods. See the "Day-Matching Analysis – An Example" below for an application of uncertainty analysis.

The use of day-matching methods to estimate ex ante load impacts and their uncertainty is even more difficult. Indeed, with standard day-matching approaches, there is no mathematical function or "bridge" that can be used to relate conditions that are in effect on a given ex ante day type to the load impact that will occur under the prescribed conditions. One can imagine ways of approaching the problem, such as regressing day-matching based impact estimates against explanatory variables such as weather, event characteristics and customer characteristics. However, if regression analysis is needed in order to build a bridge between estimated impacts using day-matching and relevant explanatory variables, it is probably better to simply use regression methods directly as the statistical properties will be much easier to calculate and interpret. For this reason, we do not recommend day-matching as a suitable approach when the primary focus is on ex ante estimation for day types that differ from those that have occurred historically. On the other hand, if the objective is a simple, straightforward way to develop an ex post estimate, day-matching methods can be useful as a way of quickly reporting DR results.

Day-matching Analysis: An Example

For day-matching methods, the protocols require estimates of uncertainty-adjusted energy impacts for each event day hour (i.e., average hourly demand), for each event-day as a whole, and for the year (Protocols 4 -8). In addition, the protocols also require several statistics that reflect the bias (or lack of bias) and predictive capability of the reference methods tested and selected. This section provides an example of how those protocols can be met using a day-matching methodology.

The example was developed using 2005 data from a random sample of 50 (out of 114) large C&I customers on SDG&E's voluntary critical peak pricing tariff. For all the CPP participants, electricity prices varied according to peak, semi-peak, and off-peak hours. In addition, the rate allowed for a maximum of 12 CPP operations. In days where a CPP event was called, participants paid roughly 33 cents/kWh from 11 am to 3 pm (CPP Period 1) and 115 cents/kWh from 3 to 6 pm.

The analysis is for illustrative purposes only. The load impacts may be better estimated by using control groups, pre-treatment data, and/or regression methods.

The first step in estimating impacts using day-matching is to select an appropriate day-matching method. Identifying the candidate day-matching methods and selecting the final method used to calculate load impact estimates is left to the evaluator's discretion. However, evaluators are required to calculate statistics that describe the accuracy (i.e., lack of bias) and precision of the day-matching method(s) tested and selected (Protocol 9). Candidate day-matching methods should first be identified based on their accuracy (lack of bias) before considering statistical precision. Put differently, selecting biased reference values with narrow confidence intervals is an example of false precision.

The accuracy and predictive capability of day-matching methods is observed by comparing the predicted and actual loads for proxy days. By using proxy days, it is possible to compare the loads predicted by candidate day-matching methods with the actual loads that occurred. In order

to select the event-like days, each day in 2005 was ranked based on the SDG&E daily peak. In total, four out the five events were among the ten highest SDG&E system load days. Of the remaining six non-event days, two were on weekends. As a result, the four remaining non-event weekdays were employed as proxy days and used to assess the accuracy and predictive capability of each method. Table 4-3 summarizes the dates, SDG&E daily peak, and event/non-event status for the ten days with the highest SDG&E daily peaks.

In total, three day-matching methods for determining reference values were evaluated:

- A three day baseline adjusted by the load an hour before the peak period;
- The 5-day adjusted by the load on the day before the event; and
- The load on the day prior to the event.

Table 4-3. Proxy Day Selection

Date	Day Type	SDG&E Daily Peak (MW)	Avg. Load During Peak Period (11am-6pm)	Daily Peak Rank for 2005	Proxy Day
Friday, July 22, 2005	Event-day	4,057.2	3,916.4	1	
Monday, August 29, 2005	Non-event weekday	4,031.5	3,869.3	2	✓
Friday, August 26, 2005	Event-day	3,995.3	3,834.4	3	
Thursday, July 21, 2005	Event-day	3,985.0	3,848.5	4	
Thursday, August 25, 2005	Non-event weekday	3,947.2	3,748.2	5	✓
Wednesday, July 20, 2005	Non-event weekday	3,821.3	3,508.9	6	✓
Saturday, August 27, 2005	Weekend or holiday	3,799.3	3,679.0	7	
Tuesday, August 30, 2005	Non-event weekday	3,753.3	3,571.7	8	✓
Thursday, September 29, 2005	Event-day	3,734.8	3,632.5	9	
Sunday, August 28, 2005	Weekend or holiday	3,712.9	3,597.3	10	

For each of the methods used to determine reference values, the statistics used to evaluate accuracy (lack of bias) and predictive capability were calculated, as prescribed by Protocol 9. Protocol 9 requires that four statistics used to assess accuracy be calculated on an hourly basis across proxy or event-like days. It also calls for two measures of predictive capability across event-like days: the Coefficient of Alienation and Theil's U. The Coefficient of Alienation describes the share of variation in loads unexplained by the method. Theil's U measures the naivety of the predictions; that is, if the Theil's U statistic is less than one, then the method used for predicting reference value is better than guessing. The closer the value gets to 0 the more accurate the projection. Table 4-4 and Table 4-5 reflects the predictive capability and accuracy statistics for each of the methods evaluated.

Table 4-4. Comparison of Day-matching Methods – Predictive Capability Statistics

Day-Matching Method	Coefficient of Alienation	Theil's U
3-day average with day-of adjustment	3.740%	0.12104
5 day average with prior-day adjustment	3.736%	0.18109
Prior day, no adjustment	3.740%	0.19428

Table 4-5. Comparison of Day-matching Methods –Accuracy Statistics

			3-day average with day-of adjustment				5 day average with prior-day adjustment				Prior day, no adjustment			
Hour Ending	Average Load (kW)	Median Load (kW)	Average Error	Median Error	Relative Average Error	Relative Median error	Average Error	Median Error	Relative Average Error	Relative Median error	Average Error	Median Error	Relative Average Error	Relative Median error
1	340.28	278.56	1.66	2.30	0.5%	0.82%	-2.04	0.00	-0.6%	0.00%	-8.22	-0.32	-2.4%	-0.11%
2	334.99	269.08	-6.01	1.35	-1.8%	0.50%	-8.52	-0.03	-2.5%	-0.01%	-13.22	-0.57	-3.9%	-0.21%
3	326.28	261.44	-1.40	2.56	-0.4%	0.98%	-6.49	0.00	-2.0%	0.00%	-7.43	-0.32	-2.3%	-0.12%
4	311.36	242.48	5.11	3.10	1.6%	1.28%	0.03	0.79	0.0%	0.33%	-4.26	-0.32	-1.4%	-0.13%
5	312.24	244.20	7.07	3.06	2.3%	1.25%	2.66	0.13	0.9%	0.05%	1.23	0.00	0.4%	0.00%
6	314.09	237.76	2.87	0.36	0.9%	0.15%	-3.95	0.03	-1.3%	0.01%	1.28	0.00	0.4%	0.00%
7	330.54	272.58	6.32	0.75	1.9%	0.28%	3.37	0.00	1.0%	0.00%	-0.31	0.00	-0.1%	0.00%
8	341.11	306.72	6.69	0.44	2.0%	0.14%	3.13	-0.08	0.9%	-0.03%	1.79	-0.02	0.5%	-0.01%
9	350.18	320.24	4.79	0.00	1.4%	0.00%	0.66	-1.49	0.2%	-0.47%	-1.99	-0.01	-0.6%	0.00%
10	353.79	331.20	12.38	-0.35	3.5%	-0.10%	5.35	-1.24	1.5%	-0.38%	0.51	-0.35	0.1%	-0.11%
11	352.13	331.20	4.93	-0.89	1.4%	-0.27%	0.91	-3.85	0.3%	-1.16%	-3.49	-0.23	-1.0%	-0.07%
12	348.06	293.42	-6.61	-1.87	-1.9%	-0.64%	-7.68	-2.98	-2.2%	-1.02%	-6.53	0.00	-1.9%	0.00%
13	336.86	249.00	-5.47	-0.18	-1.6%	-0.07%	-7.89	-2.59	-2.3%	-1.04%	-2.70	0.00	-0.8%	0.00%
14	323.29	138.70	-12.55	-0.20	-3.9%	-0.14%	-12.34	-0.35	-3.8%	-0.25%	-9.81	0.00	-3.0%	0.00%
15	320.01	126.08	-13.60	0.00	-4.3%	0.00%	-13.66	-0.17	-4.3%	-0.14%	-11.03	0.00	-3.4%	0.00%
16	324.54	195.36	-0.88	0.04	-0.3%	0.02%	-8.57	0.00	-2.6%	0.00%	-0.22	0.00	-0.1%	0.00%
17	329.25	211.50	-3.23	0.20	-1.0%	0.09%	-10.68	0.00	-3.2%	0.00%	-7.00	0.00	-2.1%	0.00%
18	326.35	217.60	-6.29	0.91	-1.9%	0.42%	-10.21	0.00	-3.1%	0.00%	-12.48	0.00	-3.8%	0.00%
19	335.90	252.00	-5.16	0.80	-1.5%	0.32%	-10.67	0.00	-3.2%	0.00%	-13.86	0.00	-4.1%	0.00%
20	351.24	296.40	-5.16	0.00	-1.5%	0.00%	-17.35	-1.36	-4.9%	-0.46%	-12.91	0.00	-3.7%	0.00%
21	342.84	267.56	6.77	0.52	2.0%	0.19%	-8.39	-1.33	-2.4%	-0.50%	-3.47	0.00	-1.0%	0.00%
22	331.87	261.12	13.81	1.02	4.2%	0.39%	-1.17	-1.06	-0.4%	-0.41%	1.15	0.00	0.3%	0.00%
23	338.63	293.40	-0.69	0.75	-0.2%	0.25%	-11.68	-1.30	-3.4%	-0.44%	-12.81	0.00	-3.8%	0.00%
24	345.11	300.96	-5.36	0.60	-1.6%	0.20%	-17.69	-1.52	-5.1%	-0.51%	-15.61	0.00	-4.5%	0.00%
Daily	334.21	263.90	0.00	0.10	0.0%	0.04%	-5.95	-0.15	-1.8%	-0.06%	-5.89	0.00	-1.8%	0.00%
Peak	329.77	211.50	-6.95	0.00	-2.1%	0.00%	-10.15	-0.09	-3.1%	-0.04%	-7.11	0.00	-2.2%	0.00%
Off-peak	336.03	274.20	2.86	0.52	0.9%	0.19%	-4.23	-0.18	-1.3%	-0.07%	-5.39	0.00	-1.6%	0.00%

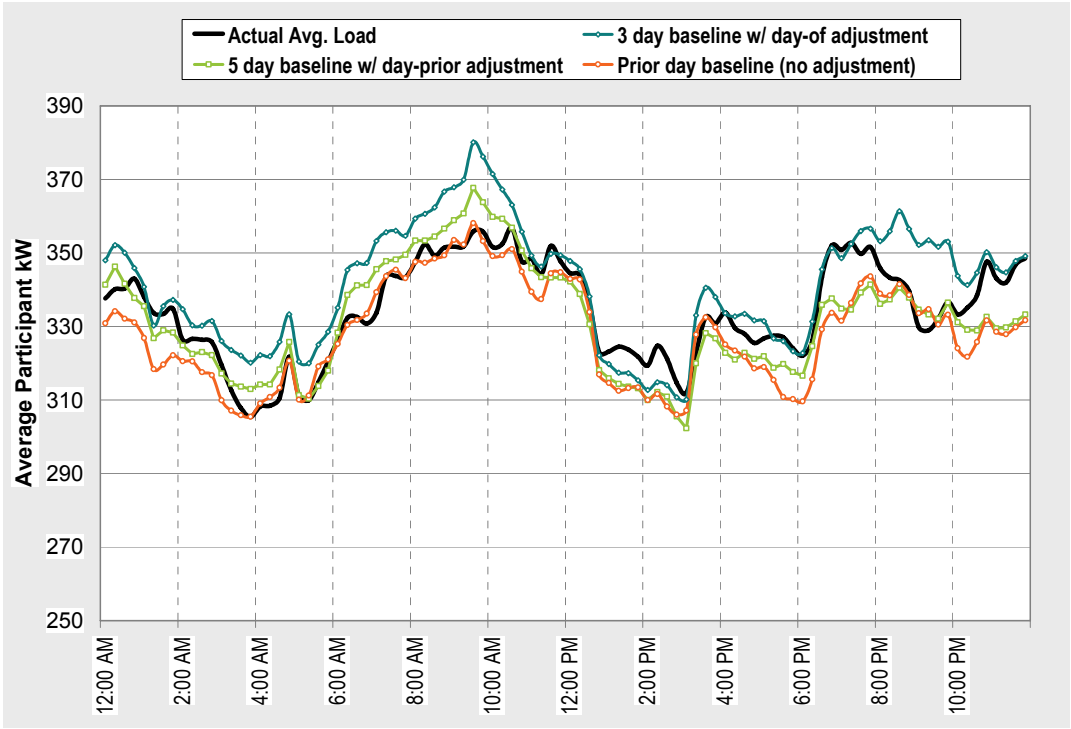
Overall, all three day-matching methods have high predictive capability and are relatively accurate for the median customer. However, all three day-matching methods underestimate the actual load during the peak period. The three day average with same day adjustment is the most accurate day-matching method and has the highest predictive capability. Although it overestimates the reference value loads in the off-peak periods and underestimated the reference value loads for the peak period hours, the 3-day adjusted average has the least bias for the day

and for the peak period. In addition, it has the lowest Coefficient of Alienation and Theil's U, though by a small amount, reflecting that it has relatively strong predictive capability across customers. Because the direction and the magnitude of the bias is known for each hour, the information can (and should) be used to adjust the load impact estimates.

Care must be taken in calculating and interpreting the accuracy statistics for three reasons. First, the average of the ratios is not the same as the ratio of the averages. Both the relative average and relative mean errors are ratios (percentages), and both are calculated as the ratio of the averages (or medians). If the average of the individual customer percent errors is estimated instead, it will yield different, incorrect, and volatile results. Second, it is possible for a reference level with a low median error to have a higher average error. For this reason, the protocols state that both the average and median errors must be included. Third, the percentage represents the amount by which the total actual load is over or under-estimated, not how much the demand response impact is over or underestimated. For example, suppose a reference level overestimates the total load by 5%, and that the actual demand response achieved is a 7% load reduction. An impact analysis using this reference level will report a load reduction of 12%, which is an error of 71% percent relative to the actual load reduction of 7%.

Figure 4-4 provides a visual comparison of the day-matching methods versus the actual loads.

Figure 4-4. Day-matching Methods Example: Comparison of Actual and Predicted Load of Proxy Days



Providing the 24 hour load impacts for an event day is straightforward using a day-matching approach. The reference load is provided by the chosen day-matching method. According to Protocol 8, both the average reduction per customer and the total load reduction for the resource

should be reported. For simplicity, only the average reduction per customer for the average event day is presented here. Table 4-6 shows the average hourly load reduction from the 50 sampled voluntary CPP customers in SDG&E's service territory across the 2005 event days. The reference level used for this example is the 3 day average adjusted by the period preceding the peak period. Note that the table in this example has not corrected for the known hourly biases reflected in Table 4-5.

In the example in Table 4-6, the calculation of the uncertainty estimates reflects the error due to sampling. Sampling error may occur due to variation among participants or variation between days or both. Importantly, the standard errors employed to calculate the uncertainty adjusted load impacts took into account clustering (the fact that we are drawing several observations for each customer) and the size of the participant population (finite population correction). Another approach from footnote 38, uses the variance and standard errors estimated between the estimated and actual load on the event-like days used in protocol 9.[39]

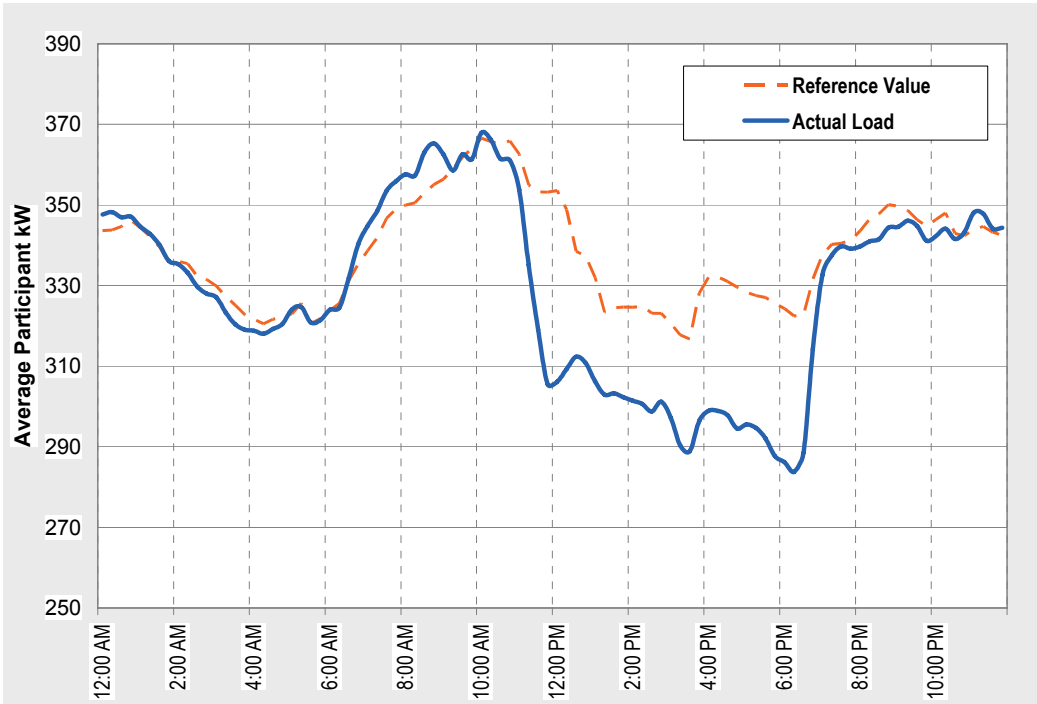
[39] In this second approach, these standard errors come from the selected proxy days rather than from actual event days, as a result the standard errors from the proxy day analysis in protocol 9 are used as the best information on the likely standard errors for the event days. Actual standard errors for the event days can not be calculated as the true reference loads for those days are never known.

Table 4-6. Day-matching Method Example – Uncertainty Adjusted Estimates

Hour Ending	Reference Load	Observed Load	Load Impact	Temp (F)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
1	344.56	347.48	-2.92	67.4	3.27	-0.32	-2.86	-5.33	-9.02
2	340.54	340.92	-0.38	67.6	5.63	2.07	-0.40	-2.78	-6.28
3	333.80	331.56	2.24	66.6	8.25	4.57	2.20	-0.22	-3.83
4	326.29	322.46	3.83	65.0	9.81	6.36	3.90	1.43	-2.10
5	321.52	319.14	2.38	65.2	8.26	4.75	2.38	-0.09	-3.63
6	322.73	322.75	-0.03	65.0	6.25	2.63	0.00	-2.49	-6.21
7	328.98	330.18	-1.20	67.0	5.64	1.72	-1.13	-3.91	-7.93
8	344.03	350.69	-6.66	74.2	0.31	-3.84	-6.63	-9.48	-13.45
9	352.14	360.82	-8.69	80.0	-1.49	-5.74	-8.70	-11.65	-15.88
10	360.34	361.24	-0.91	84.4	6.19	2.00	-0.91	-3.90	-8.19
11	365.83	364.15	1.68	88.0	8.99	4.74	1.66	-1.24	-5.54
12	355.99	328.38	27.61	89.2	35.11	30.73	27.59	24.52	19.98
13	344.54	309.65	34.89	89.4	42.44	37.97	34.87	31.79	27.32
14	326.14	303.68	22.45	90.0	29.83	25.48	22.42	19.38	14.82
15	323.93	300.49	23.43	88.0	31.09	26.55	23.47	20.42	16.07
16	320.78	293.18	27.60	87.2	35.24	30.69	27.65	24.61	20.07
17	331.22	297.54	33.68	85.4	41.05	36.57	33.53	30.54	26.19
18	327.11	292.49	34.62	83.4	41.95	37.54	34.57	31.55	27.22
19	325.21	293.14	32.08	78.4	39.33	34.96	32.17	29.19	25.02
20	339.87	337.24	2.63	74.6	9.63	5.46	2.61	-0.32	-4.47
21	346.97	341.67	5.31	73.4	12.09	8.13	5.33	2.55	-1.31
22	347.36	344.13	3.23	73.0	9.95	6.04	3.35	0.60	-3.48
23	344.92	342.80	2.12	70.6	8.62	4.74	2.13	-0.56	-4.40
24	343.55	346.10	-2.55	69.2	3.94	0.14	-2.51	-5.18	-8.95
	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
Daily	8,118.35	7,881.89	236.46	122.0	399.88	305.11	237.62	166.60	70.03

Figure 4-5 below reflects the observed loads, the reference values, and the load impact for the ex-post average event day from the example.

Figure 4-5. Day-matching Method Example: Observed versus Reference Loads on the Average Event day



4.2.2. Regression Methodologies

Regression analysis is another commonly used method for estimating the impact of DR resources. 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 use[40] 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 4-3.

$$E_i = a + bT_i + c(T_i)(D_i) + e \qquad (4\text{-}3)$$

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

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

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.

When the primary interest is ex post impact evaluation, properly specified regression models and day-matching methods often produce similar results.[41] However, for ex ante estimation of DR resource impacts, regression models are not only recommended, they may be the only feasible approach in most situations.

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.

Overview of Regression Analysis

A 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*.[42] 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 Framework 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

[41] The reader is referred to the KEMA/CEC (2003) report for a useful comparison of the relative accuracy and other attributes of a variety of regression models and day-matching methods.
[42] TecMarket Works. *The California Evaluation Framework*, June 2004. pp. 105 – 120.

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 the load impacts
- Imprecise estimates of the level of confidence that can be placed on the results
- 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 4-7 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 10 are intended to reveal the extent to which many of these issues have been addressed.

Table 4-7. 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 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 data 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. 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 individuals 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.

Because random effects models can provide biased parameter estimates when the error terms are correlated with the explanatory variables, it important to always start with the more robust fixed effects model and subsequently test whether the resulting coefficients and standard errors are the same. This is typically accomplished via a Hausman test. Interpreting the results of such a test, however, requires the evaluator's judgment. Due to the power of time-series cross sectional load data (which has more time observations than most panel data) and the sensitivity of the Hausman test, even trivial differences in results can be statistically significant when in fact the differences between the two models is virtually nil. As a result, the magnitude of difference in results may be more important that statistically significant results, i.e., is the magnitude of the difference meaningful.

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 present[43]. Because of this, the more labor 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't be explained, it will be 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

[43] Page 274-276 of J. Woolridge's textbook, **Econometric Analysis of Cross-section and Panel Data** provides an excellent discussion on serial correlation and the robust variance matrix estimator.

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.[44] 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.

The Advantages 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 enrollment and one after. If there is no pretreatment 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, 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 Section 8, with repeated measures, you may be able to use 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.[45] 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

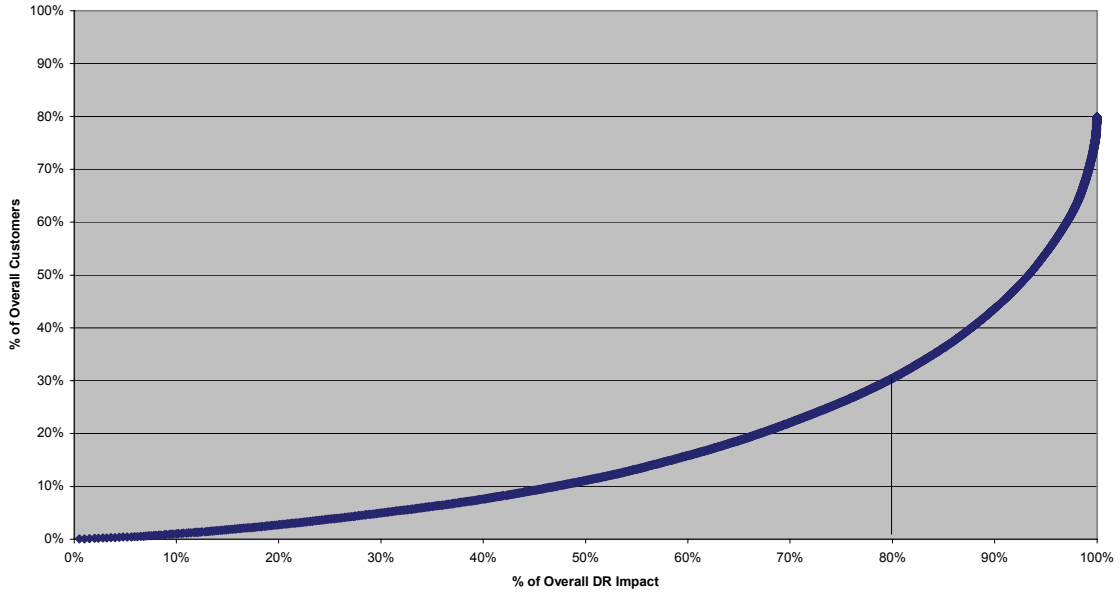
[44] In this instance, separate output tables should be reported for each market segment.

[45] 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.

your primary interest is in estimating impacts for a set of volunteers behaviorally similar to those who have participated to date.[46]

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 4-1, 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 California's Statewide Pricing Project[47] data produced the distribution of demand response impacts shown in Figure 4-6, indicating that roughly 80 percent of total demand response was provided by roughly 30 percent of participants.

Figure 4-6. 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.[48] Regressions

[46] 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. A more in depth discussion of control groups is contained in Section 5.2.

[47] Charles River Associates. "*Impact Evaluation of the California Statewide Pricing Pilot,*" Final Report. March 16, 2005, p. 66. See CEC Website: http://www.energy.ca.gov/demandresponse/documents/index.html#group2

[48] Peter Kennedy. *A Guide to Econometrics*, Fifth Edition, MIT Press, 2003. This book 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. Kennedy (2003) is also a good general reference for the regression methods and issues discussed in this chapter.

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, know as time effects.

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). This type of specification would allow you to develop ex ante estimates for specific combinations of event conditions that did not occur in the past. This ability could be quite useful for operational purposes or for longer term resource planning or resource design.

Estimating Impacts for Hours Outside of the Event Period

As indicated in Protocol 4, 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. The example in Section 4.2.2.10, equation 4-4, illustrates this type of model specification.

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 4-3 and the example shown in Section 4.2.2.10.

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 (as also illustrated in the example in Section 4.2.2.10).

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*[49] provides an example of this type of specification.

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. These topics are discussed in more detail in Section 6. 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 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.

Geographic Specificity

Knowing how impacts vary across regions can be very useful for transmission and distribution planning and for operational dispatch decisions by the CAISO, who must balance supply and demand at thousands of points on the grid and who will soon be using locational pricing to help clear markets at numerous transmission nodes. The specific locations for which impacts may be needed in the future are still unclear, and they will vary across utilities and resources. As previously discussed in Section 3, 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

[49] Charles River Associates. Op. Cit. 2005. CEC web http://www.energy.ca.gov/demandresponse/documents/index.html#group2

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.

Implementing this approach will be easier and less costly if there is prior knowledge regarding which independent variables drive demand response and if data already exists concerning how relevant variables differ across the regions of interest. California's Residential Appliance Saturation Surveys (RASS) and Commercial End Use Surveys (CEUS) provide a rich database that can be used to inform sample designs and modeling exercises. There is also a growing body of evidence concerning what customer characteristics drive demand response for many resource options. As such, there is a greater probability that sufficient prior knowledge exists in California than in many other locations so that a model based approach to location specific impact estimates is likely to be less costly than would be developing large enough samples at each location of interest to estimate impacts of comparable validity and precision.

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

Regression Analysis: An Example

As indicated in Section 4.1, protocols 4 through 7 require that uncertainty-adjusted impact estimates be developed for event-based resources for each hour of an event day. Impacts are to be reported for various day types using a format shown in Table 4-1. In this section, we provide a simple example of how those protocols can be met using a regression-based methodology.

This example was developed using residential customer data for the CPP rate from California's Statewide Pricing Pilot for the summer of 2004. Only data from climate zone 3 (the hot climate zone representing California's central valley) was used. This analysis was completed using STATA, a common statistical package. It should be noted that we did not spend a significant amount of time refining the model specification, although this should be a key area of attention

for regression-based evaluations. Our focus here is on demonstrating how to use regression techniques to meet the protocol requirements.

The estimated regression model has the following form:

$$E_{iJ} = \beta_0 + \beta_1 CPPday + \sum_{j=2}^{24} \left(\beta_{3j} hour_j \times CPPday\right) + \beta_4 (CPP_j \times CDH_{ij}) + \beta_5 (CAC_i \times CDH_{ij}) +$$

$$\beta_6 (CDHrunup_{ij}) + \beta_7 (CDH24hrlag_{ij}) + \sum_{j=2}^{24} \beta_{8j} hour_j \qquad (4\text{-}4)$$

Where:

CPPday = 1 on an event day, 0 otherwise

CPP = 1 during the event period on event days, 0 otherwise

$Hour_i$ = 1 for hour i, 0 otherwise

CAC = 1 for customers with central air conditioning, 0 otherwise

CDH_i = Cooling degree hours to base 75° F in hour i

$CDHrunup_i$ = cumulative cooling degree hours in the day prior to hour i

$CDH24lag_i$ = cooling degree hours in hour i the day before the event day

The hourly binary variables capture the non-weather dependent load shape on non-critical days whereas the hourly variables interacted with the CPP day binary variable estimate the difference in the load in each hour on CPP days relative to non-critical days. The interaction between the CPP event binary variable and the cooling degree hour variable allows one to estimate the change in the resource impact as cooling degree hours change. In order to estimate impacts on the day preceding or following an event day, binary variables representing these days interacted with the hourly binary variables could be included in the specification. For simplicity and ease of interpretation, we did not include these variables in the example.

Figure 4-7 contains the regression output for the model. As seen, the cooling degree hours variable has a strong positive relationship when interacted with central air conditioning, indicating that energy use increases with cooling degree hours for households with air conditioning. The negative sign on the interaction term between degree hours and the CPP variable indicates that energy use drops more during the event hours when the day is hotter than when it is cooler. This is logical as there is more load to drop on hotter days due to air conditioning use. The positive sign on the interaction term between the hour of the day and the CPP day binary variable for the hours immediately preceding and following the event period indicates a small amount of pre-cooling and a significant snapback effect. Tests of joint significance applied to the results from the event hours and the surrounding hours indicate that the CPP impacts are statistically significant and in the expected direction across the event period hours (2-7 pm), pre-event hours (12-2 pm), and post-event hours (7-9 pm).

Figure 4-7. Regression Output

kWh	Coef.	Std. Err.	t	P>t
CPPday	-0.0097	0.0149	-0.65	0.516
CPPxCDH	-0.0070	0.0013	-5.31	0
CACxCDH	0.0329	0.0006	53.95	0
CDHrunup	0.0017	0.0000	38.64	0
CDH24hrlag	0.0138	0.0006	24.15	0
Hourly weather independent CPP Impacts				
(1:00 am is used as the reference value)				
2:00	-0.0227	0.0150	-1.51	0.1310
3:00	-0.0549	0.0191	-2.88	0.0040
4:00	-0.0644	0.0212	-3.04	0.0020
5:00	-0.0466	0.0225	-2.07	0.0380
6:00	-0.0374	0.0233	-1.61	0.1080
7:00	-0.0240	0.0238	-1.01	0.3150
8:00	-0.0228	0.0242	-0.94	0.3450
9:00	-0.0068	0.0244	-0.28	0.7800
10:00	-0.0311	0.0245	-1.27	0.2050
11:00	-0.0128	0.0247	-0.52	0.6040
12:00	-0.0136	0.0248	-0.55	0.5820
13:00 (pre event hour)	0.0400	0.0249	1.61	0.1080
14:00 (pre event hour)	0.0496	0.0250	1.99	0.0470
15:00	-0.0643	0.0323	-1.99	0.0460
16:00	-0.0334	0.0330	-1.01	0.3110
17:00	-0.0043	0.0330	-0.13	0.8970
18:00	0.0304	0.0321	0.95	0.3430
19:00	0.0479	0.0303	1.58	0.1140
20:00 (snapback period)	0.2314	0.0242	9.54	0.0000
21:00 (snapback period)	0.3037	0.0237	12.8	0.0000
22:00 (snapback period)	0.2595	0.0229	11.35	0.0000
23:00 (snapback period)	0.1953	0.0213	9.18	0.0000
0:00	0.1071	0.0181	5.92	0.0000
Weather independent load profile				
(1:00 am is used as the reference value)				
2:00	-0.1119	0.0044	-25.55	0.0000
3:00	-0.1850	0.0057	-32.49	0.0000
4:00	-0.2152	0.0065	-33.36	0.0000
5:00	-0.2232	0.0069	-32.23	0.0000
6:00	-0.1836	0.0072	-25.36	0.0000
7:00	-0.0949	0.0074	-12.74	0.0000
8:00	-0.0231	0.0076	-3.04	0.0020
9:00	0.0187	0.0077	2.44	0.0150
10:00	0.0579	0.0077	7.48	0.0000
11:00	0.0952	0.0078	12.18	0.0000
12:00	0.1063	0.0079	13.37	0.0000
13:00	0.1243	0.0082	15.21	0.0000
14:00	0.1500	0.0085	17.70	0.0000
15:00	0.1056	0.0087	12.08	0.0000
16:00	0.1744	0.0089	19.55	0.0000
17:00	0.2763	0.0089	31.06	0.0000
18:00	0.3675	0.0087	42.45	0.0000
19:00	0.4090	0.0082	49.74	0.0000
20:00	0.5101	0.0076	66.72	0.0000
21:00	0.5976	0.0071	84.48	0.0000
22:00	0.5572	0.0065	85.47	0.0000
23:00	0.3662	0.0058	63.58	0.0000
0:00	0.1668	0.0045	37.33	0.0000
Constant	0.7133	0.0019	371.94	0.000

FE (within) regression with AR(1) disturbance	Number of obs	=	705421
Group variable (i): custidnum	Number of groups	=	203
R-sq: within = 0.0549	Obs per group: min	=	215
between = 0.0762	avg	=	3475
overall = 0.1497	max	=	3647
	F(33,705185)	=	1241
corr(u_i, Xb) = 0.0249	Prob > F	=	0

Figure 4-8 shows how the predicted values compare with actual values on the average critical event day in 2004. As seen in Figure 4-8, the model does a good job of tracking actual energy use on event days, including the substantial snapback effect that occurs following the end of the event period. The estimated impacts equal the difference in the two lines in Figure 4-8 labeled "predicted energy use without DR" and "predicted energy use with DR." The figure also illustrates a significant drop in load impacts in the last two hours of the event period. The impact estimates illustrated in Figure 4-8 are shown in Table 4-8, which is in the format required by Protocol 6 for the average event day.

Figure 4-8. Statewide Pricing Pilot 2004 Load Impacts

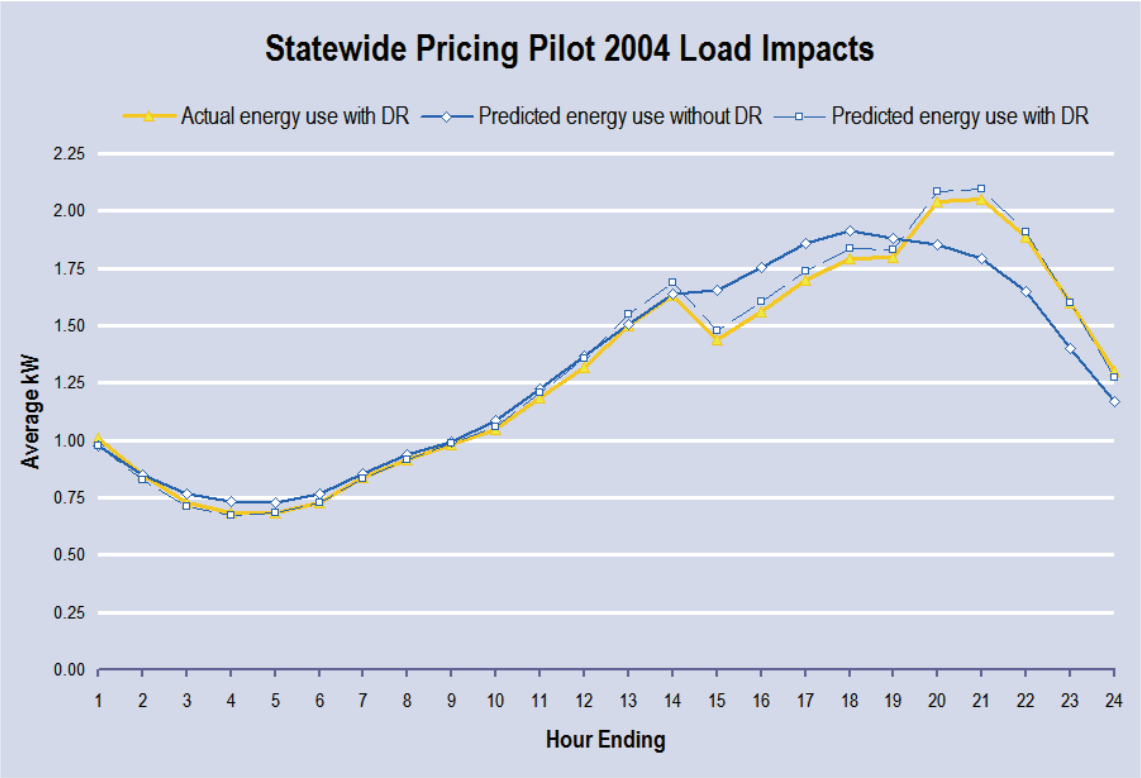


Table 4-8. Day Type: Average Event Day for 2004 SPP Residential – Climate Zone 3

		Per participant load impacts					
			Percentiles				
Hour Ending	Temp (F)	Mean (kW)	10%	30%[NC]	50%	70%[NC]	90%
1	71.2	-0.009	0.011		-0.010		-0.030
2	70.0	-0.033	-0.008		-0.032		-0.057
3	68.8	-0.064	-0.039		-0.064		-0.090
4	67.8	-0.074	-0.041		-0.074		-0.101
5	66.9	-0.057	-0.030		-0.057		-0.084
6	66.1	-0.047	-0.019		-0.047		-0.074
7	65.9	-0.034	-0.007		-0.034		-0.061
8	67.2	-0.033	-0.005		-0.032		-0.060
9	70.1	-0.017	0.011		-0.017		-0.044
10	74.4	-0.041	-0.013		-0.041		-0.068
11	78.7	-0.022	0.005		-0.022		-0.049
12	82.9	-0.023	0.004		-0.023		-0.051
13	86.4	0.030	0.058		0.030		0.002
14	89.1	0.040	0.067		0.040		0.013
15	90.8	-0.185	-0.158		-0.185		-0.212
16	91.7	-0.160	-0.132		-0.160		-0.188
17	91.6	-0.131	-0.104		-0.131		-0.159
18	90.5	-0.090	-0.062		-0.090		-0.117
19	88.2	-0.057	-0.030		-0.057		-0.085
20	84.5	0.222	0.249		0.222		0.195
21	80.2	0.294	0.320		0.294		0.267
22	76.7	0.250	0.275		0.250		0.225
23	74.3	0.186	0.210		0.186		0.162
24	72.6	0.097	0.118		0.097		0.077

NC: Not calculated.

The uncertainty adjusted load impacts shown in the right-hand columns in Table 4-8 can be generated in two ways.

One approach involves using the regression model to compute the difference in the mean predicted load with and without the DR incentive in effect, and using the standard errors of the predictions to estimate the uncertainty surrounding that difference, i.e., the confidence intervals. The exact equation to use for this calculation will vary depending on whether or not the variances of the predictions are equal and the size of the sample (small samples require adjustments). Assuming the variance of the estimates is equal and the sample size is sufficiently large, the load impact (difference of mean predictions) and the standard error of the difference are given by the following formulas:

$$\text{AVERAGE LOAD IMPACT} = \overline{X}_0 - \overline{X}_1$$

Where $\overline{X}_0$ = Energy use with DR and
$\overline{X}_1$ = Energy use without DR

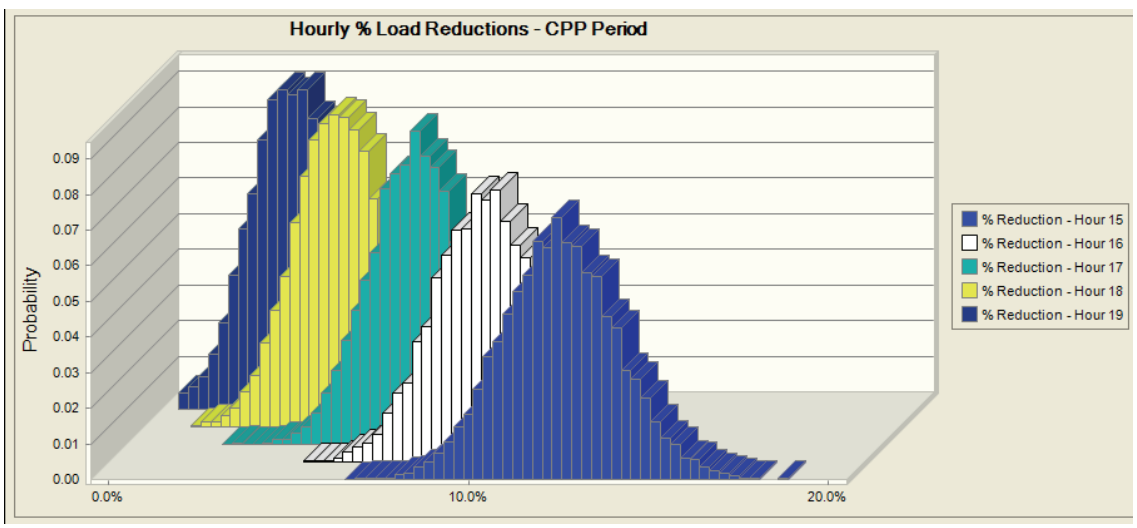
STANDARD ERROR OF THE DIFFERENCE

$$s_{\overline{x}_1 - \overline{x}_0} = \frac{s_0 + s_1}{2}$$

Where s_0 = Standard error of predicted energy use with DR, and
s_1 = Standard error of predicted energy use without DR

An alternative approach to estimating the uncertainty adjusted impacts is Monte Carlo simulation. Monte Carlo simulation will produce the same results provided the simulation includes enough draws (trials). Monte Carlo simulations employ a transparent, brute force approach in which random draws are made from the probability distributions of factors that affect the outcome. The uncertainty adjusted impact estimates in Table 4-8 were produced using Monte Carlo simulation. Figure 4-9 shows how the distribution of DR impacts (in percentage terms) looks in each hour based on the model shown in Figure 4-7.

Figure 4-9. Distribution of DR Impacts by Hour



A significant advantage of Monte Carlo simulation is that it allows you not only to incorporate uncertainty due to forecast error, but to also incorporate uncertainty in the explanatory variables. Given information on the distribution in temperature or degree hours around some mean value in each hour, for example, one can use simulation to produce a probability distribution of impacts that reflects the uncertainty in both weather and the predictive power of the underlying model

The steps outlined below illustrate how to produce probability distributions that reflect both modeling error and uncertainty in the distribution of key explanatory variables such as weather:

1. Predict the load with the DR resource in effect by temperature level and hour of day, with a 90% confidence interval for each prediction.

2. Predict the load without the DR resource in effect by temperature level and hour of day, with the 90% confidence interval for each prediction.

3. Determine the distribution for the weather variables for each hour of the day by day type based the historical data from the same type of day, including a) the mean b) standard deviation c) autocorrelation. A better approach is to fit and compare a range of distributions to the actual weather data, by hour, for that day type and include the hourly auto-correlation of weather. This is preferred because distributions may be skewed, have long tails, or be bimodal, i.e., they may be non-normal.

4. Run a simulation allowing the temperatures to vary (taking into account the actual correlations). The appropriate distribution for the load with and without the DR resource would be created by drawing the mean and standard error of the hour and temperature lookup tables. The draw from these distributions of load with and without the DR resource in effect would in turn be used to calculate the load impact for each.

5. Extract the resulting summary statistics from the resulting distributions of the Monte Carlo simulation.

4.2.3. Other Methodologies

Sections 4.2.1 and 4.2.2 provided an overview and discussion of issues associated with the two primary methods of estimating ex post impacts for event-based DR resources, day-matching and regression analysis. Regression analysis is a less transparent but more robust and flexible tool than day-matching. It is the recommended default option whenever ex ante estimation is required unless other considerations, such as erratic consumer behavior, lack of variability, data limitations, budget constraints, or the limited importance of a resource due to its small size, suggest that an alternative approach is preferred. This section covers some of the additional options that might be considered if one or more of these conditions is present.

Sub-Metering

One approach already mentioned is 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).

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).

Engineering Analysis

Another method that might be useful in limited situations is engineering analysis. As discussed previously, engineering analysis is much less useful for estimating the impacts associated with most DR resources because impacts are driven much more by consumer behavior than by technology implementation. Even some technology enabled DR resources, 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 the CAISO or utility 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 resource option 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.

Duty Cycle Analysis

Another approach 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*.[50] 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 the A/C unit, and the share of time the A/C unit 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 load utilized by hour of day and temperature bin (for non-event days).

The specific load impacts are then calculated by:

1. Identifying the average share of connected load utilized during the appropriate temperature and time bins (average duty cycle), and

[50] Quantum Consulting Inc. The Air Conditioner Cycling Summer Discount Program Evaluation Study. January 2006.

2. Calculating resource 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.

Operational Experimentation

Still another approach to impact estimation for event-based resources involves the use of what might be called operational experimentation. By operational experimentation, we mean the selective exercise of a resource on a sub-sample of participants with 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 resource.

Given the typically large number of customers participating in load control resources, there are plenty of customers from which a small sample can be drawn for experimental purposes. One could split this sample into two groups, again 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 California), 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 resource options or even for long-standing emergency resources that are not triggered very frequently. In these situations, there may not be sufficient data on event days to estimate impacts using other methods.

4.2.4. Measurement and Verification Activities

Measurement and verification (M&V) refers to data collection, monitoring and analysis activities associated with the calculation of gross energy and peak demand savings from individual customer sites or projects.[51] M&V activities typically focus on measure installation verification, installation quality, manufacturing defects, measure use and operation, equipment maintenance procedures, and in-situ measure efficiency. Such activities can be essential to process evaluations of EE or DR resources and in helping to identify ways to improve the DR resource offering. Importantly, M&V activities can help determine why estimated impacts might differ

[51] The definition of M&V used here differs from how the term is sometimes used elsewhere. In some instances, M&V is defined much more broadly and essentially is synonymous with impact estimation. It is important to keep the narrower definition in mind when reviewing this section and when encountering the term elsewhere in this document.

from goals or expectations. Put another way, M&V activities aren't often needed to understand what impacts are, but they can be useful for explaining why they are what they are.

There is an extensive literature on M&V protocols and activities in support of EE evaluation. If M&V activities are needed for DR impact estimation, evaluators can turn to the following documents to learn more about standard procedures:

- *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, TecMarket Works. Prepared for the California Public Utilities Commission, April 2006, pp. 49 - 64.

- *The California Evaluation Framework*, TecMarket Works. Prepared for the California Public Utilities Commission, June 2004.

- *The International Performance Measurement and Verification Protocol*, *Volume 1: Concepts and Options for Determining Savings*, 2002.

As previously discussed, there is probably limited need for M&V activities in support of impact estimation for most DR resource options. The prior example discussed in Section 4.2.3.2 where data on motor horsepower is used to develop impact estimates is one case where M&V activities would come into play in support of impact estimation. For technology-based DR resource options, such as load control, PCT programs, AutoDR and the like, M&V activities could be useful in helping to understand why impacts are what they are. For example, if impacts for an air conditioning load control program are not as large as expected, M&V activities could be employed to inspect load control switches to see if faulty installation, equipment deterioration or tampering might explain the result.

Select M&V methods may be appropriate if full evaluations are only expected to occur periodically and the evaluation plan called for some a sample of sites where some measurements might take place to provide validation of impacts during the period between evaluations. The benefits and costs of any M&V between evaluations would be a component of Protocol 3 in Evaluation Planning.

5. EX POST EVALUATION FOR NON-EVENT BASED RESOURCES

This section contains protocols and guidelines for ex post evaluation of non-event based, DR resource options. As delineated in Section 2, non-event based resources fall into three broad categories:

- **Non-event based pricing**—This resource category includes TOU, RTP and related pricing variants that are not based on a called event—that is, they are in place for a season or a year.

- **Scheduled DR**—There are some loads that can be scheduled to be reduced at a regular time period. For example, a group of irrigation customers could be divided into five segments, with each segment agreeing to not irrigate/pump on a different selected weekday.

- **Permanent load reductions and load shifting**—Permanent load reductions are often associated with energy efficiency activities, but there are some technologies such as demand controllers that can result in permanent load reductions or load shifting. Examples of load shifting technologies include ice storage air conditioning, timers and energy management systems.

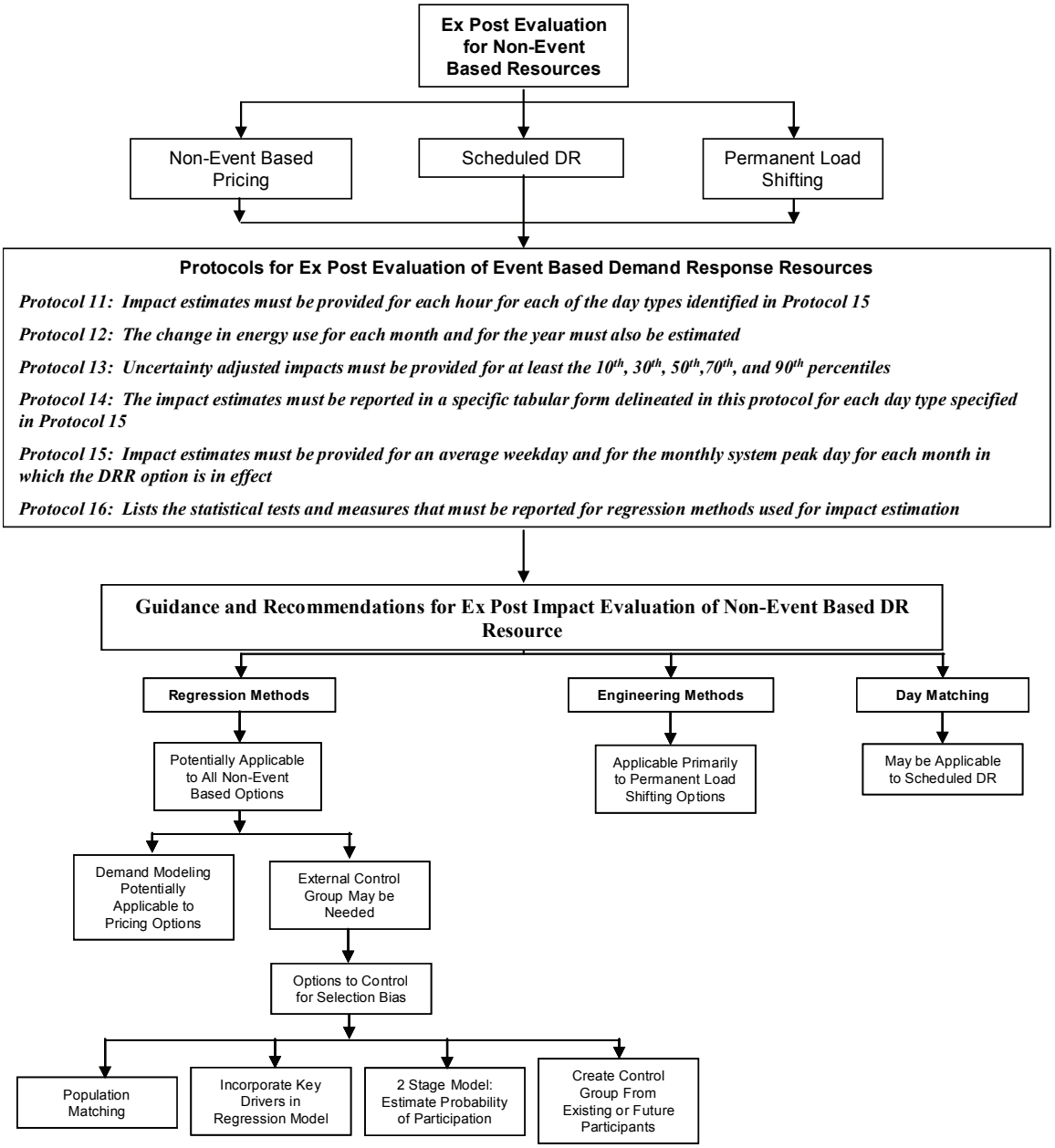
The protocols for non-event based resource options are similar to those for event-based resources—the primary difference being in the relevant day types for which impacts must be reported. Figure 5-1 summarizes the protocols that apply to non-event based resources.

Figure 5-1 also summarizes the primary guidance and recommendations discussed in the remainder of this section. Regression analysis is potentially applicable to all three primary categories of non-event based resource options. It is perhaps the only suitable method for non-event based pricing options where impacts are strongly influenced by consumer behavior. With the possible exception of scheduled DR, day-matching methods are not suitable because the influence of the demand response resource is in effect every day of the week (although it may vary across days for some resource options such as RTP). Consequently, it is not possible to estimate a reference usage level using prior days in which the DR resource is not in effect. Engineering methods may be useful for some permanent load shifting options, such as ice storage.

The guidance and recommendations concerning regression methods contained in Section 4.2.2 apply here as well. However, an important additional issue must be addressed when applying regression analysis to impact estimation for non-event based options, which arises from the fact that, in most instances, it is not possible to use participants as their own control group. Given this, for non-event based resources, it is often necessary to use an external control group, which raises the possibility that selection bias could lead to erroneous impact estimates. The guidance section discusses ways of avoiding this potential problem. Another topic that is particularly relevant to non-event based pricing options is demand modeling, which can be used to estimate

impacts in situations where there is sufficient price variation to allow for estimation of price elasticities of demand.

Figure 5-1. Section Overview



5.1. Protocols for Non-Event Based Resources (Protocols 11-16)

There are six protocols that apply to non-event based resources. As with event based resources, collectively, these protocols define the time periods, day types, measures of uncertainty, output formats and ancillary data that are to be reported when presenting impact estimates for non-event

based resources. The day types differ for non-event based resources. The statistical measures protocol associated with day-matching methods is typically not relevant for non-event based resource options because day-matching is only applicable in rare cases (e.g., scheduled DR).

Protocol 11:

The mean change in energy use per hour (kWh/hr) for each hour of the day shall be estimated for each day type and level of aggregation defined in Protocol 15. The mean change in energy use for the day shall also be reported for each day type.

Protocol 12:

The mean change in energy use per month and per year shall be reported for the average across all participants and for the sum of all participants in a DR resource option in each year over which the evaluation is conducted.

Protocol 13:

Estimates of the 10th,30th, 50th, 70th, and 90th percentiles of the change in energy use in each hour, day and year, as described in Protocols 11 and 12, for each day-type and level of aggregation described in Protocol 15, shall to be provided.

Protocol 14:

Impact estimates shall be reported in the format depicted in Table 4-1 for all required day types, as delineated in Protocol 15.

Protocol 15:

The information shown in Table 4-1 shall be provided for each of the following day types for the average across all participants sum of all participants:

- ***For the average weekday for each month in which the DR resource is in effect***[52]
- ***For the monthly system peak day for each month in which the DR resource is in effect.***

Day type definitions and additional reporting requirements for each day type are summarized below:

Average Week Day for Each Month: The average across all weekdays in each month during which the DR resource is in effect. In addition to the

[52] If a resource is seasonal, only the months in which the resource is in effect needs to be reported.

information contained in Table 4-1, the following information shall be provided:

- ***Average temperature[53] for each hour for a typical week day for each month.***
- ***Average degree hours for the typical week day for each month.***
- ***Average number of customers participating in the DR resource option each month***

Monthly System Peak Day for Each Month: The day with the highest system load in each month. In addition to reporting all of the information shown in Table 4-1, the following information shall be provided:

- ***Temperature for each hour on the system peak day for each month***
- ***Average degree hours on the system peak day for each month.***
- ***Average number of customers participating in the DR resource option on the system peak day for each month.***

Protocol 16:

For regression based methods, the following statistics and information shall be reported:

- ***Adjusted R-squared or, if R-squared is not provided for the estimation procedure, the log-likelihood of the model***
- ***Total observations, number of cross-sectional units and number of time periods***
- ***Coefficients for each of the parameters of the model***
- ***Standard errors for each of the parameter estimates***
- ***The variance-covariance matrix for the parameters***

[53] As noted in Section 4, when reporting temperatures and degree days, it is intended that the temperature be reasonably representative of the population of participants associated with the impact estimates. If participation in a resource option is concentrated in a very hot climate zone, for example, reporting population-weighted average temperature across an entire utility service territory may not be very useful if a substantial number of customers are located in cooler climate zones. Some sort of customer or load-weighted average temperature across weather stations close to participant locations would be much more accurate and useful.

- ***The tests conducted and the specific corrections conducted, if any, to ensure robust standard errors***

- ***How the evaluation assessed the accuracy and stability of the coefficient(s) that represent the load impact.***

5.2. Guidance and Recommendations

As discussed in the introduction to this section, regression methods are most applicable to non-event based pricing options in part because demand response for these options is strongly influenced by consumer behavior which is best captured using statistical analysis. Regression analysis could be used to estimate impacts for scheduled DR and permanent load shifting options as well. Day-matching is not a suitable estimation method for pricing options because there are no prior days to use for estimating reference values. Day-matching may have some limited applicability for estimating impacts for scheduled DR while engineering analysis may be suitable for permanent load shifting where technologies such as ice storage may be used.

5.2.1. Regression Analysis

As with event-based resource options, regression analysis is the preferred method for estimating load impacts in most instances. The guidance and recommendations pertaining to the regression analysis contained in Section 4.2.2 are applicable for non-event based resources as well. However, the advantages associated with repeated measures discussed in Section 4.2.2.2, in particular the option of using participants as their own control group, do not apply with non-event based resource options. While it may be possible to use participants as their own control group if sufficient pre-treatment data exists, it is more likely that an external control group will be needed in order to estimate impacts. When this occurs, selection bias is a key issue that must be addressed. As such, Section 5.2.1.1 provides guidance concerning methods for addressing this critical issue. Section 5.2.1.2 discusses demand modeling, a particular type of regression analysis that may be applicable when developing impact estimates for non-event based pricing options.

External Control Groups and Selection Bias

The primary goal of impact estimation is to develop an unbiased estimate of the change in energy use resulting from a DR resource. Impacts can be estimated by comparing energy use before and after participation in a DR resource option, energy use between participants and non-participants, or both. The primary challenge in impact estimation is ensuring that any observed difference in energy use across time or across consumer groups is attributable to the DR resource, not to some other factor—that is, determining a causal relationship between the resource and the estimated impact.

One way of ensuring that a causal relationship can be established is through random assignment of treatment and control customers within the context of a controlled experiment. Random assignment helps ensure that any estimated difference in the variable of interest is due to the treatment, not due to any preexisting differences between the treatment and control populations.

If participants in an experiment are allowed to self-select into the treatment group, any observed difference between the treatment and control groups could be due to some pre-existing difference between the two groups. A pre-existing difference of this sort will cause selection bias in the estimated impact of the treatment, if measured as the difference between treatment and control customers. Whenever random assignment is not possible, the operating assumption should be that selection bias exists.

Even though ex post estimation of DR resource impacts is rarely if ever done in the context of a controlled experiment, the environment of a controlled experiment can be closely approximated if a control group can be selected from among DR resource participants. This requires that data on the variable of interest, in this context, energy use by time period, be available both before and after the DR resource influence is in effect. However, until AMI meters are more fully deployed, pre-participation data may not be available in most instances, in which case an external control group will be needed.

When using an external control group, it is imperative that the control group either has usage characteristics that are quite similar to those of the participant population or that any preexisting differences can be controlled for. For voluntary DR resource options, there are a number of reasons to believe that those who participate might be different from those who do not.

For example, if a TOU rate is revenue neutral compared with energy use for the average consumer in a rate class, customers who use less energy than the average during the peak period relative to the off peak period will see their bills fall even without changing their usage pattern. These structural benefiters might consist of consumers who either don't have air conditioning or who have it but typically don't use it during peak periods because, for example, no one is home during that time. Structural benefiters may volunteer at a higher rate than those who use more energy during the peak period.

If participation in the DR resource program is driven by the type of selection bias described above, impact estimates based on the difference in loads during the peak period between a control group chosen from the general population of non-participants and a participant group will be comprised of two parts. One part would result from any change in behavior that the participant population makes in response to the time varying rate. However, the second part would result from any preexisting difference in load shapes between the two groups. In the above example, that preexisting difference would lead to an over estimate of resource impacts.

Figure 5-2 helps illustrate how impact estimates can be developed given various scenarios regarding the availability of data for control and participant populations.

Figure 5-2. Impact Estimation Options

	Control Group	Participant Group
Before Participation	C_1	P_1
After Participation	C_2	P_2

In Figure 5-2, P represents the participant population and C represents a control group. Subscripts 1 and 2 represent the time periods before and after a customer decides to participate in a DR resource program or tariff. The ideal situation occurs when usage data is available on participant and control customers for a sufficiently long time period before and after the point at which the group of participants being examined sign up for the resource option—that is, during periods 1 and 2 for both groups. In this situation, the resource impact can be estimated using the following calculation:

$$\text{Impact} = (P_2 - P_1) - (C_2 - C_1) \qquad (5\text{-}1)$$

That is, the impact equals the difference in energy use in the two time periods for the participant group, adjusted for any difference in energy use between the two time periods for the control group. The second term adjusts for differences in energy use due to exogenous factors, such as weather, economic activity, and the like.

Equation 5-1 above can be rewritten as follows:

$$\text{Impact} = (P_2 - C_2) - (P_1 - C_1) \qquad (5\text{-}2)$$

In this form, the equation can be interpreted as estimating impacts based on the difference in energy use for the participant and control group samples during the participation period and the difference between the two groups in the pre-participation period. The second term adjusts for any preexisting differences in load shapes between the participant and non-participant population.

If load data does not exist for customers prior to participation (e.g., there is no data in period 1), impacts could be estimated as $(P_2 - C_2)$, but this estimate will be biased unless $(P_1 - C_1)$ equals 0—that is, unless energy use for the control group is a very good proxy for energy use by the

participant population prior to participation. In the future, when AMI is widely deployed, there will be a high likelihood of having pre-participation load data in most instances. Today, however, that likelihood is quite low. Without pre-participation data, selecting a well-matched control group or otherwise controlling for differences between the control and participant population is essential.

There are a variety of strategies for choosing a good control group or otherwise controlling for relevant differences between the control and participant populations. One approach is to pick a group from the general population that has observable characteristics that match the participant population. For instance, in the above example where it is likely that participants have lower air conditioning saturations than the population as a whole, one could select a control group with the same air conditioning saturation and the same dispersion across climate zones as the participant population. With sufficient survey data on both participants and non-participants, stratification on other characteristics (e.g., pool ownership, size of house, income, etc.) could also be used to decrease the likelihood of any load shape bias influencing the impact estimates.

An alternative to the control matching procedure described above, but one that is conceptually similar, involves incorporating variables representing observable characteristics for the participant and control groups into the impact estimation procedure, and then adjusting the impact estimates to reflect the participant population characteristics. For example, one could estimate a regression model using participant and control group data that would correlate household load (or share of daily energy use) during the peak period with air conditioning ownership. Given this relationship, one could use the saturation of air conditioning for the participant population to produce an unbiased estimate of load and load impact, assuming the difference in air conditioning saturation is the primary determinant of differences in load between the participant and control group, aside from the influence of the DR resource itself.

A third approach to addressing selection bias involves developing a two stage model, where the first stage estimates the probability of participating in a DR resource option, which then becomes a variable that is included in the impact estimation model. A useful discussion of various approaches to modeling and adjusting for self-selection in the context of EE evaluation is contained in the *California Evaluation Framework*.[54]

Still another approach to addressing selection bias is to figure out a way of creating a control group from the existing participant population or from future participants. For example, it might be possible to select a sample of current resource participants and offer them an incentive to become control customers, thus no longer having them respond to the DR resource prices or incentives for some period of time. An alternative would be to focus on future volunteers, asking them to delay their transition onto the resource option so that they can be used as a control. This approach is comparable to randomly assigning volunteers in an experiment to control and treatment groups once they agree to volunteer.

The problem of selection bias discussed above is equally important for both voluntary opt in and opt out DR resource options. Given the typically high degree of customer inertia, especially

[54] pp. 142-145.

among mass market consumers, consumers who opt out of a default tariff or DR program may be as different from those who stay as are those who volunteer to participate on an opt in basis.

For a mandatory DR resource, selection bias is not an issue. Of course, there is also no possibility of selecting a control group from among non-participants. As such, if pre-participation data does not exist, there may be no alternative but to select a sample of consumers and create a control group by removing the influence of the resource for some time period.

The approaches outlined above are focused on ensuring the internal validity of the impact estimates. In this context, internal validity refers to establishing a causal relationship between the DR resource and the change in energy use for the current DR resource participants. Knowing whether the estimated impacts are also valid for potential future participants is typically also of interest. This is known as external validity. Issues associated with external validity will be discussed in Section 6, as it is a key issue for ex ante estimation.

5.2.2. Demand Modeling

For price-driven DR resources, if there is sufficient variation in prices across time or across consumers, it may be possible to estimate an energy demand model and use the model to estimate resource impacts for the day types of interest. A demand model quantifies the relationship between energy demand and price. As prices increase, the amount of energy used decreases and vice versa. Because energy use varies with other factors, such as weather and end use appliance holdings, variables representing these factors are typically also included in the demand equation.

If there is sufficient variation in price to estimate a demand model, the impact of a price-driven DR resource can be estimated by predicting energy demand based on the new tariff and what the price was prior to selecting the new tariff. The following equation represents a simple demand model.

$$\ln(E_i) = \beta_0 + \beta_1 \ln(T_i) + \beta_2 \ln(P_i) + \varepsilon \qquad (5\text{-}3)$$

where E_i = energy use in hour i

T_i = the temperature in hour i

P_i = the price in hour i

ε = the regression error term

β_0 = a constant term

β_1 = the change in load given a change in temperature

β_2 = the change in log of energy use given a change in the ln of price

ln = the natural logarithm.

One can use equation 5-3 to estimate energy demand for two price levels, one representing the resource price during hour i and the other representing the price in that same hour prior to selecting the new tariff. The difference between energy use at these two prices is an estimate of the impact of the DR resource option.

The double-log specification depicted in equation 5-3 is commonly used in empirical estimation of demand models. It is convenient in that the coefficient on the price term, β_2, represents the price elasticity of demand, which equals the percentage change in energy use given a percentage change in price.

Demand modeling works best when there are multiple prices that can be used to estimate the demand function. Real time pricing is an ideal candidate for demand modeling, as prices change hour to hour and day to day. As such, the demand equations can be estimated without using an external control group, thus eliminating the possibility of selection bias due to a mismatch between control and participant populations.

It may also be possible to estimate a demand model for a TOU tariff using only the participant population. This approach will have a higher probability of success if pre-participation load data is available and if there is seasonal fluctuation in prices. However, taking advantage of the seasonal fluctuation in prices would require normalizing for variation in weather, seasonal fluctuations in business operations, and other factors. Any omitted variables or misspecification in this regard could easily bias the price parameters and the resulting impact estimates.

5.2.3. Engineering Analysis

Engineering analysis is another approach that might be suitable for some resource options that are largely technology driven and that have much more limited behavioral variation than do pricing resources, for example. Permanent load shifting options such as ice storage and energy management systems are examples where engineering analysis may be suitable for estimating load impacts.[55]

Engineering methods use basic rules of physics to calculate estimates of energy and demand savings. The technical information required as inputs to engineering models generally come from manufacturers, research studies, and other general references combined with assumed or measured equipment operating characteristics.

In order to estimate savings via engineering methods, one must establish a baseline or reference value from which to compare the energy consumption and demand of facilities included in the evaluation. The baseline may require specification of the equipment or building characteristics and operations prior to participation, as well as an estimate or measurement of pre-participation energy consumption. The baseline may consist of the following:

- For DR programs focused on early equipment replacement (retrofit), the pre-existing and still-functioning equipment replaced as a result of participation defines the baseline. Pre-

[55] The remainder of this discussion consists mainly of selected text from *The California Evaluation Framework*, pp. 120 – 129.

participation energy consumption may need to be adjusted to reflect changes in equipment or building operations that were not a direct result of participation in the DR program.

- For equipment that is being replaced at the end of its useful life (i.e., in all situations where the customer would have been replacing the equipment in the absence of the DR program), standard-efficiency new equipment defines the baseline. The DR program's purpose in these cases is to induce customers to do the replacement with a higher-efficiency alternative than they would have selected in the absence of the program.

Engineering methods can be divided into two basic categories.

- Simple Engineering Model
- Building Energy Simulation Model

Simple engineering models and algorithms are typically straightforward equations for calculating energy and demand impacts of non-weather dependent energy efficiency measures, such as energy efficient lighting, appliances, motors, cooking equipment, etc. Simple engineering models are generally not used for weather dependent measures such as building envelope and HVAC measures; these measures are generally analyzed using building energy simulation models.

Building energy simulation models are computer programs that use mathematical representations of important energy and control processes in an attempt to realistically simulate the thermal and energy systems in a building. Energy calculations are carried out on an hourly or sub-hourly basis for a selected time period or more commonly for an entire year based on typical weather data for the selected building site. The resources are made up of a collection of mathematical models of building components, such as windows, wall sections, and HVAC equipment. The individual component models are linked together to form a complete building simulation. The results predict the performance of the building structure and energy systems under given weather conditions at a selected geographic location.

All building energy simulation models have limitations that must be well-understood before applying the model to a particular energy estimation problem.[56] For example, most resources are limited to the simulation of common HVAC system types with a predetermined system configuration. Considerable latitude is given to the user with respect to describing system performance parameters, but the basic arrangement of the system component is fixed and defined by common practice in the building design and HVAC industries. This does not present a problem for most buildings and systems, but for complex custom HVAC configurations, the judgment and experience of the user is critical.

[56] For more information on building energy simulation models, see *State-of-the-Art Review: Whole Building, Building Envelope and HVAC Component and System Simulation and Design Tools.* (Jacobs and Henderson 2002).

It also can be useful to calibrate the simulation models against metered or sub-metered energy usage information in order to ensure that the models are performing well against some empirically based data from the local population.

5.2.4. Day-matching for Scheduled DR

Although day-matching is generally not suitable for non-event based resource options, one possible exception may be scheduled DR. Scheduled DR options prearrange with customers that have flexible loads to limit use of certain equipment on regularly scheduled days. For example, an agricultural customer that does a lot of irrigation pumping might be willing to only irrigate on selected days.

With this type of resource option, it may be possible to use load from non-scheduled days as a reference value for what a customer might have used in the absence of the DR incentive on the day that they have agreed not to use electricity. However, there could also be problems with this approach if, for example, the customer uses more electricity on non-scheduled days than they otherwise would have if they were not participating in the DR resource option. In this situation, using other days would overstate the magnitude of the reduction on the scheduled day. Other types of free ridership might also be present. For example, if a customer agrees not to irrigate or otherwise use load on a day when they typically don't use electricity for those purposes, they would simply be getting paid for doing nothing. Thus, while day-matching might work in theory for scheduled DR, it should be used with caution. There may be no real substitute for having pre-participation data on a customer in this situation to determine a suitable baseline.

6. EX ANTE ESTIMATION

This section contains protocols and guidance for ex ante estimation of both event and non-event based resource options. Ex ante estimation involves determining what the load impacts are likely to be for a given set of user-defined conditions. It does not include defining what those conditions are. For example, forecasting the size or makeup of the participant population at some future point in time is not part of impact estimation. Rather, impact estimation concerns estimating demand response given assumptions about the size and makeup of the participant population that are provided to the evaluator by someone else (e.g., regulators, planners or some other stakeholder).

Having said that, the evaluator has an important role in guiding the development of data needed to make such estimates, in that he or she must tell the interested user what information is needed. For example, for a residential critical peak pricing tariff, it would be important that the evaluator tell the prospective user that air conditioning ownership is a key driver of demand response. As such, it will be necessary for the prospective user to indicate not only that they expect the number of customers who sign up for the tariff to grow from X to Y over the next five years, but also that the percent of participants who own central air conditioning is expected to change from A to B over the same period. With this information, the evaluator can predict how the average impact per customer will change as the air conditioning saturation changes and how total impacts will grow as the number of participants increases.

Ex ante estimation requires development of a model that relates changes in demand response to changes in the exogenous variables that drive demand response. Whenever possible, the model should be based on ex post analysis of existing DR resource options. As such, all of the issues associated with ex post evaluation that have been raised in prior sections apply here as well. However, there are additional issues that are unique to ex ante estimation.

- Ex ante estimation may require developing estimates for values of key drivers that are outside the boundaries of historical experience (e.g., for extremely hot days that might not have occurred over the historical period) where the relationship of demand response and the variable of interest may differ from the relationship that exists within a narrower range of values;

- Ex ante estimation may require determining how demand response might evolve over time as participants become better educated about how to modify behavior in response to demand response stimuli or, alternatively, lose interest in modifying their behavior. The persistence of demand response impacts over time may also be impacted by degradation of or improvement in enabling technology, which may also need to be factored into ex ante estimates.

- Ex ante estimates are subject not only to the uncertainty associated with ex post impact estimates (e.g., due to sample selection, model specification and the like), but also to the additional uncertainty associated with the exogenous factors that

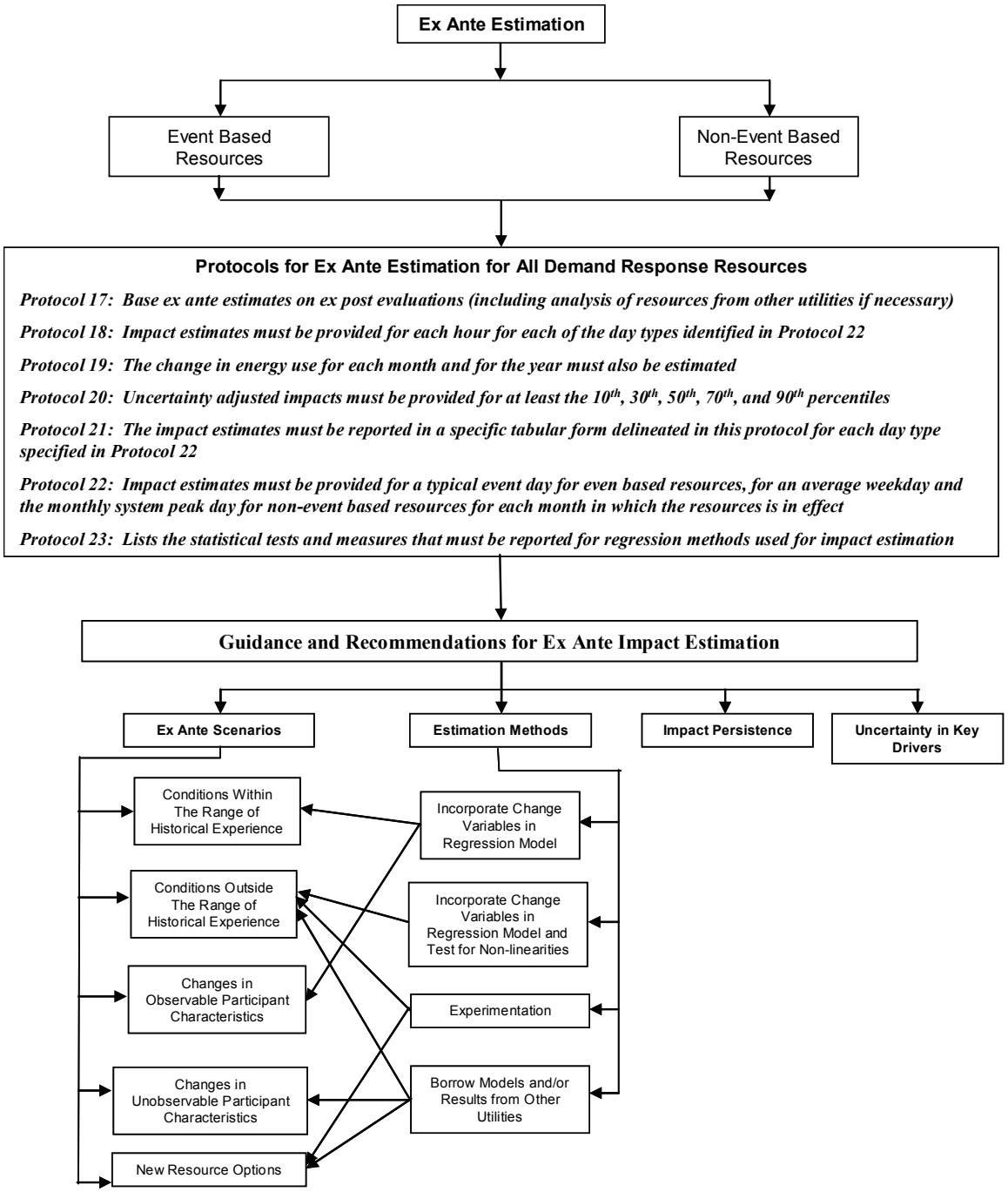
drive demand response (e.g., uncertainty in weather, participation levels and customer characteristics, etc.).

Figure 6-1 summarizes the topics covered in the remainder of this section. The protocols for ex ante estimation are similar to those for ex post estimation. Ex ante estimation should utilize information from ex post estimates. Where information is available on achieved DR impacts for similar activities this information is likely to provide an initial platform for forecasting future impacts. The best approach to ex ante estimation varies with the ex ante scenario for which estimates are needed. If estimates are needed for a scenario where the value of key drivers (e.g., weather or price conditions) differ, but are within the range of, historical experience, ex ante impact estimation is straightforward. However, if the need is for estimates under conditions that differ significantly from those that have occurred historically, or for brand new resource options, alternative methods including experimentation or borrowing impact estimates from other utilities may be required. Section 6.2 provides guidance regarding the methods that are most relevant for five different ex ante scenarios. Sections 6.3 and 6.4, respectively, discuss two other important topics associated with ex ante estimation, the persistence of demand response impacts and methods for incorporating uncertainty in key drivers into the impact estimates.

6.1.Protocols for Ex Ante Estimation (Protocols 17-23)

The protocols contained in this section are intended to apply to all types of ex ante estimation, including estimation for brand new DR resource options. It is expected that, in the vast majority of situations, ex ante estimation for resource options that are not new will be based, at least in part, on ex post evaluation studies. As such, the output requirements and protocols that apply to ex post evaluation should be able to be met for ex ante estimates developed from these studies, although there are some differences associated with the standard day types and forecast horizon and with factoring in changes in exogenous variables. Meeting the same protocols for brand new resource options may be more difficult, as the amount of available data and the statistical rigor that can be applied may be less for new resources than for existing ones. This is not always true, however, as illustrated by the example presented in Section 6-2. Information on the probability distributions associated with key drivers of demand response, or reasonable assumptions concerning the minimum, maximum and most likely estimates associated with key drivers, can be used along with Monte Carlo simulation modeling to develop uncertainty adjusted impact estimates even for new resources. As such, the same protocols apply for new resources, although it is recognized that even a “best efforts” level of commitment to meeting these requirements may fall short depending upon the nature of the new resource options and the degree to which data and/or models can be obtained elsewhere.

Figure 6-1. Section Overview



Protocol 17:

Whenever possible, ex ante estimates of DR impacts should be informed by ex post empirical evidence from existing or prior DR resource options. Evidence from resource options and customer segments most relevant to the ex ante conditions being modeled should be used, regardless of whether they come from the host utility or some other utility. If ex post estimates or models are not used

as the basis for ex ante estimation, an explanation as to why this is the case shall be provided.

Protocol 18:

The mean change in energy use per hour (kWh/hr) for each hour of the day shall be estimated for each day type and level of aggregation defined in Protocol 22. The mean change in energy use for the day shall also be estimated for each day type.

Protocol 19:

The mean change in energy use per month shall be estimated for non-event based resources and the mean change in energy use per year shall be estimated for both event and non-event based resources for the average across all participants and for the sum of all participants on a DR resource option for each year over the forecast horizon.

Protocol 20:

Estimates of the 10th, 30th, 50th, 70th and 90th percentiles of the change in energy use in each hour, day and year, as described in Protocols 17 and 18, and for each day-type described in Protocol 22, shall be provided.

Protocol 21:

Impact estimates shall be reported in the format depicted in Table 6-1 for all required day types and levels of aggregation, as delineated in Protocol 22.

It should be noted that there is a difference between Table 4-1, which applies to ex post estimation, and Table 6-1. Table 4-1 contains a column representing the observed load whereas Table 6-1 does not. Obviously, it is not possible to measure observed load in the future. However, every estimate of load impacts will have an implied reference load as the baseline against which impacts are estimated. The reference load column is included so that percent impacts can be calculated. Once again, temperature and degree hours are included primarily for comparison purposes across day types and resources. These variables may or may not have been those used in developing the estimates.

Table 6-1. Reporting Template for Ex Ante Impact Estimates

Hour Ending	Estimated Reference Load (kWh/hr)	Estimated Event Day Load (kWh/hr)	Estimated Load Impact (kWh/hr)	Temp (F)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
Daily									

Protocol 22:

The information shown in Table 6-1 shall be provided for each of the following day types using 1-in2 and 1-in-10 weather conditions for the average across participants and for the sum of all participants for each forecast year:

- ***For a typical event day for a 1-in-2 and for a 1-in-10 weather year for event-based resource options.***

- ***For the average weekday for each month in which the resource option is in effect for a 1-in-2 and for a 1-in-10 weather year for non-event based resource options*** [57]

- ***For the monthly system peak day for each month in which the resource option is in effect, for a 1-in-2 and for a 1-in-10 weather year for non-event based resources.***

Day type definitions and additional reporting requirements for each day type are summarized below:

Typical Event Day for a 1-in-2 and 1-in-10 Weather Year: This day type requirement applies primarily to event-based resources. It is meant to capture both the exogenous factors such as weather and the event characteristics for a day on which an event is likely to be called. The relevant characteristics can be defined by the evaluator. At a minimum, the following information shall be provided:

- ***An explanation of how the weather and any other relevant day-type characteristics were chosen***

- ***Detailed information on the timing and duration of the event or any other factors (e.g., notification lead time) that were explicitly factored into the impact estimates (e.g., factors that, if different than those reported, would change the estimated impacts)***

- ***The number of notified consumers included in the aggregate impact estimate***

- ***Any other factors that have been explicitly incorporated into the impact estimate, such as prices for price based resource options and population characteristics (e.g., air conditioning saturation, business type, etc.).***

Average Week Day for Each Month In A 1-in-2 and for a 1-in-10 Weather Year: This day type applies primarily to non-event based resources. It is meant to capture the weather conditions and other relevant factors for an average weekday. In addition to the information contained in Table 6-1, the following information must be provided:

- ***An explanation of how the weather and any other relevant day-type characteristics were chosen for the typical weekday in each month***

- ***The number of enrolled customers included in the aggregate impact estimate***

[57] If a resource is seasonal, only the months in which the resource is in effect must be reported.

- ***Any other factors that have been explicitly incorporated into the impact estimate, such as prices for price based resource options and population characteristics (e.g., air conditioning saturation, business type, etc.).***

Monthly System Peak Day for Each Month In a 1-in-2 and for a 1-in-10 Weather Year: This day type applies primarily to non-event based resources. It is meant to capture impacts for the day with the highest system load in each month. In addition to reporting all of the information shown in Table 6-1, the following information must be provided:

- ***An explanation of how the weather and any other relevant day-type characteristics were chosen for the typical monthly system peak day***
- ***The number of enrolled customers included in the aggregate impact estimate***
- ***Any other factors that have been explicitly incorporated into the impact estimate, such as prices for price based resources and population characteristics (e.g., air conditioning saturation, business type, etc.).***

Protocol 23:

All ex ante estimates based on regression methodologies shall report the same statistical measures as delineated in Protocols 10 and 16.

It should be noted that the day types described above, and that are incorporated in Protocol 22, are intended to be the minimum set of required day types, in part, to allow for comparisons across resources and to support long term resource planning. Additional day types may be of interest to many users. For example, impacts based on weather for a 1-in-10 year or 1-in-10 event day may be a relatively common need.

6.2. Guidance and Recommendations

Ex ante estimation concerns extrapolating the findings from ex post evaluations (of either the same resource or one similar enough so that logical inferences can be drawn) to a set of conditions that differ from those that have occurred in the past. The issues that must be addressed vary depending upon the conditions of interest and how much these conditions differ from those that have occurred in the past.

6.2.1. Ex Ante Scenarios

The five scenarios identified below are typical ex ante estimation scenarios across which issues and methods vary. There could also be scenarios of interest that combine elements from each of these scenarios.

Conditions within the Range of Historical Experience

The most straightforward scenario is when estimates are needed for a set of conditions that are within the range of those that have occurred in the past. An example would be development of an estimate for a DR resource option where the mix of customers is assumed to remain largely the same in the future as it was in the past and the weather conditions of interest—while not exactly the same as any particular day that occurred in the past—can be represented by temperatures that are below the maximum and above the minimum temperatures that occurred during the ex post evaluation period. Another example would be for an RTP tariff where estimates are needed for a set of prices that are inside the range of prices experienced previously. These examples merely require interpolation between prior extremes. Developing these estimates still requires a model that relates the variables of interest to demand response, but there is every reason to believe that the inferences drawn will be valid.

Conditions outside the Range of Historical Experience

A second scenario of potential interest might, once again, involve little change in the participant population. However, in this case, there is interest in knowing what the impacts might be for a day type where the weather or price conditions (or some other variable of interest) are outside the range that has been observed in the past. For example, one might want to know what the impacts would be for a 1-in-10 weather year or weather day, or for highly volatile market conditions where hourly prices exceed any that had previously occurred. These examples are much more challenging, as the functional relationship between the variable of interest and demand response may differ under these extreme conditions from what it was under the observed conditions.

For example, the relationship between the change in energy use associated with air conditioning and a change in temperature is reasonably linear over some range of temperatures, but highly non-linear at both the low and high end of the temperature range. A change in temperature from, say, 65 to 70 degrees will produce very little if any change in energy use because air conditioning typically is not running at either of those temperatures. Similarly, a change in temperature from, say 100 to 105 degrees, may produce little change in air conditioning energy use if most air conditioners are already running flat out at 100 degrees,[58] so higher temperatures do not increase energy use. For the same reasons, demand response may not occur at these extremes, regardless of the magnitude of the incentive provided, since thermostat adjustments at these extremes will have little impact on energy use. Consequently, if the model being used for ex ante estimation was developed from data on days that did not include these extreme conditions or, even if such conditions existed, the model assumed a linear relationship across the entire temperature range (e.g., it was miss-specified), it will not do a good job of estimating demand response impacts under these extreme conditions.

The same type of problem can arise when using demand models to estimate impacts for prices well outside the range of what has been observed historically. It may be, for example, that customers are not very price responsive at the very low end of the price range, when a change

[58] The threshold temperature above which most or all air conditioners will be running will vary depending upon the typical unit sizing practices for a location. It may be that many air conditioners will still be cycling above 100 degrees in some locations but most will be on in other locations.

has only a small impact on their bills, or at the very high end of the price range, when they have already made all of the reductions that they are willing or able to make. In between these extremes, customers may be relatively price-responsive. Recent evidence from a pricing experiment in New South Wales, Australia, for example, suggests that there is very little incremental effect associated with a change in prices when moving from a peak period price of $1.50/kWh to a price of $2.00/kWh. Implicitly, this evidence suggests that consumers in this service area have already made all of the adjustments they are willing to make at the $1.50/kWh price.

Changes in Observable Population Characteristics

A third scenario concerns estimating the change in demand response associated with a change in participant characteristics that are observable. This could occur, for example, for a demand response resource that is targeted at customers with air conditioning but open to all customers. Suppose that the initial marketing effort for this resource was quite effective at attracting customers from the target population, perhaps because it was initially only advertised in areas where the saturation of air conditioning was high. However, over time, through word of mouth or because of expansion of the DR program into other geographic regions where the saturation of air conditioning is lower, the saturation of air conditioning among participants might decrease. If demand response is tied to air conditioning ownership, this type of shift in the participant population will lead to an overall decrease in average demand response per participant, even as total demand response increases with increased participation.

Producing estimates for this type of scenario requires developing a model that relates the change in demand response to a change in the observable variables that are expected to differ over the forecast horizon. In some instances, this will be relatively straightforward while, in others, it may be more difficult. In the above example, if the early targeted marketing is so successful that the only customers currently enrolled in the resource are those with air conditioning, it will not be possible to establish a relationship between air conditioning ownership and demand response from the historical resource data. Under these circumstances, it may be necessary to use information from other utilities with similar resources but a more diverse mix of participants in order to adjust the impact estimates based on current participants so they reflect the future penetration of participants who do not have air conditioning.

Changes in Unobservable Population Characteristics

The fourth scenario is the most difficult one of all, as it involves developing estimates when there are reasons to believe that future participants will differ from those in the past in ways that are not easily tied to observable variables. This could be a reasonable expectation for any resource that is in the early stages of its lifecycle, as it may have only attracted “early adopters” who may not be terribly representative of the general population. Extrapolation to future participants may be even more challenging in a situation where a resource is changing from a voluntary, opt-in marketing approach to a voluntary, opt-out approach or to mandatory participation. Under these circumstances, it may be that current participants are more environmentally conscious, more price sensitive, or have lifestyles or business operations for which any negative aspects of demand response are less impactful than it is for the average customer. If so, extrapolating impacts derived from this group to a much broader population in

which those difficult to observe characteristics are much less prevalent will lead to an overestimate of demand response impacts.

New Resource Options

The final ex ante scenario involves estimation for brand new resource options. This scenario is similar to a scenario for an existing resource option where the future may differ significantly from the past. The two primary approaches to addressing this problem are relying on estimates from elsewhere and experimentation. California's SPP is an example of an experimental approach that developed the data necessary for the State's utilities to estimate likely impacts for critical peak pricing resources for residential and small and medium C&I customers in California that did not previously exist. Pilot resources and experiments are important methods to consider when developing ex ante estimates for new resources. However, if time or budget limits do not allow for an experimental approach to be used, the evaluator must make reasonable judgments to extrapolate results from evaluations of existing resource options.

6.2.2. Impact Estimation Methods

Developing impact estimates for a specific set of conditions that differ from those that occurred historically requires estimation of a model that will predict how demand response impacts change given a change in these conditions. As discussed in Section 4.2.2, in a regression context, this can be achieved using a model specification that includes interaction terms between exogenous variables of interest and resource variables. Equation 6-1 is an example of this type of model. This specification is one of several that were developed to estimate hourly impacts for residential critical peak prices tested in California's Statewide Pricing Pilot (SPP).[59]

$$\left[\left(\frac{kWh_i}{\sum_{i=1}^{24} kWh_i}\right)\right] = a_i + b_i\left(\frac{P_p}{P_{op}}\right) + c_i\left[\left(\frac{CDH_i}{\sum_{i=1}^{24} CDH_i}\right)\right] + d_i\left(\frac{P_p}{P_{op}}\right) * CAC + e_i\left(\frac{P_p}{P_{op}}\right) * \left[\left(\frac{CDH_i}{\sum_{i=1}^{24} CDH_i}\right)\right] + \sum_{k=1}^{n} g_k D_k + \varepsilon_i \qquad (6\text{-}1)$$

P_p = Peak period price

P_{op} = Offpeak period price

CDH_i = Cooling degree hours for hour i

D_k = Customer k's fixed effect

ε_i = Error in hour i

CAC = 1 if a household owns a central air conditioner, 0 otherwise.

[59] CRA International. Residential Hourly Load Response to Critical Peak Pricing in the Statewide Pricing Pilot. May 18, 2006. CEC website: http://www.energy.ca.gov/demandresponse/documents/index.html#group2

The above equation estimates the share of daily energy use in each hour as a function of the share of daily cooling degree hours in each hour, the peak-to-off-peak price ratio, air conditioning ownership and binary variables representing each customer (in order to control for cross-sectional differences in energy use). The model coefficients d_i and e_i, respectively, represent the change in price responsiveness given a change in air conditioning ownership and weather. This type of model can be used to produce ex ante impact estimates for any combination of weather conditions and prices that are not too far outside the boundaries of what occurred within the estimating sample, and for a mix of resource participants that have any saturation of air conditioning ownership that the evaluator might think is likely to occur over the forecast horizon. Models such as this can be used to produce ex ante forecasts for the scenarios outlined above in the sections describing conditions within the range of historical experience and changes in observable population characteristics.

This type of model can also be used for ex ante estimation for the scenario outlined in the section describing conditions outside the range of historical experience, but only under certain circumstances. Assuming that data exists for a reasonably wide range of variation in the variables of interest, the first step in model estimation should involve an exploration of different functional forms to assess whether a linear or non-linear relationship fits the data best. If nonlinearities are present within the estimating sample and can be captured in the functional form that is fit to the historical data, the model should do a better job of estimating impacts based on input variables that have values outside the historical boundary than if only a linear relationship is exhibited within the estimating sample (and assuming that there are logical reasons to believe that non-linearities exist at the extremes of the distribution even though they are not detectable from the historical data). The previous examples concerning air conditioning energy use at very low and high temperatures and incremental demand response at very high prices are cases in point. In these examples, logic and/or experience from elsewhere suggest that, at some point, impacts will not change given any incremental change in the exogenous variables.

Another approach to addressing this problem is to incorporate information from other DR resource programs or from other utilities. For example, the highest critical peak price tested in California's SPP was roughly $0.75/kWh for residential customers, which was roughly 5 times the standard price. If there was interest in knowing what the impacts would be for a price closer to $1.25/kWh or even higher, there is a risk that the price elasticities from the SPP would not apply. In this case, one could turn to other pricing experiments, such as the NSW pilot mentioned above, to see if much higher prices and/or price ratios were tested. If they were and the estimated price elasticities were comparable to those found in the SPP, there will be greater confidence in using the SPP model to produce estimates for prices outside the boundary of those tested in the pilot than if a different result were observed elsewhere.

A third approach to developing impact estimates for a scenario with conditions outside the range of historical experience is experimentation. If a resource is expected to be large and it is important to understand what happens at the extremes, it may be necessary to plan and conduct an experiment that creates the conditions of interest. For resource options such as load control, where an event might only be triggered during system emergencies and such emergencies often, but do not always, occur on very hot days, it could be useful to trigger the load control when a

hot day occurs but an emergency does not actually exist. Similarly, for an RTP tariff, if there is interest in knowing what might happen if prices go really high for a few hours, but such prices have never occurred, it might be possible to get permission to test a very high price signal on a sample of customers under market conditions where prices are typically high, just not as high as they might become at some point down the line.

As previously mentioned, the most difficult challenge occurs in the scenario described in the section about changes in unobservable population characteristics, when a DR resource program or tariff is expected to undergo a very significant transition from a small group of early volunteers to a much broader group of participants that might have unobservable characteristics that differ from those of the early participants. This might occur due to normal growth over the forecast horizon or a significant shift in marketing approach from a voluntary, opt-in tariff, for example, to an opt-out or mandatory tariff. In these circumstances, the past may not be a good guide to the future.

One approach to addressing this scenario is to explore whether or not it is possible to learn enough about the current participants to ascertain how they might differ from potential future participants—that is, to try and turn currently unobservable characteristics into observable characteristics. For example, if one is concerned that early adopters are more environmentally conscious or more budget-minded than what future participants would be, it might be possible to conduct a survey to explore whether or not the hypothesis is true. If it is not true, there will be greater confidence in extending the historical findings to future participants. If it is true, the survey data will not necessarily help you solve the problem, but at least it will confirm that you have one.

Another approach is, once again, to look elsewhere for data and information that can be used to gauge whether or not it is appropriate to extrapolate from the current population and resource characteristics to a different set of conditions. It may be that some other utility has a program or tariff with the characteristics of interest that can provide guidance into what impacts are likely to be. For example, if there was interest in knowing whether the impact estimates based on large C&I customers participating in a voluntary RTP tariff are suitable for estimating impacts given a shift to a mandatory RTP tariff, one could examine estimates based on New York's mandatory RTP tariff for large C&I customers and see how they compare to estimates from the current voluntary tariff. If they are similar, after controlling for differences due to customer mix and price variation, there will be greater confidence in using the current estimates than if they are quite different. There is a growing body of evidence from demand response resource options across the country that can and should be used whenever ex ante estimation must be done for a DR resource option that is expected to differ significantly from what has occurred in the past.

Another approach to this scenario involves experimentation. This is almost always an option, albeit a potentially expensive and time consuming one, for developing impact estimates where history or information from elsewhere is not a sufficient guide to what might happen over the forecast horizon. In the example discussed above where survey data revealed a difference between current and future participants on attitudes about the environment or cost consciousness, it is likely that demand response impacts will differ for future participants from those estimated from the current participant population. However, it might be impossible to know how impacts

are likely to change because the current participant population might not have any people who are not either environmentally or budget conscious. In this situation, it could be fruitful to conduct a small experiment in which the population of interest is recruited using some form of incentive to secure their participation and assess whether or not there is any difference in demand response between current participants and likely future participants.

Whenever it is necessary to rely on information from DR resource options or experiments conducted elsewhere, it is important to explore ways of adapting these estimates for differences in the resource option and population characteristics between the utility from which the information is obtained and the utility for which it is being used. In some situations, the available information may not be robust enough to allow for adjustments to be made, or even to obtain a thorough understanding of whether or not there are differences. Whenever differences are relevant and evident, however, they should be documented, even if they can't be adjusted for. The ideal situation occurs when it is possible to borrow a model from another jurisdiction that allows adjustments to be made.

6.3. Impact Persistence

Impact persistence refers to the period of time over which resource-induced impacts are expected to last. There are two key questions that influence how estimation of impact persistence might be approached:

- Do impacts persist beyond the life of the DR program or tariff?
- Do average impacts per customer change over time due to changes in consumer behavior and/or technology degradation?

For most demand response resources, the answer to the first question is no. In most instances, demand response can only be expected to occur for as long as the DR program is in effect. For example, for customers that are on time-varying rates or interruptible rates, or for customers who are paid an incentive to participate in a load control program, if the tariff or program is eliminated, impacts will also stop even if some technology was installed to enable the impacts to occur in the first place (e.g., like a load control switch).

An exception to this rule might be for some permanent load reduction resources, such as ice storage. If a utility implements a DR program that subsidizes ice storage, for example, the overall load shifting associated with the technology will probably persist as long as the technology remains operational, which could extend well beyond the termination of the program. Estimating persistence in this case requires estimating the effective useful life of the technology. Persistence may not extend beyond the resource life for all permanent demand response resources, however. For example, demand response associated with energy management systems or time switches may dissipate once a program incentive is eliminated, as consumers might disable the time switch or adjust their energy management system so they can operate end use equipment at times that are more convenient once an incentive is no longer provided.

For technology enabled demand response resources, such as direct load control, programmable communicating thermostats and autoDR, the average impact per participant may change over

time due to technology degradation. Unless there is a proactive effort to maintain and/or replace the technology to ensure that it remains operational, technology will eventually fail and the impacts associated with the technology will no longer exist. Persistence estimates for technology enabled resources must account for technology degradation.

For both technology and non-technology enabled resource options, changes in human behavior must be considered. For some resources, such as price-driven demand response, the average impact might increase over time as consumers become better educated and learn better ways to reduce energy use during peak periods or as they invest in equipment on their own, such as time switches or programmable thermostats in order to increase their demand response. On the other hand, responsiveness may fall over time if the savings associated with participation are not large enough to sustain the behavior initially observed while customer inertia or some other factor (e.g., mandatory participation) keeps participants in the resource even though they are no longer providing the same level of demand response.

The *EE Protocols* contain an extensive discussion of methods and protocols for estimating the effective useful life of various kinds of energy efficiency equipment. For resources that have been in place for an extended period of time and that have undergone multiple evaluations, surveys and on-site inspections of equipment can build a database over time that will allow for estimation of logistic curves and other functional forms that can be used to estimate the effective useful life of equipment. Given that most demand response resources are new and few evaluations have been done, these kinds of methods may not be an option currently, although there may be exceptions to this fact. For example, traditional load control of air conditioners has been used in the US by many utilities for many years and it may be possible to obtain data from some of these other resources that can be used to estimate annual failure rates for this type of technology.

The *EE Protocols* define a basic rigor level for degradation studies as follows:

> *Literature review for technical degradation studies across a range of engineering-based literature, to include but not limited to manufacturer's studies, ASHRAE studies, and laboratory studies. Review of technology assessments. Assessments using simple engineering models for technology components and which examine key input variables and uncertainty factors affecting technical degradation.*

These methods should also be considered for demand response impact persistence estimation for resource options where technology is a key component.

A potentially much more difficult aspect of persistence estimation concerns predicting how consumer behavior may change over time. The extent to which this is a concern will vary significantly across resource options. Resources involving the establishment of firm service levels and substantial penalties for violation of agreements are unlikely to see much degradation in demand response over time. On the other hand, price based resources such as critical peak pricing or RTP, or even technology-based options such as PCT programs that allow overrides, might experience either an increase or decrease in average response depending upon how much consumers value the benefits that are actually received relative to the discomfort, inconvenience or other customer costs that might occur. Most dynamic rate options have not been in place long

enough anywhere in the US to obtain good information on which direction these behavioral changes might go, or whether there is likely to be any change at all compared with the response that was estimated over a relatively short program history.

6.4. Uncertainty in Key Drivers of Demand Response

With ex ante estimation, it is important to consider not only the degree of uncertainty associated with the ex post evaluation parameters, which is largely tied to the accuracy and statistical precision of model parameters, but also the uncertainty associated with any significant drivers that underlie the ex ante estimates. Everything is uncertain in the future, and providing point estimates based on specific values for key variables can significantly overstate the true confidence that underlies the estimates.

Incorporating uncertainty into input values into estimates of demand response is straightforward using Monte Carlo simulation methods or similar approaches.[60] With Monte Carlo analysis, each variable that drives demand response can be represented by a probability distribution defined by an explicit set of characteristics. Standard software packages, such as Crystal Ball, can also accommodate correlations among exogenous variables (e.g., the fact that both price elasticities and reference values may increase with weather). The analysis software will pick a value from each input distribution and predict the demand response associated with that set of input values. This process will be repeated many times (1,000 draws from each distribution is relatively common) in order to produce the distribution of impact estimates that reflects the uncertainty associated with the driving variables as well as the model parameters.

The challenge in employing this (or any) method to represent the uncertainty in ex ante forecasts is developing probability distributions for the input values and incorporating the interdependencies in the relationships. In some cases, data exists that will allow for empirical estimation of the distributions. This is often the case for weather variables, and it might be true for other factors such as market prices (in a competitive wholesale market with a reasonably long history, for example). For other important variables, such as resource participation, it might be possible to develop reasonable estimates of minimum, maximum and most likely values. If so, the information can be used to create a triangular or beta distribution to represent the uncertainty, as these types of distributions can be fully defined with just these three values. Regardless of the method used to develop distributions for key drivers of demand response, the shape of those underlying distributions should be clearly described.

6.4.1. Steps for Defining the Uncertainty of Ex Ante Estimates

In the case of regression based impact evaluations, incorporating uncertainty in the regression parameters and in the input values for ex ante estimates is relatively straightforward, and involves the following steps:

1. Obtain the regression output, recording the parameters and their respective standard errors

[60] Monte Carlo simulation is a straightforward, widely used approach for reflecting uncertainty in key model parameters, but there may be other approaches that can be used to accomplish the same objective.

2. Obtain the variance-covariance matrix of the parameters

3. Convert the co-variances between parameters into correlations

4. Create a Monte Carlo model that replicates the parameters, their distributions, and inter-correlations

5. Incorporate the uncertainty associated with key drivers (e.g. temperature, participant characteristics) of the ex ante estimates and the inter-correlations among these drivers.

6. Run the simulation many times and obtain the confidence intervals.

An accurate estimate of the uncertainty associated with the model precision requires obtaining the full variance-covariance matrix of the regressions and incorporating any inter-correlations among the parameters. Simulating each parameter independently provides an inaccurate estimate of the confidence intervals. Likewise, correlations among the load impact drivers should be incorporated; otherwise the uncertainty estimates will be inaccurate.[61]

Nearly all statistical packages provide the full variance-covariance matrix of parameters if requested explicitly, and many easily provide it in the form of a correlation matrix, which is the format required by most Monte Carlo simulation software packages.

In cases where statistical packages do not translate the parameter co-variance matrix to correlations, the correlations can be obtained by the following equation:

$$\mathrm{corr}(\beta_i, \beta_j) = \frac{cov(\beta_i, \beta_j)}{\sigma_i \sigma_j} \mathrm{T} \qquad (6\text{-}2)$$

where

σ_i and σ_j are the standard error for coefficients β_i and β_j

6.4.2. Defining the Uncertainty of Ex Ante Estimates: Example

To illustrate how incorporating the uncertainty associated with key drivers affects the load impact estimates, the example from Section 4.2.2.10 is extended here. The example was developed using residential customer data for the CPP rate from California's Statewide Pricing Pilot for the summer of 2004. Only data from climate zone 3 (the hot climate zone representing California's central valley) treatment group was used. As such, the impacts reflect the incremental impact of a CPP rate layered on top of a TOU rate.

The key difference between ex post and ex ante uncertainty adjusted load impact estimates in the example is the fact that weather is uncertain in the future. In an ex ante setting the historical weather for the defined scenario, a typical event day, can be used to create the distributions by hour. For this example, the process was simplified by using a hypothetical distribution of

[61] Section 7.2.3 provides a detailed example of how failure to account for correlations can distort uncertainty estimates.

weather across event days rather than by mining historical data. Importantly, the hour-to-hour correlation for weather had to be incorporated in order for the uncertainty adjusted load impacts to be accurate.

Table 6-2 presents the uncertainty adjusted load impacts for both a fixed scenario and one that incorporates the uncertainty in weather.

Table 6-2. Uncertainty Adjusted Load Impacts from Regression Analyses

Day Type:	Average Event day for 2004 - SPP Climate Zone 3												
		Fixed scenario						Stochastic Scenario (incorporates uncertainty associated with weather)					
			Percentiles						Percentiles				
Hour Ending	Temperature	Mean Impact (kW)	10%	30%	50%	70%	90%	Mean Impact (kW)	10%	30%	50%	70%	90%
1	71.2	-0.01	-0.03	-0.02	-0.01	0.00	0.01	-0.01	-0.04	-0.02	-0.01	0.00	0.01
2	70.0	-0.03	-0.06	-0.04	-0.03	-0.02	-0.01	-0.04	-0.06	-0.04	-0.03	-0.02	-0.01
3	68.8	-0.06	-0.09	-0.07	-0.07	-0.05	-0.04	-0.07	-0.09	-0.08	-0.07	-0.06	-0.04
4	67.8	-0.07	-0.10	-0.08	-0.07	-0.06	-0.05	-0.07	-0.10	-0.08	-0.07	-0.06	-0.05
5	66.9	-0.06	-0.08	-0.07	-0.06	-0.05	-0.03	-0.06	-0.08	-0.07	-0.06	-0.05	-0.03
6	66.1	-0.05	-0.07	-0.06	-0.05	-0.04	-0.02	-0.05	-0.07	-0.06	-0.05	-0.04	-0.02
7	65.9	-0.03	-0.06	-0.04	-0.03	-0.02	-0.01	-0.03	-0.06	-0.04	-0.03	-0.02	-0.01
8	67.2	-0.03	-0.06	-0.04	-0.03	-0.02	-0.01	-0.03	-0.06	-0.04	-0.03	-0.02	-0.01
9	70.1	-0.02	-0.04	-0.03	-0.02	-0.01	0.01	-0.02	-0.05	-0.03	-0.02	-0.01	0.01
10	74.4	-0.04	-0.07	-0.05	-0.04	-0.03	-0.02	-0.06	-0.10	-0.07	-0.05	-0.04	-0.02
11	78.7	-0.05	-0.07	-0.06	-0.05	-0.04	-0.02	-0.06	-0.11	-0.07	-0.05	-0.03	-0.01
12	82.9	-0.08	-0.11	-0.09	-0.08	-0.07	-0.05	-0.08	-0.14	-0.10	-0.08	-0.05	-0.02
13	86.4	-0.05	-0.08	-0.06	-0.05	-0.04	-0.02	-0.05	-0.12	-0.07	-0.05	-0.02	0.01
14	89.1	-0.06	-0.09	-0.07	-0.06	-0.04	-0.02	-0.06	-0.13	-0.09	-0.06	-0.03	0.01
15	90.8	-0.18	-0.21	-0.19	-0.18	-0.17	-0.16	-0.18	-0.25	-0.21	-0.18	-0.16	-0.12
16	91.7	-0.16	-0.18	-0.17	-0.16	-0.15	-0.13	-0.16	-0.23	-0.19	-0.16	-0.13	-0.08
17	91.6	-0.13	-0.16	-0.14	-0.13	-0.12	-0.11	-0.13	-0.21	-0.16	-0.13	-0.10	-0.05
18	90.5	-0.09	-0.11	-0.10	-0.09	-0.08	-0.06	-0.09	-0.17	-0.12	-0.09	-0.06	-0.01
19	88.2	-0.05	-0.08	-0.06	-0.05	-0.04	-0.03	-0.06	-0.14	-0.09	-0.05	-0.02	0.02
20	84.5	0.16	0.12	0.14	0.16	0.17	0.19	0.15	0.07	0.12	0.16	0.19	0.22
21	80.2	0.26	0.23	0.25	0.26	0.27	0.28	0.25	0.19	0.23	0.26	0.28	0.30
22	76.7	0.24	0.21	0.23	0.24	0.25	0.26	0.23	0.18	0.21	0.23	0.25	0.27
23	74.3	0.19	0.16	0.18	0.19	0.20	0.21	0.17	0.13	0.16	0.18	0.19	0.20
24	72.6	0.10	0.08	0.09	0.10	0.11	0.12	0.09	0.06	0.08	0.09	0.10	0.11

Figure 6-2 reflects the uncertainty associated with the load reduction, presented as percent change in energy use, during the peak period hours for the fixed scenario. Figure 6-3 reflects the uncertainty adjusted load impacts that incorporate the uncertainty of weather. Both figures employ the same horizontal scale in order to allow for easy comparisons.

As seen in the figures, the difference is not trivial. The ex ante estimate under a fixed scenario presents substantially narrower distributions. If, for example, a planner was interested in the load impacts that could be obtained with 90% confidence (i.e., the 10th percentile) the fixed scenario produces an estimated reduction of 0.09 kW per customer. The stochastic scenario, on the other hand, produces a 10th percentile estimate of .05 kW per customer. This is a difference of roughly 80%.

Figure 6-2. Uncertainty Associated with Load Reductions

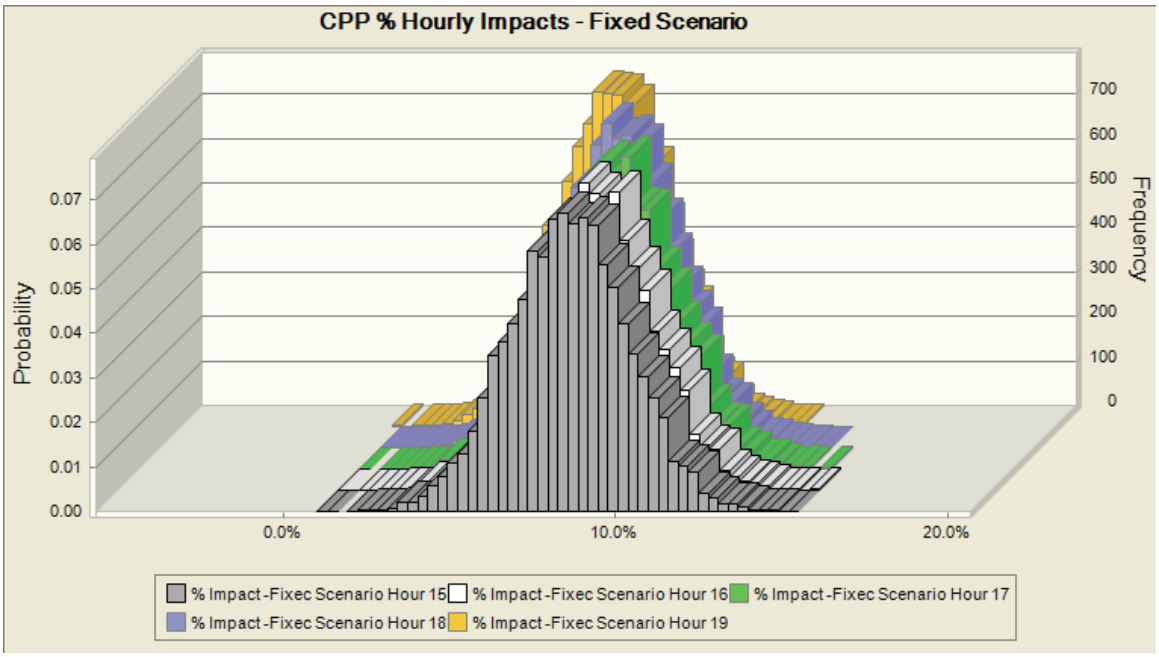
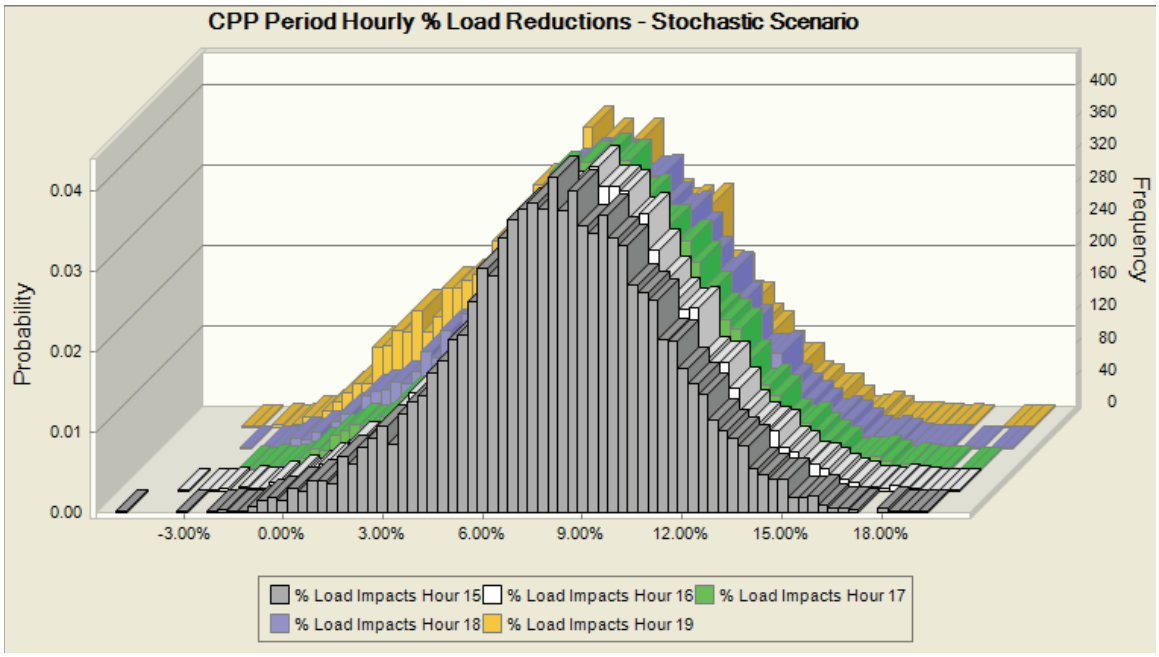


Figure 6-3. Uncertainty Adjusted Load Impacts



7. ESTIMATING IMPACTS FOR DEMAND RESPONSE PORTFOLIOS (PROTOCOL 24)

The methods and guidance provided in prior sections all focus on estimation of load impacts for individual DR resources. It is often important to also estimate the aggregate impact of a portfolio of DR resources managed by utilities, the CAISO or the state as a whole. This section discusses issues related to estimating the load impacts for DR portfolios and quantifying the uncertainty associated with using DR portfolios for operations and planning purposes.

To date, there has been little work done on estimating the aggregate impact of a portfolio of DR resources. As such, we believe it is premature to propose a set of prescriptive protocols regarding how best to develop impact estimates for DR portfolios. However, it is recognized that there can be overlaps in participation across DR activities (e.g., a nonevent-based pricing program and an event-based load curtailment program) and the success of one DR activity may influence the impacts that can be achieved by another DR activity. As a result, a protocol that is meant to identify synergies (positive or negative) and overlaps across programs is included.

Protocol 24:

The evaluation of a DR resource should identify correlations, synergies and overlaps across the set of DR resource options offered in a region or being proposed for a region. A judgmental determination of the impact of the magnitude of adjustment in program impacts should be made for all programs. In some cases, a zero adjustment may be recommended. In other cases, identified correlations, synergies and overlaps may result in a recommended adjustment to the ex ante estimate of program impacts.

These protocols acknowledge that synergies between programs may exist and that the identification of these synergies is a useful first step to developing DR impacts at the portfolio level. For planning, the overall resource contribution of the DR portfolio is likely to be an important consideration in deferring other new resources.

The balance of this section provides guidance regarding important issues that should be addressed as part of a more detailed DR portfolio analysis. This analysis is presented as a straw man, five-step process for developing portfolio impact estimates that could be used in evaluations that want to aggressively pursue a full analysis of portfolio impacts, but the only analysis required is protocol 23, which is based on a judgmental approach.

This section provides guidance regarding important issues that should be addressed as part of DR portfolio analysis and presents a straw man, five-step process for developing portfolio impact estimates.

Among the issues that should be considered when developing impact estimates for DR portfolios are:

- The quantity of demand response that can occur varies within resources and across participants based on conditions that vary systematically with weather, day-of-week, etc. For portfolio analysis, it is essential that common values for key drivers that affect multiple programs be used to develop individual program impacts prior to aggregation.

- Interactions between DR resources need to be explicitly considered. In practice, participants can enroll in multiple DR resource options that may be triggered under similar conditions. For example, a customer may be enrolled in both a demand bidding resource and a curtailable resource with a firm load level. The customer can submit bids at any point, but the load impacts for demand bidding should take into account whether or not curtailment notices were sent and, in response, some customers already reduced load.

- Individual DR resources may or may not be deployed at the same time and, even if deployed at the same time, may not be deployed to full potential. Portfolio analysis must define a set of scenarios according to a variety of characteristics (e.g., weather conditions, notification lead time, event window, day of week, etc.) and a determination made concerning which DR resources are likely to be called and available for event-based options and are likely to provide demand response for non-event based resource options.

- When developing uncertainty adjusted impacts across a portfolio of programs, it is typically not valid to simply add up the percentile impact estimates (e.g., 10th, 30th, 50th, 70th, or 90th percentiles) across the different DR programs.

- The value of a portfolio is not simply represented by instances in which participants reduce load, it also includes the option value of having the DR as a resource, making it important to obtain an accurate assessment of the total DR resources and the uncertainty/confidence surrounding that estimate.

Figure 7-1 outlines a five-step process for developing load impact estimates for a portfolio of DR resources.

1. *Define Scenarios:* The first step is to characterize the conditions for which portfolio estimates are needed. There are a wide variety of conditions that may characterize a scenario, including weather, day of week, the start and stop time and available notification lead time for event-based resources, and many others.

2. *Determine Resource Availability:* Given the conditions outlined in Step 1, the availability of each DR resource must be determined. Depending upon the conditions, some resources may not be fully available, or available at all.

3. *Estimate Uncertainty Adjusted Average Impacts per Customer for Each Resource:* In this step, it is important to use the same input values for key drivers of demand response

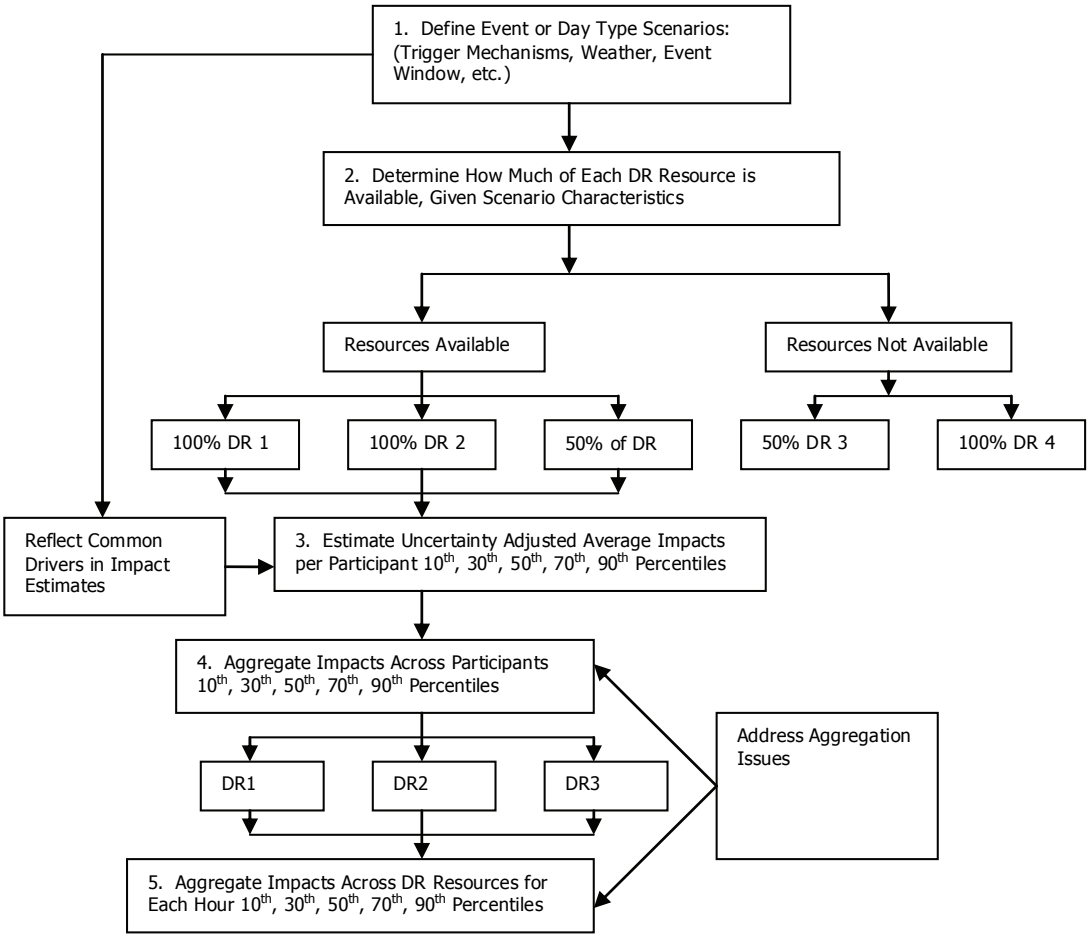
for each DR resource. It is also important to incorporate the uncertainty associated with model parameters and the underlying drivers.

4. *Aggregate Impacts Across Participants:* This step simply involves multiplying the average values developed in Step 4 by the number of customers notified for event-based programs.

5. *Aggregate Impacts Across Resources:* The final step involves aggregating the load impacts for each resource option by hour. It is not correct to simply add up the 10th percentile values for each resource, for example, in order to arrive at the 10th percentile estimate for the portfolio. The aggregation process must properly account for the underlying distributions.

A more detailed discussion of these five steps is contained below in Section 7.2. Prior to that discussion, we address a number of issues that, if not properly addressed, can lead to errors in portfolio impact estimates.

Figure 7-1. Estimating Impacts for DR Portfolios



7.1.Issues in Portfolio Aggregation

The challenge in aggregating load impact estimates from individual resources to arrive at portfolio level load impact estimates is not in calculating the expected value (mean) of the load impact. Rather, it is in describing the level of uncertainty associated with the aggregated estimates. Estimating risk requires more than knowledge of the mean, it requires an accurate description of the uncertainty in the underlying estimates.

Key considerations in calculating a portfolio's uncertainty are:

- Ensuring proper aggregation of individual resource load impacts,
- Correctly modeling the form of statistical distribution of the load impact (e.g., normal, beta, gamma, etc.), and
- Correctly taking account of correlations among the effects of the various resources.

With energy efficiency portfolios, the portfolio analysis framework typically calls for aggregating individual DR resource impacts assuming that the impacts associated with each resource option are independent and normally distributed. This simplifies the calculations necessary for aggregation. In the case of DR portfolios, these simplifying assumptions are probably not valid or, at least, should not be assumed to be valid without question.

Producing aggregate load impacts from probability distributions for individual resource options that have different shapes and that may be correlated with common factors such as weather and other factors will probably require the use of Monte Carlo simulation, a transparent approach grounded on real (not hypothetical) distributions and correlations whenever possible.[62] With Monte Carlo analysis, the load impacts from each resource in the portfolio should be represented by a probability distribution defined by an explicit set of characteristics. Standard software packages, such as Crystal Ball, can accommodate correlations among the distributions of load impacts for individual resources.

The following subsections discuss a number of issues that, if not properly addressed, will lead to erroneous impact estimates for DR portfolios. The most common problems typically arise from making incorrect assumptions about the form of the probability distribution of load impacts or by failing to include correlations among the impacts produced by the resources in the portfolio.

The following discussion relies on a simple, hypothetical example of a DR resource portfolio consisting of four resource options: an interruptible/curtailable tariff; a Critical Peak Pricing tariff for small and medium commercial customers; a residential A/C cycling program; and a two-way programmable thermostat program. Keep in mind that the examples used here are strictly hypothetical.

It is assumed that some of the resource options have load impact estimates that are not normally distributed and some of the load impacts from the resources are correlated. The distributions and correlations for the example are intentionally exaggerated to better illustrate the three basic complexities in portfolio aggregation. For each of the three key considerations, the example proceeds by presenting the expected load impacts, the standard deviation, and the distribution shape for the DR portfolio using both the correct and incorrect approach to aggregation, as explained below. A visual depiction of the distributions and the correct and incorrect portfolio aggregation estimates are shown in the right hand column of each table below.

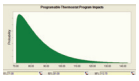
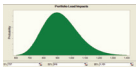
7.1.1. Errors Resulting from Improper Aggregation of Individual Resource Load Impacts

A common mistake made in DR planning is to de-rate individual resources and then sum the de-rated values to produce a de-rated portfolio estimate. For example, for a planning application, a utility may be interested in only counting the DR load impacts that are 90% certain to be delivered when called upon. The intuitive but incorrect approach would be to de-rate individual resources, and sum the 90th percentile values for each resource to reach an estimate of the 90th

[62] In theory, the convolutions of the underlying distributions of load impacts from different DR resources could be accomplished with calculus, but it is much easier to do so with Monte Carlo simulation.

percentile value for the DR portfolio. This approach incorrectly calculates the uncertainty in the portfolio and undercounts DR when 90% certainty is required over the portfolio. Table 7-1 contains estimates using the correct and incorrect methods.

Table 7-1. Comparison of DR Portfolio Aggregation: Correct Approach versus Summing Across the Percentile Estimates for Each Resource

PROGRAM	Expected Load Impact Capacity	σ	Percentiles 10%	50%	90%	Distribution Visual
Interruptible/Curtailable	401.2	98.7	279.3	395.0	531.0	
CPP - Small and Med C&I	120.7	31.6	88.5	113.1	162.9	
A/C Cycling	320.5	33.5	284.8	313.5	365.4	
Programmable Thermostat	85.0	14.0	72.0	81.0	103.4	
PORTFOLIO	927.4	129.6	770.5	916.6	1,098.3	
PORTFOLIO - INCORRECT AGGREGATION	927.4		724.6	902.6	1,162.6	

The difference between the incorrect and correct aggregation method may not be trivial. The incorrect approach underestimates the load impacts for the DR portfolio with 90% certainty by roughly 46 MW (e.g., 770.5 – 724.6). The DR resource can provide 6.3% more demand response than the estimate based on the incorrect method. The disparity associated with the inaccurate method would be even greater if more of the distributions were normal or skewed to the right.

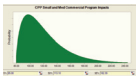
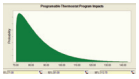
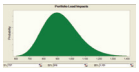
Importantly, the incorrect aggregation method also incorrectly describes the magnitude and distribution of the uncertainty. This can be seen by looking at the distribution depicted at the right hand column of Table 7-1 for the rows associated with the correct and incorrect approaches, both of which are on the same scale. The example highlights the importance of properly aggregating individual load impacts and risk.

7.1.2. Errors Resulting from Incorrect Assumptions About Underlying Probability Distributions

Estimation error can also occur by assuming that the probability distributions for the individual resource impact estimates are normal when in reality they are not. In this example, we assume that portfolio estimates must reflect the influence of weather which is a stochastic variable. Once again, we are interested in calculating the DR load impacts that are 90% certain to be delivered when called upon at the portfolio level. In this example, the uncertainty estimates must account for both the statistical precision of the model parameters and the stochastic component associated with the weather variable. Generally, the uncertainty of statistical estimates is normally

distributed and non-normal distributions will arise because non-fixed variables that drive demand response, such as weather, may not be normally distributed. In the hypothetical example, the A/C cycling and programmable thermostat resources are most skewed since their impacts can be expected to increase with temperature, the driver of load impact variation that is not normally distributed. Under these circumstances, as shown in Table 7-2, the difference between the estimates using the correct and the incorrect approaches is smaller (e.g., 16.2 MWs) but, clearly, the shape and amount of uncertainty around the estimate is not properly represented if normal distributions are assumed.

Table 7-2. Comparison of DR Portfolio Aggregation: Correct Approach versus One That Assumes Normal Distributions

PROGRAM	Expected Load Impact Capacity	σ	Percentiles 10%	50%	90%	Distribution Visual
Interruptible/Curtailable	401.2	98.7	279.3	395.0	531.0	
CPP - Small and Med C&I	120.7	31.6	88.5	113.1	162.9	
A/C Cycling	320.5	33.5	284.8	313.5	365.4	
Programmable Thermostat	85.0	14.0	72.0	81.0	103.4	
PORTFOLIO	927.4	129.6	770.5	916.6	1,098.3	
PORTFOLIO - INCORRECT	927.4	109.8	786.7	927.4	1,068.2	

7.1.3. Errors Resulting from a Failure to Capture Correlations across Resources

Correlations across individual resources are particularly important to incorporate into the portfolio impact estimates. If load impacts across resource options are positively correlated, this will increase the variation in the portfolio level load impacts. If the load impacts are negatively correlated, the correlation has the opposite effect – that of narrowing the variance for the portfolio impacts.

In practice, portfolios may have a mix of positive and negative correlations among individual resource options. As detailed later, the preferred approach is to incorporate the interdependencies among individual resources by using a common set of input values when estimating the uncertainty adjusted impact estimates for each resource option, a bottom-up approach. In this manner, the correlation due to the common drivers is accounted for, as long as the drivers are identified. However, it may not be possible to capture all correlations in this manner. If factors correlated with load impacts of multiple resources are not accounted for in the estimates for the DR portfolio aggregation scenario, it could potentially lead to correlation between the individual resource impact estimates.

In the hypothetical example shown in Table 7-3, the individual resource estimates already factor in a key driver of demand response, temperature. However, we assume that there correlations due to other factors that were not incorporated in producing the resource level impact estimates. For example, geographical distribution could be a common factor unaccounted for in the load impacts if, say, customers in some regions are more or less pre-disposed to provide larger amounts of load impacts entirely separate from the temperature effect. Admittedly, the correlations employed in this example are unrealistic, but it illustrates a key point. Figure 7-2 provides a visual display of the assumed correlations.

For the example, the analysis that did not incorporate the correlations incorrectly stated the amount and shape of the uncertainty for the portfolio's DR resources.

Table 7-3. Comparison of DR Portfolio Aggregation: Correct Approach versus One that Does Not Correct for Correlations

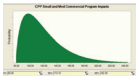
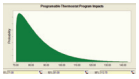
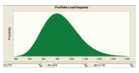
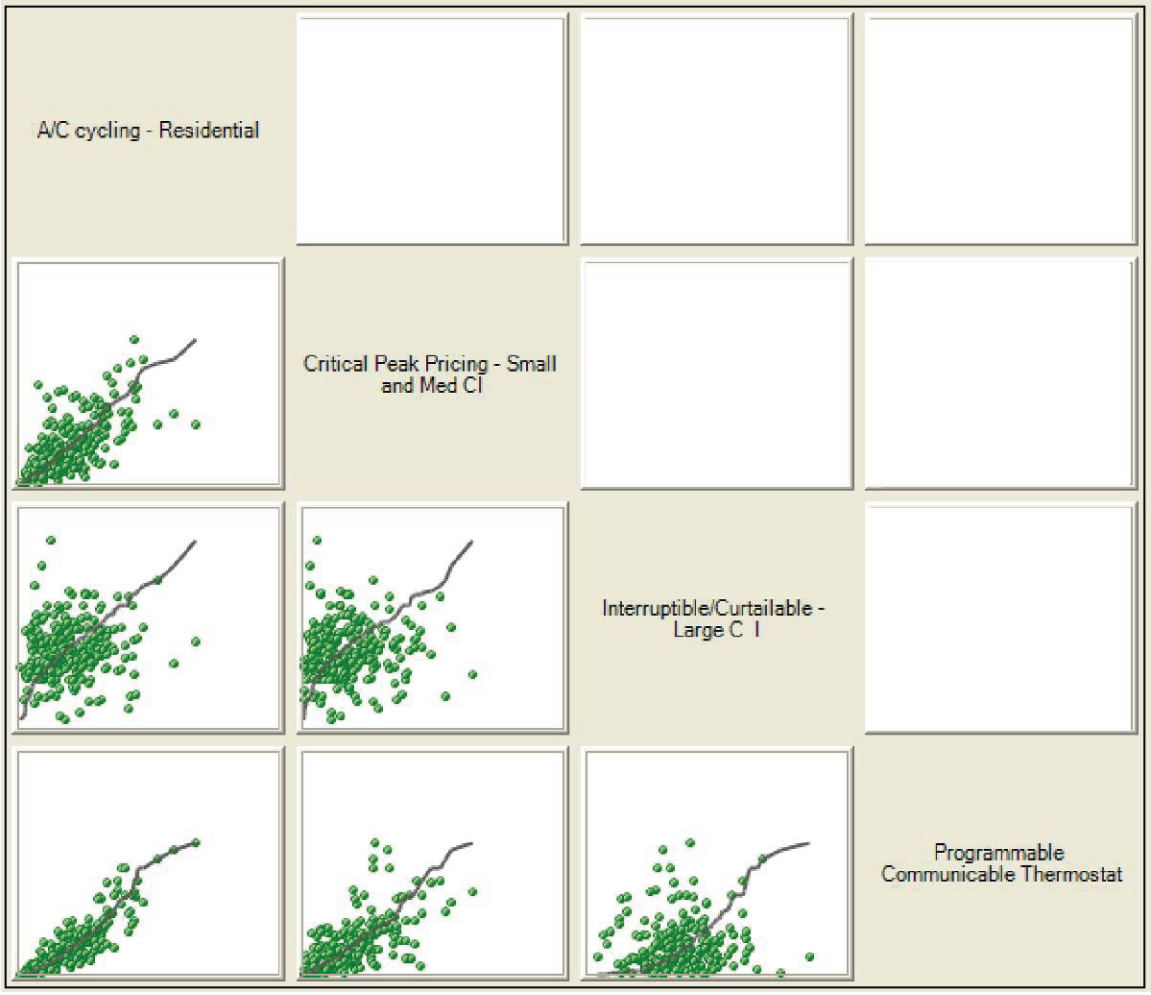
PROGRAM	Expected Load Impact Capacity	σ	Percentiles 10%	50%	90%	Distribution Visual
Interruptible/Curtailable	401.2	98.7	279.3	395.0	531.0	
CPP - Small and Med C&I	120.7	31.6	88.5	113.1	162.9	
A/C Cycling	320.5	33.5	284.8	313.5	365.4	
Programmable Thermostat	85.0	14.0	72.0	81.0	103.4	
PORTFOLIO	927.4	129.6	770.5	916.6	1,098.3	
PORTFOLIO - INCORRECT	927.4	109.2	792.2	922.4	1,072.6	

Figure 7-2. Correlations Underlying Example in Table 7-3



7.2.Steps in Estimating Impacts of DR Portfolios

In the introduction to this section, we outlined a five-step process that could be used for impact estimation for DR portfolios. The remainder of this section provides additional detail regarding that process.

7.2.1. Define Event Day Scenarios

Since demand response load impacts can vary by temperature, day type and other factors, using a common event or day type definition for all portfolio elements is a necessary first step in calculating DR portfolio load impacts. Day types can be defined in a variety of ways, including: based on temperature; system load; and ISO system-wide or zonal emergencies.

System load levels are useful in defining event days and scenarios because the need for Demand response tends to coincide with high-load days, although the relationship is not perfect. Generation outages, transmission outages, level of imports, wind generation, and the accuracy of

the load forecast are also key factors affecting whether or not a resource shortage occurs on a particular day.

Weather patterns are useful in helping define suitable event days or scenarios for DR portfolio aggregation because, in many cases, they are directly related to the amount of demand response that individual resources may deliver. For example, an A/C cycling resource may deliver more load reduction on a hotter day than on a cooler one, while a demand bidding resource may deliver less if events had been called for the days leading up to that day.

ISO called emergencies may also serve to define the common DR portfolio event days or scenario. Importantly, they will reflect the conditions under which DR resources are most needed and, by default, factor in the other drivers of resource shortages. The one drawback is that a larger set of data may be needed in order to define the common event day or scenario for purposes of DR portfolio aggregation.

The definition of event days or scenarios should be grounded in historical data if possible. If the level of load response for a resource is affected by temperature, it will be necessary to compute a weighted temperature for the scenario that reflects the geographical distribution of the participant population. While the scenario remains common, the temperature used to obtain estimates from individual resources may differ from resource to resource because of different participant characteristics and geographical distribution.

7.2.2. Determine Resource Availability

Individual DR resources have different triggers, event durations, notification periods, restrictions on operations and hourly impacts. At times, they may directly interact, for example, when a participant is enrolled in two demand response resources. In other instances, the load impacts from individual resources may be correlated, which affects the certainty of the portfolio load impacts.

The first step in analyzing the DR portfolio's resources based on a common scenario is to determine the likelihood that individual resources could operate simultaneously given the scenario characteristics and the resource trigger and notification requirements.

The second step is to assess whether the resources share the same participants and, if so, whether the load reduction in question is sufficiently large so as to require attention and resources for untangling those load impacts.

The third step is to assess whether or not the load impacts are correlated, and, if so, in what way. This step requires some attention and caution, as correlations in load impacts across resources affect the uncertainty associated with the portfolio load impacts. Specifically, there are two types of correlation that must be accounted for, correlations among the load response drivers (e.g., temperature and compliance), and, in the case of regressions, correlations between model parameters.

7.2.3. Estimate Uncertainty Adjusted Average Impacts per Participant for Each Resource Option

With the scenarios defined and the participants properly allocated, the next step is to estimate the individual impacts and the level of certainty around those estimates for each resource. Preferably, this is where any relationships between resources must be accounted for. In cases where the common scenario is not fixed, calculating the confidence intervals for the DR portfolio will require a Monte Carlo simulation approach.

The first task is to identify the common factors that drive load response for multiple resources, e.g., weather and day of week. These drivers should already be identified and accounted for in the individual evaluation studies, since factors that affect load are of particular interest.

The second task is to model the load impacts for individual resources taking into account 1) the uncertainty in the parameters and the stochastic components of the scenario, and 2) correlations between model parameters. By including the drivers of demand response in individual resources, which may be common across resources, the relationships across resources are accounted for in the certainty bands. It is critical for evaluations to model and account for factors that influence the customer load response. If this is not done, it will not only provide less accurate load impact estimates for the resource, but the correlations with other demand response resources cannot be easily controlled for regardless of the attentiveness paid in aggregating the portfolio.

7.2.4. Aggregate Impacts across Participants

Given unbiased estimates of the average hourly load impacts, estimating an individual resource's load impacts is straightforward.

To obtain the load impact estimates of a resource for any given hour of a day type, the average hourly load impacts and confidence bands are multiplied by the number of participants enrolled or the number notified. Multiplying the average by the number of enrolled customers produces an estimate of resource potential. Multiplying by the number of individuals notified provides an assessment of what the load impacts would be, ex ante, in an operations setting.

$$\text{Load Impact Capacity}_{dt} = \text{average load impact}_{dt} \times \text{total enrollees}_{d} \qquad (7\text{-}1)$$

Where d = day type
t = hour of day

7.2.5. Aggregate Impacts across Resources Options

The final step involves aggregating the load impacts, by hour, across DR resources. This requires both adding up the mean values and calculating the uncertainty or confidence bands around the portfolio level estimate.

Aggregating the mean load impact of a DR portfolio is straightforward. The sum of the average expected load impacts of individual resources produces the average load impacts for the entire portfolio, provided that the individual resources operate simultaneously. If the resources do not operate simultaneously under the defined scenario, it is necessary to weight the individual resource load impact capacity by the likelihood that the resources would operate under the scenario.

The more complex task is best calculating the certainty of the DR portfolio's load impacts. How best to approach this depends on whether or not the distribution for all resource load impact estimates are normally distributed. The methods for aggregating load impacts under the two different circumstances is discussed below, but first it is important to clarify the difference between the two potential cases, so that the correct method is identified and employed in the aggregation.

The individual resource impacts are likely to be normally distributed when the underlying scenario is fixed and the relationships of load impact to, for example, weather and day of week are already accounted for in the hourly individual resource predictions. This is because the certainty is largely tied to the accuracy and statistical precision of model parameters.

The individual resource impacts are less likely to be normally distributed when the underlying scenario is stochastic. For example, in the stochastic scenario, the expected ISO weighted temperature for hour 1700 may be between 95-104 degrees with a median of 98.5. The implications are twofold. The confidence bands for the DR portfolio load impacts must incorporate uncertainty in weather since the scenario is not fixed. Second, the estimates and confidence bands will vary, by hour, for resources with weather sensitive load impacts and those impacts will be correlated. With non-normal distributions, the DR portfolio uncertainty can be computed via either calculus or Monte Carlo methods.

The impact of correlations and stochastic scenarios on the certainty of the estimates are preferably accounted for in step 3, as described above. If, however, the evaluations do not calibrate the load impact estimates for factors that drive load impacts for multiple resources, it might be necessary to incorporate the correlations at the aggregation stage, although it will likely be difficult to obtain empirically based, accurate estimates of correlations across resources at this stage.

When the certainty around impact estimates for all DR resources are normally distributed, they can be accurately described by the standard errors, which can be aggregated and used to produce the certainty around the DR portfolio estimates. This approach mirrors the discussion on integrating the results from multiple evaluation studies presented in Chapter 12 of *The California Evaluation Framework.*

The standard error from multiple resources can obtained be by the following equation:

$$\sigma_{\text{Portfolio}} = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots \sigma_n^2} \qquad (7\text{-}2)$$

The standard error of the portfolio, along with the mean, can then be used to recreate the distribution and compute confidence intervals as described in any basic statistics textbook. An alternative method is to create a joint probability surface, incorporating the load impact uncertainties for individual resources via Monte Carlo simulation.

In cases where the common scenario is not fixed, the certainty around individual resource estimates may or may not be normally distributed. Calculating the confidence intervals for the DR portfolio requires either calculus or a Monte Carlo method approach. If done properly, both will produce the same results, but Monte Carlo methods are less prone to error and more transparent to the reviewers.

8. SAMPLING

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 offerings, and it is costly to install interval meters that are needed to estimate impacts. 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.

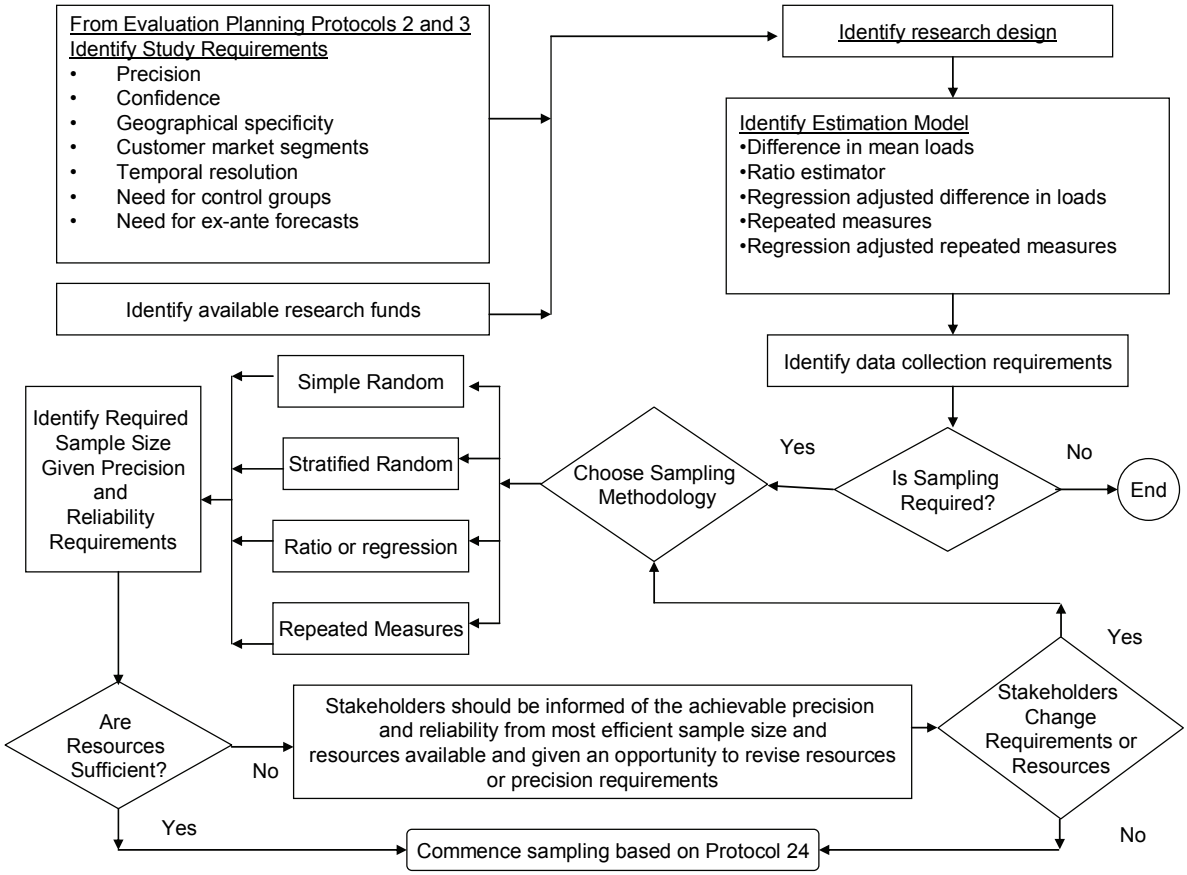
The decision to employ sampling and the numerous technical decisions required in sample design are driven by the broader research issues that are addressed during evaluation planning. These issues were discussed in detail in Section 3 and must be addressed in meeting the requirements associated with Protocols 2 and 3. 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 8-1 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);
- 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 8-1. Sample Design Process Diagram



8.1.Sampling Bias (Protocol 25)

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*.

> **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.

A variety of methods and procedures can be used to help ensure that the effects of selection bias on load impact estimation are minimized and that the impacts of any bias are clearly understood. Protocol 25 is intended to help ensure that these procedures are applied.

Protocol 25:

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.

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 these 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. ***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;***

6. ***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. This study should focus on discovering and describing any sampling bias that may have occurred as a result of selection. This should be done by comparing the known characteristics of the observed sample with the known characteristics of the population. Known characteristics would include such variables as historical energy use, time in residence, geographical location, reason for attrition from sample, and any other information that may be available for the population and sample.***

7. ***If selection bias is suspected, the evaluator must describe it as well as any efforts made to control for it.***[63]

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 some significant way. *Whether bias is induced by selection depends on whether the selection is somehow related to the*

[63] 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

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.

8.2.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 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 8-1:

$$\sigma_m^2 = \frac{\sigma^2}{n} \qquad (8\text{-}1)$$

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 8-2:

$$\bar{x} - z\sigma_m^2 \le \mu \le \bar{x} + z\sigma_m^2 \qquad (8\text{-}2)$$

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 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 estimates of the wrong answer.

8.2.1. 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 really two related issues that must be decided in this conversation – identification of an acceptable level of sampling precision (e.g., plus or minus 30%, 20% or 10%, or whatever precision is deemed appropriate in the evaluation planning phase)[64] and identification of the desired reliability of the estimate (e.g., 80% 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.

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.

[64] A level of precision that is quite high may be inappropriate for programs that are expected to have smaller impacts either due to the design of the program, or due to the program not yet attaining its target level of participation. If the DR impacts are small, achieving increasing high precision levels may likely to cost more than achieving the same levels of precision for programs with sizeable impacts and a large number of participants.

8.2.2. 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.[65] 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 favor 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.

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.

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 8-3 identifies the formula for determining the sample size required to obtain a given level of precision under simple random sampling:[66]

$$n = \frac{z^2 \sigma^2}{r^2 \bar{x}^2} \qquad (8\text{-}3)$$

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

[65] 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

[66] 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 minimus when the population of interest is large (e.g., more than a few thousand).

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). However, there some negatives associated with 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 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.

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 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.[67] 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 Hodges[68] and Neyman,[69] 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

[67] Ibid.
[68] See "Minimum Variance Stratification" Dalenius T. and Hodges J. L., *Journal of the American Statistical Association*, 1959, 4, pp. 88-101
[69] 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.

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);
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.

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 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 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 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 Tabachanick and Fidell (2001)[70] 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/(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.

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

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. 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.

As explained in Section 4, 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.

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.[71]

[71] 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.

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.

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).

8.3. Conclusion

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 favor the use of these analysis techniques in DR load impact 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*.

9. REPORTING PROTOCOLS (PROTOCOL 26)

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 resource goals and the impacts that have been used to report progress toward goals, and explain 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 impact estimates that are reported;
- Reporting any deviations from the requirements of these protocols and the reasons why it was not possible to meet them;
- Providing recommendations regarding resource modifications and modifications to the impact estimates used for resource progress reports; and
- Providing recommendations concerning future evaluation activities.

Evaluation reports should generally be written for a wide range of individuals, including people who are not familiar with evaluation approaches or the field's specialized terminology. Technical information associated with the 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 wider audience, the more technical details relating to these reporting categories must also be provided.

Protocol 26 outlines in detail the required content of the evaluation reports. Protocols 4 through 25 describe the primary output requirements and formats for the impact estimates developed under these protocols. Table 9-1 contains a template for impact estimates for ex post estimation and Table 9-2 contains a template for ex ante estimates. A separate table must be provided for each of the day types summarized in Table 9-3.

Table 9-1. Reporting Template for Ex Post Impact Estimates*

Hour Ending	Estimated Reference Load (kWh/hr)	Observed Load (kWh/hr)	Estimated Load Impact (kWh/hr)	Temp (F)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
Daily									

*This table is the same as Table 4-1 of the report.

Table 9-2. Output Template for Ex Ante Impact Estimates*

Hour Ending	Estimated Reference Load (kWh/hr)	Estimated Event Day Load (kWh/hr)	Estimated Load Impact (kWh/hr)	Temp (F)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
	Reference Energy Use (kWh)	Observed Energy Use (kWh)	Change in Energy Use (kWh)	Cooling Degree Hours (Base 75)	Uncertainty Adjusted Impact - Percentiles				
					10th	30th	50th	70th	90th
Daily									

*This table is the same as Table 6-1 of the report.

Table 9-3. Day Types to be Reported for Each DR Type*

	Event Based Resources			Non-Event Based Resources		
Day Types	Event Driven Pricing	Direct Load Control	Callable DR	Non-event Driven Pricing	Scheduled DR	Permanent Load Reductions
Ex Post Day Types						
Each Event Day	X	X	X			
Average Event Day	X	X	X			
Average Weekday Each Month				X	X	X
Monthly System Peak Day				X	X	X
Ex Ante Day Types						
Typical Event Day	X	X	X			
Average Weekday Each Month (1-in-2 and 1-in-10 Weather Year)	X	X	X	X	X	X
Monthly System Peak Day (1-in-2 and 1-in-10 Weather Year)	X	X	X	X	X	X

- This table is the same as Table 1-2.

Protocol 26:

Evaluation reports shall include, at a minimum, the following sections:

1. ***Cover***
2. ***Title Page***
3. ***Table of Contents***
4. ***Executive Summary - this section should very briefly present an overview of the evaluation findings and the study's recommendations for changes to the DR resource***
5. ***Introduction and Purpose of the Study - this section should briefly summarize the resource or resources being evaluated and provide an overview of the evaluation objectives and plan, including the research issues that are addressed. It should also provide a summary of the report organization.***
6. ***Description of 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 or tariff, a summary of resource goals (both in terms of enrollment and demand impacts), tables showing reported progress toward goals, projections of future goals and known changes and other information deemed necessary for the reader to obtain a thorough***

understanding of how the resource has evolved over time and what changes lie ahead.

7. ***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.)***

8. ***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.)***

9. ***Detailed Study Findings - this section presents the study findings in detail. (See additional content requirements below.)***

10. ***Recommendations - this section should contain a detailed discussion of any recommended changes to the resource as well as recommendations for future evaluation efforts.***

The Study Methodology section shall include the following:

1. ***Overview of the evaluation plan study methodology;***

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 impact estimates; and the statistics required to be reported in Protocols 9, 10, 16 and 23;***

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 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; and***

6. ***Sampling methodology and sample descriptions (including all 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. All

potential threats to validity given the methodology used must be assessed and discussed. This section should also discuss the evaluator's opinion of how the types and levels of uncertainty affect the study findings. Findings also must 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). The data and statistics outlined in Protocol 24 should be reported in this section.

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 protocols 4 through 10 for ex post evaluation of event based resources, protocols 11 through 16 for non-event based resources, and protocols 17 through 23 for ex ante estimation. 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. Detailed data tables should also be provided in electronic format.***

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), event trigger (e.g., emergency, temperature, etc), start and stop times for the event, event duration in hours, notification lead time, number of customers notified, and number of customers 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 for each forecast year, including but not necessarily limited to, the number of customers enrolled and notified (for event based resources), participant characteristics, weather conditions (if relevant), prices and price elasticities (if relevant), other changes in demand response over time due to persistence related issues and the reasons underlying the changes for the average customer. Information describing the probability distributions for these exogenous variables should be provided whenever such uncertainty is included in the ex ante impact estimates.***

A comparison of impact estimates derived from the analysis and those previously obtained in other studies and those previously used for reporting of impacts toward resource goals, and a detailed explanation of any significant differences in the new impacts and those previously found or used.

10. PROCESS PROTOCOL (PROTOCOL 27)

The protocols include a process protocol that would provide for public review and comment. This will occur at three stages in the evaluation effort.

Protocol 27:

A review and comment process will be used at three stages in the implementation of the Load Impact estimation effort. These stages are:

1. ***The evaluation plan used to develop the research questions to be answered and the corresponding methods to be used to answer them;***
2. ***The interim and draft final reports for all load impact studies conducted for demand response resources; and***
3. ***Review of Final Reports to determine how comments were addressed.***

This process protocol is meant to ensure that the products of each of the two stages in the estimation effort benefits from a public review by stakeholders, Joint Staff, and the CAISO (California Independent System Operating). The Demand Response Measurement Evaluation Committee[72] (DRMEC) would be used to initiate evaluation planning, review the final evaluation plan, and review draft load impact reports.

Two processes are set out below for comments – one for review and comment on the Evaluation Planning effort and a second for the review of interim and draft impact reports.

10.1. Evaluation Planning—Review and Comment Process

The DRMEC will be responsible for working with the utilities (or another identified lead entity) in developing evaluation plans for all statewide or local DR programs that are to have load impacts estimated. The DRMEC will develop a process to determine which demand response programs/activities or tariffs should be evaluated and how frequently meetings should be held. The DRMC is responsible for finalizing the process of deciding which DR programs or tariffs should have impact evaluations within 90 days of this order. The DRMEC will also be responsible for ensuring the issues identified in the evaluation planning sections of the load impact protocols are covered during this planning process. The following actions will be undertaken:

[72] The DRMEC was established by the CPUC in Decision # D.06-11-049 as an informal group charged with developing evaluation plans for demand response resources and reviewing interim evaluations of ongoing demand response programs. Here is an excerpt from that decision: "In D.06-03-024, we authorized the Working Group 2 Measurement and Evaluation subcommittee to continue its work in providing oversight of demand response evaluation, and we continue that authorization for the program augmentations we approve here under the more appropriate name of the Demand Response Measurement and Evaluation Committee. Due to the importance of monitoring and assessing the progress of these programs, the IOUs will provide all data and background information used in monitoring and evaluation projects to Energy Division and the CEC, subject to appropriate confidentiality protections."

1. DRMEC members will identify utility or state staff leads that will be responsible for developing draft evaluation plans for selected projects. The DRMEC will also review draft and final research plans for local utility programs.
2. The DRMEC is to oversee the drafting of the evaluation plans. These drafts should be sent to interested utility program managers and/or evaluators and to the service list (preferably the list established for the review and authorization of DR programs in the last round) or for those who want to participate on the DRMEC for comment.
3. The Utility or DRMEC member responsible for drafting the evaluation plan is responsible for ensuring that comments are solicited from key stakeholders and publishing a small summary of comments received and how or if they were incorporated into the final evaluation plan for each load impact study. The comment period, including responses to them, will be set by the DRMEC taking into account the complexity and length of the documents. Absent good reason, the period for comments on evaluation plans will be 15 business days.
4. The final evaluation plan will be made available to Joint Staff and parties to previous DR proceedings upon request.

10.2. Review of Interim and Draft Load Impact Reports

The utility or contract manager is responsible for facilitating the production of a readable first draft of the load impact report. There may also be interim reports specified in the evaluation plan that will also be subject to a review and comment process. Interim reports may be useful to the impact estimation effort by ensuring interim work products are to be consistent with the protocols. The review and comment process will consist of:

1. The interim or draft load impact report will be sent to both the members of the DRMEC and the service list with a request for comments in at least 5 business days or more, within the time limit determined by the DRMEC. The DRMEC can, at its discretion, choose to meet to discuss the study or conduct the study review by e-mail.

10.3. Review of Final Load Impact Reports

The utility or research manager is responsible for reviewing the comments received and identifying which comments have been incorporated or responded to in the final report.

Copies of the final load impact report should be filed on the CALMAC website and a notice of its availability should be sent out to the service list for the previous demand response rulemaking.

10.4. Resolution of Disputes

Joint Staff (CPUC and CEC) is responsible to resolve any disputes that arise related to evaluation plans or evaluation results. For example, if a party disagrees with a chosen baseline method for evaluation of a particular program, the Joint Staff should have the authority to decide how to resolve it. Elevating these types of technical disputes to the Commission will be too time-

consuming and these technical disputes do not need formal venues such as advice letters for resolution.

(END OF ATTACHMENT A)