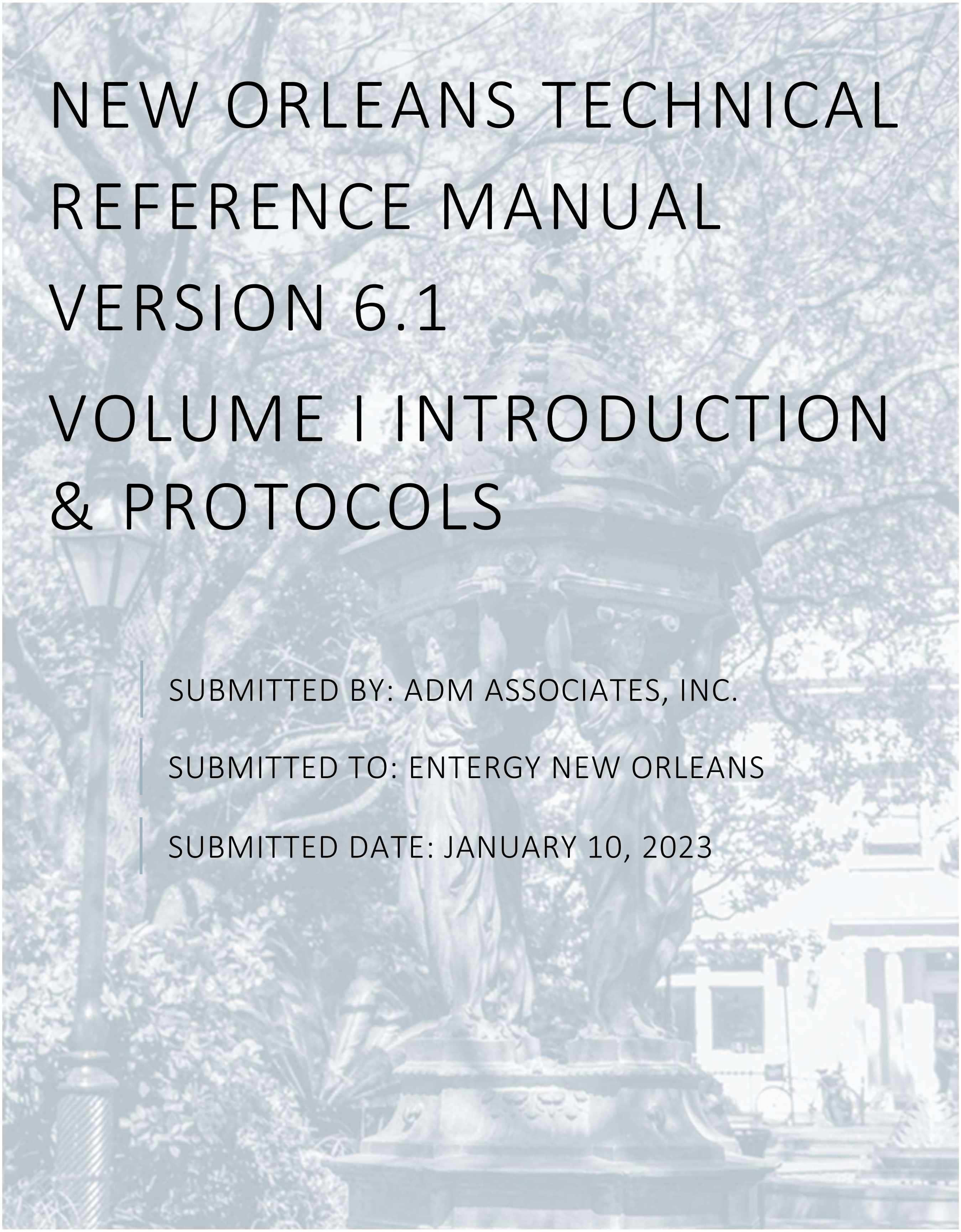

NEW ORLEANS TECHNICAL REFERENCE MANUAL VERSION 6.1

VOLUME I INTRODUCTION & PROTOCOLS

SUBMITTED BY: ADM ASSOCIATES, INC.

SUBMITTED TO: ENTERGY NEW ORLEANS

SUBMITTED DATE: JANUARY 10, 2023

ACKNOWLEDGEMENTS

ADM Associates, Inc. (ADM) would like to acknowledge the many talented individuals who contributed to this sixth version of the Technical Reference Manual for Program Year 13 (2023).

The Entergy New Orleans staff participated in ongoing evaluation deliverable reviews and discussions, attended regular meetings, and responded to follow-up questions, data requests and document requests. They are an ongoing partner in our evaluation efforts.

We also wish to thank the implementation firms: APTIM, Franklin Energy Services, Energy Wise Alliance, Green Coast Enterprises, Honeywell, and their staff for their insights and information.

Additionally, we would like the evaluation staff who supported the creation of this document.

ADM Staff

Richard Haley | Lead Residential Field Staff

John Coleman | Lead C&I Field Engineer I

Joe Marquez | Project Manager

Michael Nicholson, CMVP, CEM | Engineer II

Chris Johnson | Senior Analyst

Melissa Kosla | Senior Analyst

Zephaniah Davis | Senior Manager

Blake Heckendorn, PE | Senior Engineer

Melissa Culbertson | Director

Adam Thomas, PMP | Principal

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ACRONYMS/ABBREVIATIONS

Table 1 Acronyms/Abbreviations

Acronym	Term
AC	Air Conditioner
AOH	Annual operating hours
APS	Advanced Power Strip
AR&R	Appliance Recycling & Replacement
BP	Behavioral Program
BYOT	Bring Your Own Thermostat
C&I	Commercial and Industrial
CEE	Consortium for Energy Efficiency
CF	Coincidence factor
CFL	Compact fluorescent lamp (bulb)
CFM	Cubic feet per minute
CRE	Commercial Real Estate
DI	Direct install
DLC	Direct Load Control
DLC	Design Lights Consortium
EER	Energy efficiency ratio
EFLH	Equivalent full-load hours
EISA	Energy Independence and Security Act
EL	Efficiency loss
EM&V	Evaluation, Measurement, and Verification
ES	ENERGY STAR
EUL	Estimated Useful Life
GPM	Gallons per minute
HDD	Heating degree days
HID	High intensity discharge
HOU	Hours of Use
HP	Heat pump
HPwES	Home Performance with ENERGY STAR
HSPF	Heating seasonal performance factor
HVAC	Heating, Ventilation, and Air Conditioning
IEER	Integrated Energy Efficiency Ratio
IEF	Interactive Effects Factor
IPLV	Integrated part load value
IQW	Income Qualified Weatherization
ISR	In-Service Rate
kW	Kilowatt

Acronym	Term
kWh	Kilowatt-hour
LCDR	Large Commercial Demand Response
LCIS	Large Commercial & Industrial Solutions
LCA	Lifecycle Cost Adjustment
LED	Light Emitting Diode
M&V	Measurement and Verification
MFS	Multifamily Solutions
MW	Megawatt
MWh	Megawatt-hour
NC	New Construction
NTG	Net-to-Gross
PCT	Participant Cost Test
PFI	Publicly Funded Institutions
PY	Program Year
QA	Quality Assurance
QC	Quality Control
RCA	Refrigerant charge adjustment
RIM	Ratepayer Impact Measure
RLA	Retail Lighting and Appliances
ROB	Replace on Burnout
RR	Realization Rate
RUL	Remaining Useful Life
SCDR	Small Commercial Demand Response
SCIS	Small Commercial & Industrial Solutions
SEER	Seasonal Energy Efficiency Ratio
SK&E	School Kits and Education
TA	Trade Ally
TPI	Third-Party Implementer
TPE	Third-Party Evaluator
TRC	Total Resource Cost Test
TRM	Technical Reference Manual
UCT	Utility Cost Test
VFD	Variable Frequency Drive

SAVINGS TYPES

Table 2 Savings Types

Savings Types	**Definition**
Energy Savings (kWh)	The change in energy (kWh) consumption that results directly from program-related actions taken by participants in a program.
Demand Reductions (kW)	The time rate of energy flow. Demand usually refers to electric power measured in kW (equals kWh/h)
Expected / *Ex ante* Gross	The change in energy consumption and/or peak demand that results directly from program-related actions taken by participants in a program, regardless of why they participated.
Verified / *Ex post* Gross	Latin for “from something done afterward” gross savings. The energy and peak demand savings estimates reported by the evaluators after the gross impact evaluation and associated M&V efforts have been completed.
Net / *Ex post* Net	Verified / *ex post* gross savings multiplied by the net-to-gross (NTG) ratio. Changes in energy use that are attributable to a particular program. These changes may implicitly or explicitly include the effects of free-ridership, spillover, and induced market effects.
Annual Savings	Energy and demand savings expressed on an annual basis, or the amount of energy and/or peak demand a measure or program can be expected to save over the course of a typical year. The TRM provides algorithms and assumptions to calculate annual savings and are based on the sum of the annual savings estimates of installed measures or behavior change.
Lifetime Savings	Energy savings expressed in terms of the total expected savings over the useful life of the measure. Typically calculated by multiplying the annual savings of a measure by its EUL. The TRC Test uses savings from the full lifetime of a measure to calculate the cost-effectiveness of programs.

1. TRM PURPOSE AND SCOPE

ADM Associates, Inc. (ADM) is contracted as the Third-Party Evaluator (TPE) for New Orleans Energy Smart Programs administered by Entergy New Orleans (ENO) and their Third-Party Implementer (TPI) team. The purpose of the Technical Reference Manual (TRM) is to provide a single common reference document to estimate energy and peak demand savings from energy efficiency measures promoted by ENO. This document is a compilation of deemed savings values approved by the New Orleans City Council (City Council) and their Advisors for use in estimating savings for energy efficiency measures. The TRM is updated annually through a collaborative process between Stakeholders and the TPE, ADM. The data and methodologies in this document are to be used by program planners, administrators, implementers, and evaluators for forecasting, reporting, and evaluating energy and demand savings from energy efficiency measures installed in New Orleans.

The selection of measures for inclusion in this TRM was based on historical implementation rates of measures; identification of measures in other programs that may warrant inclusion; and an assessment of whether a measure is an appropriate candidate for deemed savings or if it warrants custom analysis. Some viable measures have been excluded from this TRM as they are more appropriate for custom analysis.

1.1 Deemed Savings / Unit Energy Savings (UES)

Deemed savings refers to an approach for estimating average or typical savings for efficiency measures installed in relatively homogenous markets with well-known building characteristics and usage schedules. Previous market research and building simulation tools have been used to develop estimates of “average” or deemed energy or peak savings per measure as a function of building type, capacity, weather, building schedules, and other input variables. Using this approach, program savings can be estimated by multiplying the number of measures installed by the deemed or estimated savings per measure based on previous research on the average operating schedules, baseline efficiencies, and thermal characteristics of buildings in each market.

The deemed savings approach provides reasonably accurate estimates of savings in mass markets where building operating conditions, system characteristics, and baseline efficiencies are relatively well-defined. This approach is not normally used to estimate savings in less homogenous and more site-specific applications, especially in non-residential facilities where the range of operating conditions and energy using processes is significant and can vary widely from one project to another for a similar measure. Developing energy savings estimates for these more complex facilities require the use of one or more of the International Performance Measurement and Verification Protocol (IPMVP) options that require some form of on-site measurement.

Deemed savings estimates require the development of engineering algorithms, tools, or models to estimate average savings as a function of one or more average inputs, including baseline usage, equipment efficiency levels, and building thermal characteristics. This document organizes the methods and sources used to develop these average and default values by measure category and sector and lays out the resulting savings per measure estimates in the form of savings values, algorithms, and/or calculation tools for energy efficiency measures offered by utility program administrators for claiming and reporting energy savings impacts to the City Council.

program-level cost-effectiveness. By establishing clear qualification criteria for the development of projected and claimed savings estimates, the TRM provides transparency of savings for all interested stakeholders.

The TRM document also provides guidance on the update frequency for key inputs and/or equations based on the vintage of the input parameters, as well as the EM&V team's assessment of the level of variability in likely savings estimates across the range of measure applications. The intent is to help participants in the energy efficiency market save money and time by providing a single source to guide savings estimates and equations.

Finally, the EM&V team provides clear criteria for deciding whether future efficient technologies or systems are good candidates to be included in the TRM as a deemed savings measure estimate or a deemed algorithm with stipulated or variable parameters. Changes to the TRM are driven by new technologies, code or baseline changes, a change in high impact measures (HIM) or a request from the implementation team.

The data and algorithms in the TRM are to be used for projecting program savings for the next year and reporting program savings for the previous year. The specific process for updating the TRM and related guidance is discussed in Section 2.10 *Protocol and Guidance for Updating the TRM*.

1.3 TRM Updates Between V5.0 and V6.1

For 2023, the Energy Smart Portfolio is in Program Year 12 (PY12) and the New Orleans TRM (NO TRM) is well-established: All measures offered have TRM sections to support them and all major measures are based in primary New Orleans data collection. While there are two new measures to retrofit/NC measures, the NO TRM Version 6.1 (NO TRM V6.1) focuses on updating existing measures to comply with a variety of new codes and regulations:

Energy Independence Security Act (EISA) – The first of two advances of lighting standards from EISA 2007 Regulations were phased in from January 2012 to January 2014 and dictated higher efficiency for General Service Lamps (GSLs). Phase II (known as the 'EISA backstop') takes effect on July 25, 2022, stipulating that all GSLs sold in the United States (US) must achieve a minimum efficacy of 45 lumens/watt. The ruling also significantly expands the definition of GSLs, extending the covered lumen range, base types, and shapes, while reducing the types of bulbs exempted. This has major implications for lighting nationwide, necessitating a near-complete revision of the Residential Lighting Efficiency chapter, the retirement of four existing residential lighting chapters and minor updates to the commercial lighting chapter and standard wattage table:

- Residential Lighting Efficiency
- ENERGY STAR® Omni-Directional LEDS (retired)
- ENERGY STAR Directional and Specialty LEDS (retired)
- ENERGY STAR Omni-Directional CLFs (retired)
- ENERGY STAR Specialty CFLs (retired)
- Commercial Lighting Efficiency
- Standard Wattage Table Appendix

ENERGY STAR – ENERGY STAR is a government-backed program who provide information on the energy consumption of products and devices using different standardized methods. The standards set by ENERGY STAR for consumer products and buildings are similar to other Department of Energy requirements in that criteria are revised and updated as newer technologies, processes and conditions are introduced. Seven measures will be subject to increased efficiency criteria during 2023, necessitating updates of those chapters:

- Residential Ceiling Fans
- Residential Refrigerators
- Residential Water Coolers
- Commercial Combination Ovens
- Commercial Convection Ovens
- Commercial Dishwashers
- Commercial Solid Door Refrigerators

International Energy Conservation Code (IECC) 2021 – The IECC is referred to as a model energy code because building codes are state or local laws; there is no national building energy code in the US. Regardless of when any state adopts a code, every three years, the IECC is updated to incorporate new building technologies and practices as they evolve over time and ensure that new homes and commercial buildings meet modern-day minimum levels of efficiency. For 2023, New Orleans will be adopting IECC 2021, having previously used IECC 2009. Seven measures in the TRM are affected by the change in code and have been updated accordingly:

- Residential Ceiling Insulation
- Residential Central Air Conditioner Replacement
- Residential Air Source Heat Pump Replacement
- Residential Ground Source Heat Pump Replacement
- Residential Ductless Heat Pump Replacement
- Commercial Packaged Terminal Air Conditioners and Heat Pumps
- Commercial Unitary and Split Air Conditioners and Heat Pumps

Seasonal Energy Efficiency Ratio 2 (SEER2) – The SEER measures air conditioning and heat pump cooling efficiency, which is calculated by the cooling output for a typical cooling season divided by the total electric energy input during the same time frame. The Department of Energy (DOE) is changing the way HVAC systems are tested. New M1 testing procedures are thorough, demanding a lower SEER2 equipment rating (with corresponding analogous test procedures for Energy Efficiency Ratio 2 (EER2) and Heating Season Performance Factor 2 (HSPF2). Effective January 1, 2023, cooling products will be subject to regional minimum efficiencies using the new system and prohibits the sale of units which do not meet the minimum criteria. Further, IECC 2021 uses the new rating system when stipulating new codes. The new rating system and standards required updating six TRM sections:

- Residential Central Air Conditioner Replacement
- Residential Air Source Heat Pump Replacement
- Residential Ground Source Heat Pump Replacement
- Residential Ductless Heat Pump Replacement
- Commercial Packaged Terminal Air Conditioners and Heat Pumps
- Commercial Unitary and Split Air Conditioners and Heat Pumps

Specifics of these updates, as well as additions and other changes to TRM sections are discussed in detail below.

1.4 New and Revised Measures

1.4.1 COMMERCIAL MEASURES

The non-residential updates are listed below.

- Refrigerator and Freezer Case Solid Doors (New) - This measure is retrofitting existing vertical, open, refrigerated display cases by adding and installing doors.
- Water Cooler Timers (New) - This measure involves installing a timer on an existing water cooler to shut down operation during unoccupied hours.
- Window Film (Update) – Savings for film applied to south-facing windows was developed for three heating types and added to existing east and west-facing values.
- Demand Control Ventilation (Update) – Building types and savings values for 'Small Office General' and 'Small Office Densely Occupied' were removed and replaced with 'Small Office (<30,000 ft^2)' and 'Large Office (≥30,000 ft^2) to better reflect the types of office buildings found in New Orleans.
- Variable Speed Drives (Update) – Building types and savings values for 'Small Office General' and 'Small Office Densely Occupied' were removed and replaced with 'Small Office (<30,000 ft^2)' and 'Large Office (≥30,000 ft^2) to better reflect the types of office buildings found in New Orleans.
- ENERGY STAR Combination Ovens (Update) – Updated to reflect ENERGY STAR 3.0 qualification criteria. Deemed savings updated.
- ENERGY STAR Convection Ovens (Update) – Updated to reflect ENERGY STAR 3.0 qualification criteria. Deemed savings updated.
- ENERGY STAR Commercial Dishwashers (Update) – Updated to reflect ENERGY STAR 3.0 qualification criteria. Deemed savings updated.
- ENERGY STAR Solid Door Refrigerators and Freezers (Update) – Updated to reflect ENERGY STAR 5.0 qualification criteria. Deemed savings updated.
- Air Conditioner and Heat Pump Tune-Up (Update) – Air Conditioner and Heat Pump tune-ups can involve multiple steps during the tune-up process, not all of which are necessary for every system. The revised chapter now allows for savings to be assigned to partial tune-ups, based on individual components of the tune-up.
- Packaged Terminal Air Conditioner and Heat Pump – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements. Deemed savings tables updated.
- Unitary and Split System Air Conditioner and Heat Pump – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements. Deemed savings tables updated.

1.4.2 RESIDENTIAL MEASURES

The residential updates are listed below.

- ENERGY STAR Ceiling Fans (Update) – Updated to reflect ENERGY STAR 4.0 qualification criteria. Deemed savings updated.
- ENERGY STAR Refrigerators (Update) – The savings algorithm has been updated and no longer includes the Site/Lab Factor (SLF).
- ENERGY STAR Water Coolers (Update) – Updated to reflect ENERGY STAR 3.0 qualification criteria. Deemed savings updated.

- Window Film (Update) – Eligibility wording has been revised to include south-facing windows.
- Central Air Conditioner Replacement (Update) – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements. Deemed savings tables updated.
- Heat Pump Replacement (Update) – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements. Deemed savings tables updated.
- Ground Source Heat Pump (Update) – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements.
- Ductless Heat Pump (Update) – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements.
- Central Air Conditioner and Heat Pump Tune-Up (Update) – Air Conditioner and Heat Pump tune-ups can involve multiple steps during the tune-up process, not all of which are necessary for every system. The revised chapter now allows for savings to be assigned to partial tune-ups, based on the individual components of the tune-up, and to which certain levels of the component are applied (such as the % of refrigerant charge adjust). Deemed savings values have been calculated and included for applications common in the Energy Smart portfolio.
- Ceiling Insulation (Update) – Baseline and efficiency levels updated to reflect SEER2 and IECC 2021 requirements. Baseline level of insulation updated with past three years' of Energy Smart program data. Deemed savings tables updated. A baseline has been provided for new construction projects.
- Air Infiltration (Update) – A baseline has been provided for new construction projects.

1.4.3 PROTOCOLS

Language in the Demand Response protocol has been changed to state that ENO participants in MISO but does not currently participate in the Demand Response market.

1.4.4 PRIMARY DATA COLLECTION

The following EM&V studies have been completed, allowing for incorporation of primary data into the TRM:

- Metering of residential air conditioning runtime, applied to AC replacement and duct sealing
- Field assessment of average SEER for air conditioning units in duct sealing projects
- Billing analysis to support reductions achieved from residential air conditioning tune-ups
- Measurement of residential domestic hot water (DHW) temperature setpoints, incorporated into DHW replacements and low flow devices
- Metering of residential lighting run-time
- Metering of commercial lighting run-time for the following facility types:
  - K-12 Education
  - Exterior Lighting (all commercial)
  - Food Preparation
  - Food Sales: Non-24 Hour Supermarket
  - Food Service: Fast Food
  - Food Service: Sit-down Restaurant
  - Health Care: In-Patient
  - Lodging: Common Areas
  - Lodging: Guest Rooms

- Multifamily: Common Area
- Religious Assembly/Worship
- Retail: Freestanding
- Warehouse: Non-Refrigerated

The data collected for these studies is summarized in Table 1-1 below.

Table 1-1 Parameters Validated with Primary Data Collection in New Orleans

Parameter	Measures Affected	Value	Sample Size
Residential Cooling Equivalent Full-load Hours	Duct Sealing, AC replacement, AC tune-up	1,637	68 homes
Residential Cooling Peak Coincidence Factor	Duct Sealing, AC replacement, AC tune-up	77%	68 homes
Residential Heating Equivalent Full-load Hours	Duct Sealing, Central AC and Heat Pump Tune-Up, Ductless Heat Pump, Ground Source Heat Pump and Heat Pump Replacement	HP: 396 ER: 600	295 homes
Lighting hours of use	CFLs, Specialty CFLs, Directional LEDs, Omnidirectional LEDs	2.38	40 homes, 355 loggers
Residential Lighting Peak Coincidence Factor	CFLs, Specialty CFLs, Directional LEDs, Omnidirectional LEDs	11.74%	40 homes, 355 loggers
Residential DHW Setpoint (deg. F)	Water Heater Replacement, Faucet Aerators, Low Flow Showerheads	122.24	37 homes
Residential AC Tune-Up Annual % Savings	AC Tune-Up	10.1%	260
Commercial Lighting Hours of Use	Commercial Lighting	Original values created for 10 facility types.	59 premises, 210 loggers
Commercial Lighting Peak Coincidence	Commercial Lighting	Original values created for 10 facility types.	59 premises, 210 loggers
Average Duct Sealing Leakage Reduction	Duct Sealing	SF: 471 MF: 443	SF: 4,939 MF: 325
Deemed Net-to-Gross Ratios	Residential: Duct Sealing Air Sealing AC/HP Tune-Up ENERGY STAR Window AC ENERGY STAR Refrigerator	Varies: Duct Sealing – 95% Air Infiltration – 95% AC/HP Tune-Up – 82% Window AC – 62% Refrigerator – 44%	Varies: Duct Sealing – 282 Air Infiltration – 78 AC/HP Tune-Up – 135 Window AC – 30 Refrigerator – 44
Deemed ISRs for Mailer kit Items	LED lighting, Faucet Aerators and Low-Flow Showerheads	LEDs: 71% Aerators: 45% Showerheads: 62%	4,572 participant responses

Primary data collection has continued during all PY evaluations, including PY12. After PY12 program close, the following data will be analyzed to either develop or refine important savings inputs:

- Results from the Residential PY12 Smart Thermostats M&V analysis will be used to update the Residential Smart Thermostat chapter, which has previously relied on the results of the PY8 pilot analysis.

- Results from the Commercial PY12 Smart Thermostats M&V analysis will be used to update the Commercial Smart Thermostat chapter, which has previously relied on deemed reductions from similar programs administered in other territories.

1.5 Incremental Costs

Incremental costs mean the difference between the cost of the efficient measure and the cost of the most relevant baseline measure that would have been installed (if any) in the absence of the efficiency program. Installation costs (material and labor) shall be included if there is a difference between the efficient measure and the baseline measure.

Note that the TRM includes at least one deemed incremental cost(s) as a default value(s) for most measures. However, consistent with previous versions, in instances where the TPA has better information on the true incremental cost of the measures (e.g., direct install programs), the TPA-specific incremental cost value should be used for the purposes of cost-effectiveness analysis.

Examples of incremental cost calculations include:

- The incremental cost for an efficient measure that is installed in new construction or is being purchased at the time of natural installation, investment, or replacement is the additional cost incurred to purchase an efficient measure over and above the cost of the baseline/standard (i.e., less efficient) measure (including any incremental installation, replacement, or O&M costs if those differ between the efficient measure and baseline measure).
- For a retrofit measure where the efficiency program caused the customer to update their existing equipment, facility, or processes (e.g., air sealing, insulation, tank wrap, controls), where the Customer would not have otherwise made a purchase, the appropriate baseline is zero expenditure, and the incremental cost is the full cost of the new retrofit Measure (including installation costs).
- For the early replacement of functioning equipment with a new efficient measure, where the customer would not have otherwise made a purchase for a number of years, the appropriate baseline is a dual baseline that begins as the existing equipment and shifts to the new standard equipment after the expected remaining useful life of the existing equipment ends. Thus, the incremental cost is the full cost of the new efficient measure (including installation costs) being purchased to replace a still-functioning equipment less the present value of the assumed deferred replacement cost (including installation costs) of replacing the existing equipment with a new baseline measure at the end of the existing equipment's life. This deferred credit may not be necessary when the lifetime of the measure is short, the costs are very low, the measure is highly cost-effective even without the deferred credit, or for other reasons (e.g., certain direct install measures, measures provided in kits to customers).
- For services, such as facility energy audits, energy assessments, and retro-commissioning, the incremental cost is the full cost of the services. Even if the service is performed entirely by a TPA, the full cost of the service charged by the TPA is the incremental cost, because this is assumed to be the cost of the service that would have been incurred by the customer if the customer were to have the service performed in the absence of the efficiency program. In some cases, this will be at the measure level; in others, it will be at the program level. Such costs should be included in measure-level cost-effectiveness calculations only

when they are inseparable from the efficiency improvements – i.e., when the provision of the service is what produces energy savings (e.g., retro-commissioning).

- For the early retirement of functioning equipment before its expected life is over (e.g., appliance recycling programs), the incremental costs are composed of the customer's value placed on their lost appliance, any customer transaction costs, and the pickup and recycling cost. The incremental costs include the actual cost of the pickup and recycling of the equipment because this is assumed to be the cost of recycling the equipment that would have been incurred by the customer if the customer were to recycle the equipment on their own in the absence of the efficiency program. The payment a TPA makes to the customer serves as a proxy for the value the customer places on their lost appliance and any customer transaction costs.

1.6 Simulation Modeling

The savings for some weather sensitive measures were developed via simulation modeling. The model software platforms included are as follows: eQuest©; BEopt™; EnergyGauge USA; and EnergyPlus™.

1.7 Weather

Various measures in the TRM refer to Typical Meteorological Year version 3 (TMY3) weather data. This data is publicly available from the National Renewable Energy Laboratory (NREL) National Solar Radiation Database (NSRDB).

This data reflects the typical year of New Orleans weather based off historical data and is the common practice for projecting average annual savings of weather sensitive measures. Inputs from the TMY3 dataset for New Orleans included the following: temperature; humidity; wind speed and direction; cloud cover; and solar radiation.

1.8 Application of Values in this TRM

It is the intent to have the values in this TRM provide parameters to stipulate ex-post gross energy savings (kWh) and demand reduction (kW) estimates. The values in this TRM do not account for free-ridership, as that is a parameter that may vary based on a program delivery mechanism (for example, the free-ridership rates for residential lighting differ significantly between retail markdown in the Consumer Products versus direct install). The measurement of free-ridership and the application of net-to gross is discussed in detail Section 2.3.7 *Impact Protocol 4.0: Net-to-Gross Analysis*.

The values in this TRM will be used to verify *ex post* gross energy savings (kWh) and demand reductions (kW), except when specified otherwise in the EM&V Plan.

1.9 Future Studies

Each measure section includes a discussion of future studies suggested by the authors of this TRM. For many measures, no studies are recommended, and suggested updates include only updating when codes and standards affecting the specific measure change. The suggestion of future studies is focused on areas of high impact in the Energy Smart portfolio (such as duct sealing) and for the identification of potential future high impact measures (such as ductless mini-split HVAC systems and smart thermostats).

The studies detailed are suggestions on the part of the authors of the TRM and guidance and feedback on these issues is welcomed as part of the stakeholder advisory process.

The general guidelines that are provided for when a study is warranted are as follows (though occasionally subject to modification as specified in a measure-specific chapter):

- Measures should be flagged for further review if they exceed 1% of savings within the residential or non-residential portfolio. In such instances, it should be determined whether:
- Primary data has been collected in Energy Smart evaluations to support the deemed savings;
- The data is sufficiently recent to support its continued use; and
- If data collection to support a deemed savings revision is cost-effective or cost-feasible given the implementation and EM&V budgets for Energy Smart programs.
- Measures that are not over the high-impact threshold should be considered for impact or market assessment studies if:
- Stakeholders (the Council and their Advisors, ENO, Third-Party Administrators (TPA), interveners, the Third-party Evaluator (TPE), and/or other appropriate parties) conclude a measure is of strategic importance to future program implementation efforts; or
- A measure is high-impact within an important market sub-segment (such as low-income, multifamily, or municipal government).

1.10 Overall TRM Layout

This document is divided into separate documents for ease of use:

Volume 1

- Section 1: *TRM Purpose and Scope* covers the process for TRM updates and version rollouts, weather zones, peak demand definitions, TRM structure, and the format of the TRM measure overviews.
- Section 2: *Evaluation Protocols* contains guidance on the application of the TRM that have been reviewed and approved by the EM&V team.

Volume 2

- Section 1: *Residential Measures* contains the measure descriptions and deemed savings estimates and algorithms for measures installed in residential dwellings.
- Section 2: *Non-Residential Measures* contains the measure descriptions and deemed savings estimates and algorithms for measures installed in nonresidential businesses.

Volume 3

- Appendix 1: Engineering Inputs, Methods and Assumptions, and Prototypical Building Characteristics
- Appendix 2: Examples of Existing Baseline Methods for Settlement and Examples of Baseline Adjustments
- Appendix 3: Prior Work in DR M&V Methods
- Appendix 4: Information Sources and References

2. EVALUATION PROTOCOLS

This chapter describes the recommended EM&V Protocols that should be incorporated in process and impact evaluations of the programs pursuant to the Energy Smart Implementation Plan for 2023-2025 (Docket R-20-257 and UD-20-02).

2.1 Protocols Introduction

This section provides protocols for various activities related to performing Evaluation, Measurement, and Verification (EM&V) for the Energy Smart programs that ENO is offering to its residential, commercial, and industrial customers.

The first section introduces the Protocols and explains general principles and concepts, then following sections provide protocols for specific topics:

- 2.3 Protocols for Impact Evaluation
- 2.4 Protocols for Process Evaluations
- 2.5 Protocols for Evaluation of New Construction Projects
- 2.5 Protocols for Evaluation of Retrocommissioning Projects
- 2.7 Protocols for Evaluating Behavioral Programs
- 2.7 Protocols for Evaluating Demand Response Programs and Projects
- 2.8 Protocols and Guidance for Establishing Quality Assurance / Quality Control for Programs
- 2.9 Protocol and Guidance for Updating the TRM

2.2 Description of the Energy Smart Portfolio

Through Energy Smart programs, ENO offers energy efficiency programs to New Orleans residents and businesses. Any residential, commercial, or industrial Entergy New Orleans electric customer is eligible to participate. The NO TRM V6.1 will be applied to the PY13, or 2023 and is active from January 01, 2023, going forward. The table below outlines the portfolio of programs[1] offered in PY13.

[1] Per the filing of Entergy New Orleans, LLC's Energy Smart Program Application for Approval of the Implementation Plan for Program Years 13-15 (Docket Nos. UD-20-02 and UD-08-02). This Plan has not yet been approved; these are subject to change.

Table 2-1 PY13 Portfolio of Programs

Program Name	Sector	Type
Home Performance with ENERGY STAR	Residential	EE
Income Qualified Weatherization	Residential	EE
Multifamily Solutions	Residential	EE
A/C Solutions	Residential	EE
Retail Lighting and Appliances	Residential	EE
School Kits and Education	Residential	EE
Appliance Recycling & Replacement Pilot	Residential	EE
Behavioral	Residential	Behavioral
EasyCool - Bring Your Own Thermostat	Residential	DLC/DR
Peak Time Rebate Pilot	Residential	DLC/DR
Electric Vehicle Bring Your Own Charger Pilot	Residential	DLC/DR
Small C&I Solutions	C&I	EE
Large C&I Solutions	C&I	EE
Publicly Funded Institutions	C&I	EE
C&I Construction Solutions	C&I	EE
Large Commercial Automated Demand Response	C&I	DLC/DR

2.2.1 PURPOSE OF EVALUATION

As defined by the American Evaluation Association, evaluation of an offering involves “assessing the strengths and weaknesses of programs, policies, personnel, products and organizations to improve their effectiveness.” The role of evaluation is two-fold:

- Quantify Results: Document, measure and estimate the energy and demand savings of an offering to determine how well it has achieved its goals and managed its budget.
- Gain Understanding: Determine why certain effects occurred (or didn’t occur) and identify ways to improve and refine current and future offerings; also, to help select future offerings (NAPEE 2007).

Figure 2-1 below provides a visual representation of the role of evaluation activities during the lifecycle of a typical program. As the figure shows, program evaluation should be viewed as an ongoing process that provides information regarding changes in direction and adjustments to goals and objectives over time.

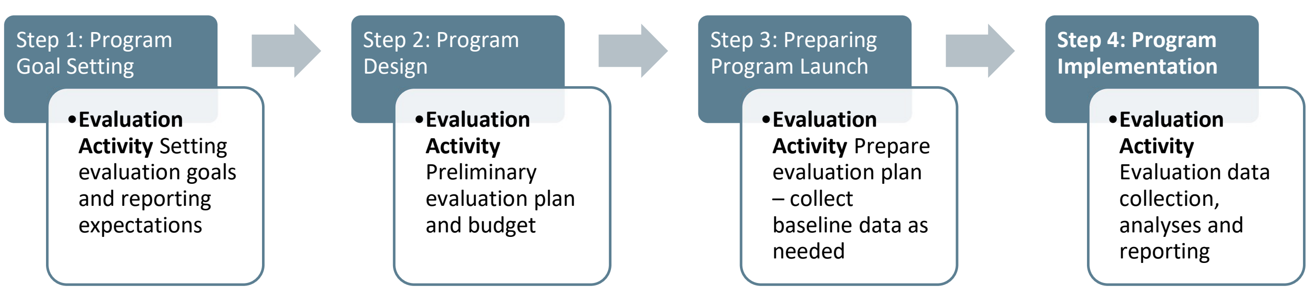

Figure 2-1 High-Level Evaluation Activities in Program Implementation Cycle (NAPEE, 2007)

2.2.2 PURPOSE OF EM&V PROTOCOLS PRESENTED IN THIS VOLUME

The protocols are intended to provide a common framework and set of reference points for conducting cost-effective evaluations. Protocols describe the types of information that must be collected to conduct a comprehensive examination of the overall effectiveness, the recommended frequency for conducting these evaluations, and the key metrics that must be reported during these evaluation activities.

2.3 Protocols for Impact Evaluation

This chapter provides guidance and protocols pertaining to impact evaluation activities for measures and projects that are not included on the list of prescriptive measures for Energy Smart programs. Protocols are presented as follows:

- Impact Protocol 1.0: Impact Evaluation Timing
- Impact Protocol 2.0: Level of Rigor for Impact Evaluations
- Impact Protocol 3.0: Evaluation of Savings for Non-prescriptive Measures or Projects
- Impact Protocol 3.1: Evaluation Approach for 100% Custom Measures
- Impact Protocol 3.2: Impact Evaluation of Non-Prescriptive Measures Whose Savings May Be Treated as Prescriptive
- Impact Protocol 3.3: Impact Evaluation of Information-Based Programs
- Impact Protocol 4.0: Net-to-Gross Analysis

2.3.1 IMPACT PROTOCOL 1.0: IMPACT EVALUATION TIMING

The decision regarding the appropriate time frame for impact evaluation has two components:

- When and over what period of time the evaluation effort will take place?
- What is the level of detail or “granularity” required for the evaluation analyses?

2.3.1.1 Evaluation Occurs When and over What Period of Time?

A standard evaluation begins before program implementation begins to collect important baseline data and then continues for some time after the program is completed to analyze persistence of savings and other program elements.

The actual timing of evaluation efforts influenced by several factors, including:

- What will be the period of analyses (i.e., how many years)?
- Will persistence of savings be determined, and if so, how?
- What is the timing for policy decisions and evaluation planning?
- What is the need for early feedback for program implementers?
- Where is the program in its lifecycle?
- What are the evaluation data collection time lags?
- What are the other regulatory and/or management oversight requirements to be addressed in this evaluation?
- What information or data are needed to update specific energy and demand savings from the measure, and to quantify life estimates?

- What is the timing and format required for the reporting process? Is a single, final program report needed, or are more frequent reports required?

In general, program evaluations are conducted with a three-year plan. Process evaluations are usually conducted at the end of the first year of program operations and at the conclusion of the program period. Impact evaluations may be conducted annually or at the conclusion of Program Years 2 and 3, and generally free ridership and spillover no more frequently than once every three years provided there are sufficient data to determine energy savings estimates and adjustments and no significant changes in a program design. The timing for the EM&V activities should be specified in EM&V plans for the programs to be evaluated.

2.3.1.2 What Level of Detail is Required?

This relates to whether 15-minute, hourly, monthly, seasonal, or annual data collection and savings reporting are necessary. The granularity decision is based how the information will be used for evaluation purposes. Annual savings data provide an overview of program benefits. More detailed data are usually required for both cost-effectiveness analyses and resource planning purposes.

If demand savings are to be calculated, the choice of definition (e.g., annual average, peak summer, coincident peak, etc.) is related to time granularity. When evaluating energy or demand savings, it is important to properly define the project boundaries (i.e., what equipment, systems, or facilities will be included in the analyses). Ideally, all primary effects (the intended savings) and secondary effects (unintended positive or negative effects), and all direct (at the project site) and indirect (at other sites) will be captured in the evaluation. The decision concerns whether savings will be evaluated for specific pieces of equipment. For example, the "boundary" may include motor savings or light bulb savings estimates, the end-use system (e.g., the HVAC system or the lighting system), the entire facility, or the entire energy supply and distribution system (Modified NAPEE, 2007).

The EM&V plan for each program should stipulate the confidence and precision levels necessary to provide for a robust EM&V analysis of the savings estimates and describe the sampling strategy that will be used. Sampling strategies will vary by program and across the program portfolio. The sampling strategy for a particular program should therefore be fully described in the EM&V plan for that program.

2.3.2 IMPACT PROTOCOL 2.0: LEVEL OF RIGOR FOR IMPACT EVALUATIONS

Impact evaluation of gross savings can be performed under different levels of rigor, depending on available evaluation resources, uncertainty in expected savings, magnitude of expected savings, program budget, and other criteria.

The level of effort necessary to assess savings impacts is driven by the equipment type and data collection needs. The International Performance Measurement and Verification Protocol (IPMVP) is an important and widely used guidance document that provides guidelines about the "level of effort" required to document energy efficiency savings. The IPMVP presents various M&V options, summarized in Table 2-2, that help guide savings verification methods and levels of effort.

Table 2-2 IPMVP M&V Options

IPMVP Option	Measure Performance Characteristics	Data Required
Option A: Engineering calculations using spot or short-term measurements and/or historical data	Constant performance	Verified installation Nameplate or stipulated performance parameters Spot measurements Runtime hour measurements
Option B: Engineering calculations using metered data	Constant or variable performance	Verified installation Nameplate or stipulated performance parameters End-use metered data
Option C: Analysis of utility meter (or sub-meter) data using techniques from simple comparison to multivariate regression analysis	Variable performance	Verified installation Utility metered or end-use metered data Engineering estimate of savings input into SAE model
Option D: Calibrated energy simulation / modeling; calibrated with hourly or monthly utility billing data and / or end-use metering	Variable performance	Verified installation Sport measurements, runtime hour monitoring, and/or end-use metering to prepare inputs into models Utility billing records, end-use metering, or other indices to calibrate modeling

In the California Energy Efficiency Evaluation Protocols, IPMVP M&V options are used to identify two levels of rigor for evaluation of gross energy savings.

- Basic rigor level, which is consistent with IPMVP Option A (or, in some cases, Option C).
- Enhanced rigor level, which is consistent with IPMVP Options B or D (or, in some cases, Option A).

The levels of rigor for evaluating impacts of a program can be assigned by using this correspondence between IPMVP M&V options and levels of rigor by determining which IPMVP option should be applied to assess savings for measures or projects in a program. For example, Lawrence Berkeley National Laboratory (LBNL) maintains a webpage on its Measurement & Verification portal that allows use of interactive tools to identify the IPMVP option that is best suited to evaluating savings for a particular project. (See http://mnv.lbl.gov/interactive/ipmvp-1a-2.) This tool can be used to assign an IPMVP Option and corresponding level of rigor (basic, enhanced) to measures or projects included in a program.

The LBNL application (which is adapted from IPMVP 2012 Volume 1) identifies an appropriate M&V option based on responses to questions about the energy conservation measure/project that's being considered for evaluation. Items of information needed include the following:

- Claimed kWh / kW
- Count of measures in the projects
- Count of installed measures
- Descriptions of any equipment changed or of new equipment installed
- Interactive effects between measures
- Percentage of savings vs. baseline

For Energy Smart programs, there are prescriptive and non-prescriptive measures. Prescriptive measures are explicitly listed as such in program materials. Non-prescriptive measures are those that are not included on the list of prescriptive measures for the Energy Smart programs. Within the set of non-prescriptive measures, a distinction can be made between 100% custom measures and measures where deemed calculation methods might be used but data need to be collected or developed to be put into the calculation algorithms.

This distinction is shown in Table 2-3. For Prescriptive Measures that are included on the list of prescriptive measures, savings are deemed. (These deemed savings values are provided in the current Technical Reference Manual.) Protocols for assessing savings for Non-Prescriptive and 100% Custom Measures are discussed in Section 2.3.4 Impact Protocol 3.1: Custom Measures Evaluation.

Table 2-3 Spectrum of Measures: 100% Prescriptive to 100% Custom[2]

Types	100% Prescriptive	Non-Prescriptive		100% Custom
	Exclusive Source	**Primary Source**	**Used as a Source**	**May be used as a Source**
Deemed Calculation(s)	No	Yes	Yes	No, unless custom measure EM&V protocols are included
Deemed Variables or Factors	No	Mix of site-/project-specific and deemed data	None or minimal	None or minimal
Site- or Project-Specific Variables or Factors	No	Mix of site-/project-specific and deemed data	Exclusively or mostly	Exclusively or mostly
Deemed Savings Values	Fully deemed savings values	Partially deemed savings values	No, savings determined per deemed calculations, resulting in site/ project specific savings	No, savings determined per project/measure analyses and data collection, resulting in site/project specific savings
EM&V Method	Deemed savings	M&V Option A	M&V Option B, C, or D	M&V Options B, C, or D (e.g., for individual building projects) or control group methods (e.g., for residential projects)

2.3.3 IMPACT PROTOCOL 3.0: EVALUATION OF NON-PRESCRIPTIVE MEASURES

As discussed in Section 2.3.2 Impact Protocol 2.0: Level of Rigor for Impact Evaluations, levels of rigor with which savings for non-prescriptive and custom measures are assessed are determined depends on the methods chosen for the analysis of savings. Protocols pertaining to the choice of methods are presented in this section. In general, documentation information is used to determine (1) what methods of savings analysis to use and (2) specifications of assumptions and sources for these specifications. Protocols are provided for the following:

- 100% custom measures
- Non-prescriptive measures that are not 100% custom
- Measures promoted through mass market programs

[2] Carroll 2013, as adapted and presented in SEE Action Guide for States: Guidance on Establishing and Maintaining Technical Reference Manuals for Energy Efficiency Measures

2.3.4 IMPACT PROTOCOL 3.1: CUSTOM MEASURES EVALUATION

Types of measures that can be 100% custom include (1) measures or projects that site-specific but that are considered too complex or unique to be included in the list of standard measures provided in the TRM or (2) measures that may involve metered data, but that require additional assumptions to arrive at a 'typical' level of savings as opposed to an exact measurement.

Most measures in this category are custom measures installed in both retrofit and new construction situations in C&I facilities. In general, these custom measures are more complex measures that require site-specific information and detailed calculations to estimate energy and demand savings. These measures do not comply with a prescriptive calculation approach or may benefit from having more detailed savings analysis.

Because custom measures are often unique, their savings are evaluated using a site-specific M&V approach, with more reliance placed on using site-specific engineering analysis and end-use metering as methods to estimate savings. The site-specific approach involves (1) selecting a representative sample of custom projects or measures that participated in a program; (2) determining the savings for each project or measure in the sample, usually by using one or more of M&V Options defined in the IPMVP; and (3) applying the results of estimating the savings for the sample projects or measures to the entire population in the program. Further information on the EM&V methods recommended for custom measures in provided in Table 2-4. Methods to determine gross savings for 100% custom measures depend on the type of measure and the end use affected (e.g., lighting, HVAC, industrial process).

Table 2-4 Summary of Recommended EM&V Methods for 100% Custom Measures[3]

Characteristic	Approach	Additional Comments
Program Tracking	Initial gross estimates of energy and demand savings and initial net impacts as applicable. Measure description with, as applicable, unit quantities, sizes/ capacities, baseline and installed efficiencies, and operating hours.	Any additional parameters that could be useful for quality control or for evaluation design, such as sampling that are described in the EM&V plan.
Recommended M&V Method	On-site inspections with partial (Option A) or complete (Options B,C,D) measurements on a census or sample of program participants. Site visits with short-term metering is the most appropriate approach for C&I Custom measures. A detailed engineering spreadsheet model can be used to capture the dynamics and interactions on an hourly basis. Data collected from Energy Management Systems (EMS) may also provide cost-effective information and should be included in EM&V plans if available.	Metering methods often include time-of-use loggers, interval kW recorders, and spot power measurements.
Alternative M&V Method	If the Custom measure involves significant HVAC equipment and/or controls, calibrated simulation modeling (Option D) offers a viable alternative for capturing measure dynamics and interaction.	Metering can be used to calibrate the model. Such metering may include whole premise interval kW recorders with some end-use metering.

[3] NEEP EM&V Protocols, 2010

Evaluating savings impacts for custom measures or projects requires that baseline conditions be defined. The baseline reflects the conditions, including energy consumption, that were occurring before the installation of the measure. Baseline definitions consist of site-specific issues and broader, policy-oriented considerations.

- Site-specific issues include the characteristics of equipment in place before an efficiency measure is implemented as well as how and when the affected equipment or systems are operated. When defining the baseline, it is also important to consider where in the life cycle of the existing equipment or systems the new equipment was installed. The options are:
    - Early replacement of equipment that had not reached the end of its useful life;
    - Failed equipment replacement, with new energy efficient equipment installed; or
    - New construction.

For each option, there are two generic approaches to defining baselines.

- Project-Specific Baseline. With the project-specific procedure (used with all or a sample of the projects in a program), the baseline is defined by a specific technology or practice that would have been pursued, at the site of individual projects, if the program had not been implemented. For energy efficiency programs, the baseline is established by:
    - Assessing the existing equipment's energy consumption rate, based on measurements or historical data;
    - Completing an inventory of pre-retrofit equipment; or
    - Comparing to a control group's energy equipment (used where no standard exists or when the project is an "early replacement," i.e., implemented prior to equipment failure).

  The most widely accepted method, and recommended for these EM&V Protocols, is to define the baseline by determining what technologies the new equipment replaces. That is, the baseline is related to actual historical base year energy consumption or demand and carried forward to future years (NAPEE, 2007).

- Performance Standard Baseline. For the Performance Standard Baseline approach, a performance standard is developed that provides an estimate of baseline energy and demand for all the projects in a program. The assumption is that any project activity will produce additional savings if it has a "lower" baseline than the performance standard baseline. Performance standards are sometimes referred to as "multi-project baselines" because they can be used to estimate baseline savings for multiple project activities of the same type.

  Under the performance standard procedure, baseline energy and demand are estimated by calculating an average (or better-than-average) consumption rate (or efficiency) for a blend of alternative technologies or practices. These standards are used in large-scale retrofit (early replacement) programs when the range of equipment being replaced and how it is operated cannot be individually determined. This would be the case, for example, in a residential compact fluorescent lamp (CFL) incentive program, where the types of lamps being replaced and the number of hours they operate cannot be determined for each home. Instead, studies are used to determine typical conditions. Another common use of performance standards is to define a baseline as the minimum efficiency standard for a piece of equipment as defined by a law, code, or standard industry practice. This is commonly used for new construction or equipment that replaces failed equipment (NAPEE, 2007).

This approach is especially important when it is difficult to determine baselines, such as in new construction programs since no comparison period exists. However, the concepts of project and performance standard baseline definitions can still be used in these circumstances. The industry-accepted methods of defining new construction baselines are based on:

- Specifications of buildings that would have been built or equipment installed, without the influence of the program, at the specific site of each construction project. This might be evaluated by standard practice evaluation or building plans and specifications that were prepared prior to the program being launched.
- Existing building codes and/or equipment standards; and
- Performance of equipment, buildings, etc., in a comparison group of similar program non-participants.

Because custom projects or measures are usually site-specific, site visits are generally required to collect appropriate information to analyze savings. This includes collecting information on the quantity, sizing, servicing, and scheduling for HVAC, lighting, refrigeration, motors, process, and other equipment. Information may also be collected on the capabilities of control systems (e.g., whether centralized or distributed, capabilities for control monitoring, automation possibilities, and expansion possibilities).

2.3.5 IMPACT PROTOCOL 3.2: IMPACT EVALUATION OF NON-PRESCRIPTIVE MEASURES WHOSE SAVINGS MAY BE TREATED AS PRESCRIPTIVE

Energy Smart programs may include non-prescriptive measures that are not 100% custom measures. Savings for these measures are not deemed. However, savings can be assessed using savings calculation algorithms with stipulated and “open variables”. Examples of open variables include the following:

- Capacity of an AC unit
- Change in connected load
- Square footage of insulation
- Hours of operation of a facility or of a specific electric end-use
- Horsepower of a fan or pump motor

Essentially, the savings calculation algorithms can be considered deemed, but the algorithms require customer-specific input for open variables to calculate the energy and demand savings. With customer-specific information used for open variables, savings values for the same measure can differ across customers.

Information on open variables can be collected from program participants or through site visits. For some open variables, a default value may have to be used when data for the open variable cannot be collected. For example, an average value can be provided that can be considered the default value for input to the algorithm and that can be used when customer-specific information is not available.

Some issues that should be considered in evaluating savings for non-prescriptive measures include the following.

- Algorithms and definitions of terms should be reviewed to verify that accepted industry standards are being used to reasonably estimate savings. This review should be used to ensure that the deemed methodologies for calculating savings are clearly defined and can be implemented practically and effectively.
- High-impact measures should be identified for review and clarifications or modifications.
- Low-impact measures with unrealistic and inaccurate savings values should be reviewed. This review can be done periodically to adjust the level of EM&V rigor based on market adoption.
- For nonresidential measures, consider establishing energy impact thresholds by measure type in the TRM, above which customer-specific data collection is required for open variables. The intent of this is to reduce the overall uncertainty of portfolio savings estimates by increasing the accuracy of project-level savings estimates for extremely high-impact measure installations.
- When to use default values for open variables in the *ex ante* and/or *ex post* savings calculations should be determined considering the savings impact and the uncertainty associated with the measure. Default values for open variables can be used if customer-specific or program-specific information is unreliable or cannot be easily obtained. The default values are appropriate for low-impact and low-uncertainty measures (e.g., lighting retrofits in a small business facility). In contrast, customer-specific values are appropriate for high-impact and high-uncertainty measures, (e.g., HVAC or lighting retrofits in universities or hospitals that have diverse facilities) and where those types of projects represent a significant share of program savings for a year.
- For key open variables where default values are provided that are based on evaluations completed in other jurisdictions or taken from industry or other associations, the literature supporting use of the default values should be reviewed and assessed. This may include reviewing the population from which source data were used for deriving the default values and providing recommendations as to what populations or technologies the derived default values can be applied.

Because customer-specific data for open variables are collected and used to estimate savings, there will be a variety of savings values for the same measure. Customer-specific or program-specific data for the *ex ante* and/or *ex post* savings calculations should be used for as many open variables as possible to improve the accuracy of estimated savings. Site-specific data or information should be used for measures with important variations in one or more input values (e.g., delta watts, efficiency level, equipment capacity, operating hours). Customer-specific data can come directly from measure application forms, be collected during the application process, or collected through site visits.

To guide the customer-specific data collection, measures can be grouped into various end-use categories (e.g., lighting, HVAC, motors & VFDs) and kWh savings thresholds established for each end-use category level that can be used to determine whether customer-specific information should be used for estimating *ex ante* and/or *ex post* savings. If a project involves multiple measures or types of technology that fall under the same end-use category, the savings for all those measures/technology types should be grouped together to determine if the project falls below or above a particular threshold.

2.3.6 IMPACT PROTOCOL 3.3: IMPACT EVALUATION OF INFORMATION-BASED PROGRAMS

Through the Energy Scorecard program, ENO provides information to customers that they can use to adjust their use of electricity. The protocol provided here is intended to give guidance on evaluating the impacts of this and other information-based programs that might be used to provide information to customers.

There are several evaluation approaches that can be used to determine the savings impacts of an information-based program that provides customers with information that they can use to voluntarily take actions to adjust their energy. The approaches differ in their ability to produce accurate and robust results and are therefore discussed in descending order of desirability. Because of differences in performance, Option 1 is the preferred approach. Option 2 should be used only when Option 1 is infeasible. Option 3 should only be used when both Option 1 and Option 2 are infeasible.

If available, interval meter data should be used to estimate load impacts. Where advanced metering infrastructure (AMI) data is not available for all participants, estimates based on a sample of metered homes may be used.

The three options for estimation of impacts from information-based programs are as follows.

- Option 1 uses an analysis based on an experimental design that makes appropriate use of random assignment so that the reference load is estimated using a representative control group of program participants. The most common type of design satisfying this criterion is a randomized control trial (RCT), but other designs may also be used. An evaluation contractor can select a specific design, based on their professional experience.
- Option 2 uses a comparison group analysis where the loads of a group of non-participating customers that are similar to participating homes with respect to observable characteristics (e.g. electricity consumption) are used to estimate the reference load. Because there is a variety of matching techniques that are available, an evaluation contractor can choose the technique used to select the comparison group based on their professional judgment. Difference-in-differences estimators should be used in the analysis to control for any differences that may remain after matching.
- Option 3 is a ‘within-subjects’ analysis where the reference energy use of participating customers is estimated using data on their energy use during a period before their participation in the information-based program began.

The analysis for all three options can be accomplished through regression analysis that relates energy use to weather conditions (particularly temperature) and other variables that influence usage. Panel regression modeling is the recommended technique.

2.3.7 IMPACT PROTOCOL 4.0: NET-TO-GROSS ANALYSIS

NTG analysis is directed at quantifying those savings attributable to a program. This protocol presents general definitions and methods that can be employed as part of a sound NTG analysis.

There are five approaches commonly used for determining NTG.

- Self-Reporting Surveys: From participants and non-participants without independent verification;

- Enhanced Self-Reporting Surveys: Self-reporting surveys are combined with interviews and independent documentation review and analysis. They may also include analysis of market-based sales data;
- Econometric Methods: Statistical models are used to compare participant and non-participant energy and demand patterns. These models often include survey inputs and other non-program-related factors such as weather and energy costs (rates);
- Deemed: NTG is estimated using information available from evaluation of similar programs; and
- Stipulated: The stipulation of NTG may be used when the expense and uncertainty of the results are considered significant barriers (NAPEE, 2007). Use of stipulated values is not recommended if they yield results that are uncertain and/or costly; instead, the Protocol will support the usage of literature reviews.

These approaches for assessing the energy savings attributable to a program are based on determining NTGRs that have two main components: free ridership and spillover.

- Free ridership refers to program participants who received an incentive but would have installed the same efficiency measure on their own had the program not been offered. This includes partial free riders, defined as customers who, at some point, would have installed the measure anyway, but the program persuaded them to install it sooner or customers who would have installed the measure anyway, but the program persuaded them to install more efficient equipment and/or more equipment. For the purposes of EM&V activities, participants who would have installed the equipment within one year will be considered full free riders; participants who would have installed the equipment later than one year will not be considered to be free riders (thus no partial free riders will be allowed).
  Free ridership is the share of gross program savings that is generally the savings accounted for in program records and then adjusted for the naturally occurring adoption; the free ridership rate is based on actions participants "would have taken anyway" (i.e., actions that were not induced by the program). Each energy efficiency program covers a range of energy efficiency measures and is designed to move the overall market for energy efficiency forward. However, it is likely that some participants would have wanted to install some high efficiency measures (possibly a subset of those installed under the program) even if they had not participated in the program or been influenced by the program in any way.
- Spillover refers to energy savings that are due to the influence of a program but are not counted in program records. For example, a customer installs a set of efficiency measures in one of his/her buildings. These measures were promoted (and incented) under a DSM program. The customer then decides to install the same measures at another site, where there is no program incentive. In this case, the program had an influence on the market beyond the energy savings in this customer's first building. Spillover can be broken out in three categories:
  - *Participant Internal Spillover* represents energy savings from additional measures implemented by participants at participating sites not included in the program but directly attributable to the influence of the program.
  - *Participant External Spillover* represents energy savings from measures taken by participants at non-participating sites not included in the program but directly attributable to the influence of the program.
  - *Non-Participant Spillover* represents energy savings from measures that were taken by non-participating customers but are directly attributable to the influence of the program.

Spillover adds to a program's measured savings by incorporating indirect (i.e., not incented) savings and effects that the program has had on the market above and beyond the directly incented or directly induced program measures.

Total spillover is a combination of several factors that may influence non-reported actions to be taken at the project site itself (inside spillover) or at other sites by the participating customer (outside spillover). Each type of spillover is meant to capture a different aspect of the energy savings caused by the program, but not included in program records. Because a primary goal of most DSM programs is to transform markets through a variety of strategies – including education, promotion, and increasing awareness of the benefits of energy efficiency – one would expect spillover to occur to some extent in the market.

The overall NTG is meant to account for both the net savings at participating projects and spillover savings that result from the program (but are not included in program records). When the gross program savings multiplies the NTG ratio, the result is an estimate of energy savings that are attributable to the program (i.e., savings that would not have occurred without the program). The basic equation is:

$$NTG = 1 - \text{Free ridership} + \text{Spillover}$$

The underlying concept inherent in the application of the NTG formula is that only savings caused by the program should be included in the final net program savings estimate, but this estimate should include all savings caused by the program (i.e., the net program savings should account for free ridership and include spillover).

2.3.7.1 Estimating Free Ridership: Survey Techniques

Data to assess free ridership should be gathered through a series of survey questions asked of end-use customers and trade allies who participated in the program. Free ridership can be evaluated by asking direct questions, aimed at obtaining respondent estimates of the appropriate free ridership rate that should be applied to them, and by supporting, or influencing questions used to verify whether the direct responses are consistent with participants' views of the program's influence.

The direct free ridership questions ask respondents to estimate the share of measures that would have been incorporated at high efficiency if not for the technical and financial assistance of the program. The questions also ask respondents to estimate the likelihood that they would have incorporated measures "of the same high level of efficiency" if not for the technical and financial assistance of the program. This flexibility in how respondents conceptualize and convey their views on free ridership will allow respondents to provide their most informed response, thus improving the accuracy of the free-ridership estimates.

The "program influence" questions clarify the role that program interventions (e.g., financial incentives and technical assistance) played in decision-making and provide supporting information on free ridership. Responses to these questions are analyzed for each respondent and used to identify whether the direct responses on free ridership are consistent with how each respondent rated the "influence" of the program.

These results will then be compared to free ridership estimates based on on-site inspections/audits, and/or estimates derived from similar surveys completed in other jurisdictions.

2.3.7.2 Estimating Spillover: Survey Techniques

The basic method for assessing participant (inside and outside) spillover employs a three-step approach to determine the following:

1. Whether spillover exists at all. These are yes/no questions that ask, for example, whether the respondent incorporated energy efficiency measures or designs that were not recorded in program records. Questions relate to extra measures installed at the project site (inside spillover) and to measures installed in non-program projects (outside spillover).
2. Extent of the spillover. These questions request information about the number or share of projects/jobs/facilities into which additional measures or technologies are installed (these questions are not asked for inside spillover because the value is simply the one project on which the interviewee focuses).
3. Amount of savings per spillover project. These questions ask respondents to estimate the energy savings associated with the non-recorded measures relative to the savings from the participating project itself.

The outcome of these inquiries is an estimate of the share of those non-recorded savings that can be attributed to the influence of the program.

2.3.7.3 Timing of Data Collection for Free Ridership vs. Spillover

Where possible, a staggered data collection approach should be used to collect information in support of NTG analysis. The rationale for this approach is that free ridership and spillover data are best collected at different points in time.

Free ridership data are most accurate when collected as closely as possible to the point in time when the participation decision is made. Doing so helps to ensure accurate participant recall of motivating factors and relative program influence while also producing other benefits, including near-term feedback for program staff regarding program influence effects.

Conversely, spillover data are considered most accurate when collected sometime after the participating project has been completed. Allowing a reasonable amount of time to pass before asking participants about spillover effects ensures that participants have sufficient time to: a) install the incented equipment, b) experience its operating parameters and costs, and c) then decide whether to install additional energy efficiency measures at the project site or some other location independent of any program support or financial incentive (Johnson et al., 2010).

2.3.7.4 Hierarchical Approaches for Determining When to Update NTG Values

A decision tree with several steps can be used to determine the timing for updating attribution analysis. The framework for updating net savings follows the hierarchical approach presented visually in Figure 2-2 Each step in the decision is discussed in the following paragraphs.

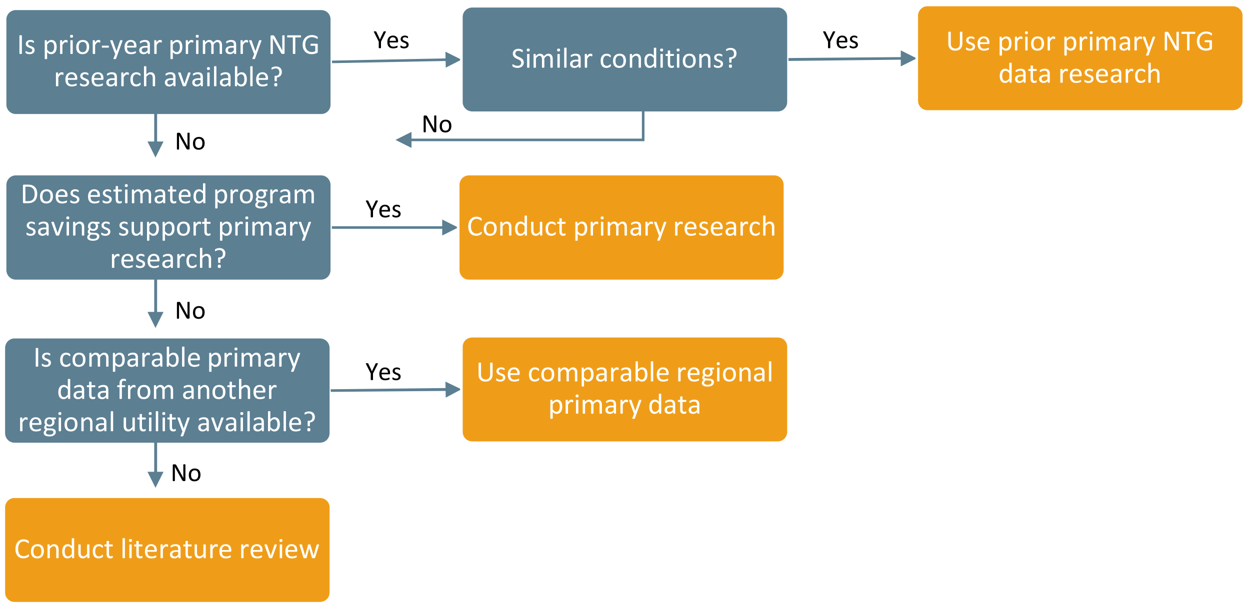

Figure 2-2 Decision Tree for Timing and Selection of NTG Research

Has NTG research been conducted on the same program in a prior year? The first step to determining whether primary NTG research should be conducted in a given program year is to assess whether primary data collected for the same program are available from a prior year. If prior data are available, it should be determined whether the prior values are applicable in the current year. There are at least two overarching components of this decision.

- First, determine if the current program is similar to the program in which the primary data was collected: Is the mix of measures the same? Is the contribution to savings for each measure similar? Are the incentive levels comparable? Is (are) the delivery method(s) similar?
- Second, determine if the market conditions are similar to the time period in which the prior data were collected: Has there been a substantial change in incremental cost for the efficient measures? Has there been a substantial change in the supply or availability of the efficient measures? Has there been a substantial change in the market share of efficient measures (i.e., the ratio of efficient measures sold to total comparable standard and efficiency measures)? Are the local or federal codes and standards the same as when the prior NTG values were estimated?

If the program and market conditions are comparable to the time period(s) in which the prior primary NTG research was conducted, these prior values can be considered applicable to the current program year.

- If prior year primary data are not available or are determined not to be applicable due to changes to either program or market conditions. The TPE should then determine whether estimated savings from the program support primary research. In general, programs that represent at least 5-10 percent of the portfolio estimated savings in any given year should use NTG ratios that are estimated via primary data research for that specific program.
- If prior year data for the program are not available or applicable, and the program savings does not support primary data collection. The evaluation should then consider if NTG values derived from Arkansas-

based comparable programs are available. A comparable program is defined as one that is similar in terms of program maturity, incentive levels, delivery mechanism, and measure types. Ideally, NTG values derived in the same program year would be used, but values from prior years may also be used if the comparability conditions are met.

- For existing and new programs that do not meet any of the above specifications. A literature review may be undertaken to locate a similar program (or programs) that has (or have) an established NTG value(s). This approach requires that the research be well documented. A program may be identified as similar if it meets the following conditions:
- Program similarity: maturity, incentive levels, delivery mechanism, and measure types are similar; and
- Market similarity: demographic, household, and business characteristics are similar (or as similar as possible) to those for New Orleans.

With this hierarchical approach, evaluation resources can be directed towards programs that could benefit most from primary research, thus avoiding unnecessarily repeating NTG research every year for the same programs. However, to prevent NTG values from being repeated too many years and becoming potentially “stale”, NTG values for programs that meet the contribution to savings threshold should be updated at least via primary research at least once during every three-year program cycle.

The steps along this decision tree should be clearly presented and discussed as part of program evaluation plans.

Evaluations using trade ally responses should be collected for programs where the trade allies play a key role in the installation decision. The evaluation work plan should present a discussion of the representation from the trade ally respondents. If use of information supplied by trade allies is applicable, evaluation plans should include details regarding how trade ally responses will be integrated with customer survey responses to determine overall program attribution.

EM&V reports should include robust reporting related to NTG research, methods, and findings. To ensure consistency and transparency, the report should include the following information regarding NTG analyses.

- Summary of each programs NTG source. For example, a table could show which programs received updated NTG research versus those where NTG analysis used previous values, deemed values, or secondary research.
- Discussion of rationale for use of previous estimate or literature review. EM&V reports should cite evidence that the delivery, incentives, measures, and program design were unchanged.
- If unique NTG values are assigned to distinct program components, then each component should be reported with gross and net savings contributions. Where different program components (e.g., measures) have different NTG values, savings for each program component should be presented along with the respective NTG values.

It is recommended that an appendix be included in the report that details NTG approach and methods. This appendix should include the following:

- High-level discussion of approach and methods. A methods section should detail the overarching NTG approach across programs, especially if the same algorithms and logic are used across multiple programs.

- Detailed discussion of logic (including questions, full battery of survey question). Complete survey battery logic, flow-charts, and comprehensive details of the program NTG approach should be included in the appendix.
- Discussion of program-specific logic in each section. If individual program NTG research includes customized logic that is distinct from the overall approach included in the methods section, then the differences in approach should be reported within each individual program section.

2.4 Protocols for Process Evaluations

This protocol provides guidance regarding scope and timing for process evaluation of a programs. A process evaluation involves examining the process of implementing a program and determining whether the program is operating as planned. The goal of a process evaluation is to recommend ways to improve processes to increase a program's effectiveness. A process evaluation focus on determining the overall effectiveness of program delivery, identifying opportunities for program improvements, and assessing key program metrics, including participation rates, market barriers, and overall program operations.

2.4.1 PROCESS PROTOCOL 1.0: PROCESS EVALUATION DETERMINATION

Two major criteria can be applied to determine if a process evaluation of a program is needed.

- The first criterion is to determine if it is time for a process evaluation;
- The second criterion is to determine if there is a need for a process evaluation.

Table 2-5 addresses the first criterion, setting out conditions for determining what timing is appropriate for conducting a process evaluation.

Table 2-5 Determining Appropriate Timing to Conduct a Process Evaluation

1. No Previous Process Evaluation: If a program has not had a comprehensive process evaluation, conducting a process evaluation should be considered.
2.New and Innovative Components: If a program has new or innovative components that have not been evaluated previously, then a process evaluation should be considered for assessing their level of success in the current program and their applicability for use in other programs.
3. New Vendor or Contractor: If a program is a continuing or ongoing program but is now being implemented, in whole or in part, by a different vendor than in the previous program cycle, then a process evaluation should be considered to determine if the new vendor is effectively implementing the program.
If any of these criteria are met, it is time to conduct a process evaluation.
If none of these criteria are met, proceed to Table 2-6 (Step 2) in the Process Evaluation Decision Map.

Figure 2-3 provides a flow chart for determining whether it is time to perform a process evaluation of a program.

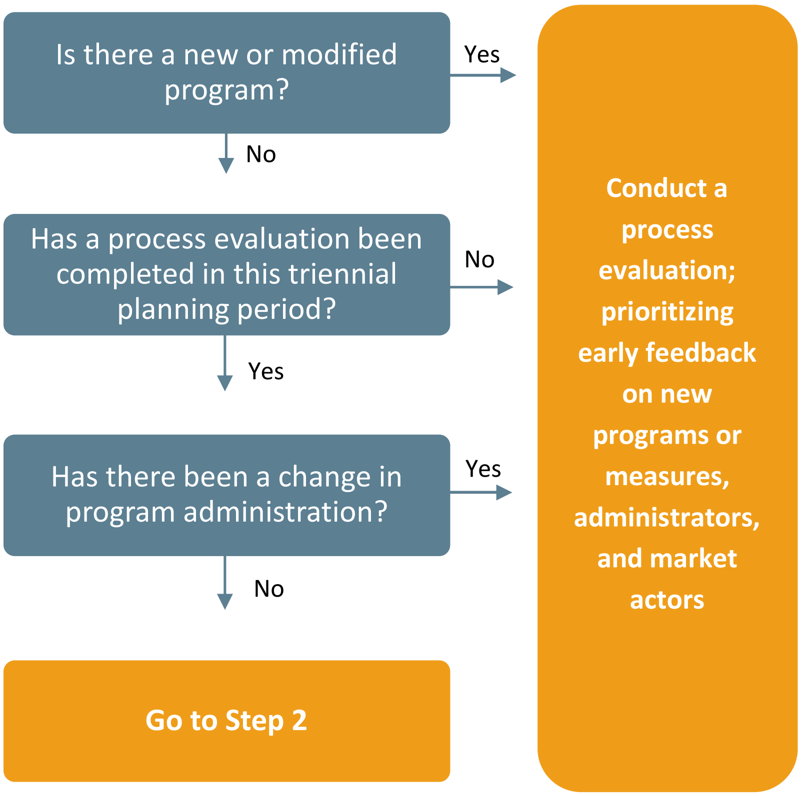

Figure 2-3 Determining Timing for a Process Evaluation

Process evaluations may be used to diagnose areas where a program is not performing as expected. Conditions to consider are outlined in the table below.

Table 2-6 Determining Appropriate Conditions to Conduct a Process Evaluation

Conditions appropriate to conducting a process evaluation may include the following:
1. Impact Problems: Are program impacts lower or slower than expected?
2. Informational/Educational Objectives: Are the educational or informational goals not meeting program goals?
3. Participation Barriers: Are the participation rates lower or slower than expected?
4. Operational Challenges: Are the program's operational or management structure slow to get up and running or not meeting program administrative needs?
5. Cost-Effectiveness: Is the program's cost-effectiveness less than expected?
6. Negative Feedback: Do participants report problems with the program or low rates of satisfaction?
7. Market Effects: Is the program producing the intended market effects?
If any of the criteria is met, a process evaluation is needed to identify ways to address and correct these operational issues.
If none of these criteria is met in either Step 1 or Step 2, then a process evaluation is not needed at this time.
Re-evaluate the need for a process evaluation at the end of the program year.

Figure 2-4 defines the method to identifying problems in program administration that may warrant a process evaluation.

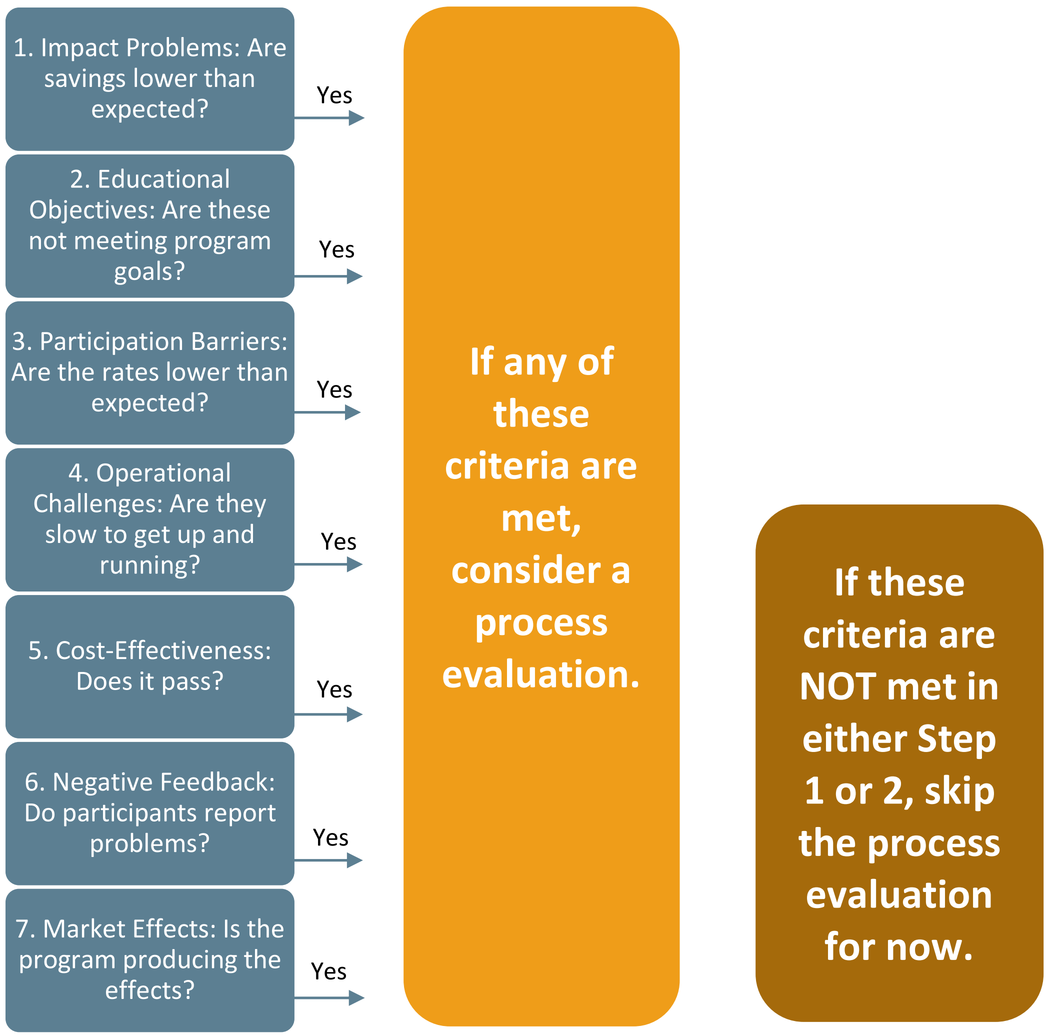

Figure 2-4 Determining Need to Conduct a Process Evaluation

Throughout an evaluation cycle, limited or focused process evaluation activities (e.g., review of program database, interviews of staff) may be used to determine interim progress for a program. Findings from focused process evaluation activities serve several purposes.

- Provide a progress report for each recommendation for program improvement made in previously conducted evaluations. For each evaluation recommendation, the report should indicate whether the recommendation has been accepted and implemented, rejected, or is still under consideration. If the recommendation is rejected, an explanation of the reason for rejection should be provided. If a recommendation is still under consideration, then an explanation should be provided for the steps underway to reach an implementation decision for that recommendation, which should include:
- Identify progress made towards achieving program objectives; and
- Identify any issues that may need to be explored more fully in future program evaluations.

2.4.2 PROCESS PROTOCOL 2.0: PLANNING PROCESS EVALUATION

This protocol provides guidance on the key issues that should be addressed in planning process evaluation activities. Aspects of program operations to address any deficiencies identified in Figure 2-4.

Three tables are provided that outline the key researchable issues that should be addressed in a process evaluation.

- Table 2-7 provides a general outline of the key elements that should be included in a process evaluation plan.
- Table 2-8 provides more detailed information regarding the key areas for investigation that need to be addressed in a process evaluation.
- Table 2-9 identifies those topic areas that should be covered in all process evaluations, those areas that should be investigated when the program is experiencing specific operational issues or challenges, and those areas that are most applicable to new programs or pilot programs.

Table 2-7 Recommended Elements of a Process Evaluation Plan[4]

Element	Description
Introduction	Description of the program or portfolio under investigation; specific characteristics of the energy organization providing the program including current marketing, educational or outreach activities and delivery channels
Process Evaluation Methodology	Process evaluation objectives, researchable issues, and a description of how specific evaluation tactics will address the key researchable issues including the proposed sampling methodology for program/third-party staff, key stakeholders, trade allies/vendors, and customers. The sampling methodology should be clearly explained with specific targets of completed surveys or interviews clearly described in the EM&V Plan.
Timeline	Summarized by key tasks identifying the length of the process evaluation and key dates for completion of major milestones
Budget	Costs of conducting the process evaluation by specific tasks and deliverables

[4] California Evaluation Protocols, 2006

Table 2-8 Recommended Areas of Investigation in a Process Evaluation

Program Design	**Additional Guidance**
▪ Program mission, vision and goal setting and goal setting process ▪ Assessment or development of program and market operations theories ▪ Program design and design characteristics, and program design process ▪ Use of new or best practices	▪ This area is especially important to address in first-year evaluations and evaluations of pilot programs.
Program Administration	**Additional Guidance**
▪ Program management process ▪ Program staffing allocation and requirements ▪ Management and staff skill and training needs ▪ Program tracking information and information support systems ▪ Reporting and relationship between effective tracking and management, including operational and financial management	▪ This area should be covered in all process evaluations, but it is especially important to address in those evaluations where operational or administrative deficiencies exist.
Program Implementation and Delivery	**Additional Guidance**
▪ Description and assessment of program implementation and delivery process ▪ Program marketing, outreaching, and targeting activities ▪ Quality control methods or operational issues ▪ Program management and management's operational practices ▪ Program delivery systems, components, and implementation practices ▪ Program targeting, marketing and outreach efforts ▪ Program goal attainment and goal-associated implementation processes and results ▪ Program timing, timelines, and time-sensitive accomplishments	▪ This is critical to gathering the information necessary to assess the program's operational flow. ▪ These are areas that should be addressed if program is not meeting its participation goals or if the program is under-performing. ▪ All marketing and outreach materials should be reviewed and assessed as part of document review task. ▪ These areas should be addressed in all process evaluations but are especially important if the program is under- performing regarding savings or participation rates.

Table 2-9 Recommended Areas of Investigation in a Process Evaluation

Areas of Investigation	Additional Guidance
▪ Documentation of program tracking methods and reporting formats	▪ This is a key element of the review of the program database and the TPE should request copies of the program records or extracts along with the data dictionary.
▪ Customer interaction and satisfaction (both overall satisfaction and satisfaction with key program components, including satisfaction with key customer- product-provider relationships and support services) ▪ Customer or participant's energy efficiency or load reduction needs and ability of program to deliver on those needs ▪ Market allies' interaction and satisfaction with program ▪ Reasons for low level of market effects and spillover ▪ Intended or unanticipated market effects	▪ These topics should be investigated in customer surveys and should be a priority if the program is experiencing negative feedback or lower-than-expected participation rates or energy savings.

2.4.3 PROCESS PROTOCOL 3.0: PROCESS EVALUATION REPORT AND RECOMMENDATIONS

The suggested reporting requirements for a process evaluation report are given in Table 2-10..

Table 2-10 Suggested Reporting Requirements for Process Evaluation Report[5]

Suggested Reporting Requirement	Description
1. Detailed Program Description	Process evaluation report should present a detailed operational description of the program that focuses on program components being evaluated. The use of a program flow model is highly recommended. Report should provide sufficient detail so that a reader can understand program operations and likely results of recommended program changes.
2. Program Theory	Process evaluation report should include a presentation of the program theory. If the program theory is not available or cannot be provided in time for the evaluation report due date, a summary program theory built from the evaluation team's program knowledge may be included instead. However, it should be complete enough for a reader to understand the context for program recommendations. It does not need to be a finely detailed program theory or logic model.
3. Support for Recommended Program Changes	All recommendations need to be adequately supported. Each recommendation should be included in the Executive Summary and then presented in the Findings text along with the analysis conducted and the theoretical basis for making the recommendation. The Findings section should also include a description of how the recommendation is expected to help the program, including the expected effect that implementing the change will have on the operations of the program.
4. Detailed Presentation of Findings	A detailed presentation of the findings from the study is essential. The presentation should convey the conditions of the program being evaluated and should provide enough detail so that any reader can understand the findings and the implications of the overall operations of the program and its cost-effectiveness

Table 2-11 provides guidance on structuring recommendations from a process evaluation.

[5] CA Evaluation Protocols, 2006

Table 2-11 Suggested Structuring of Recommendations from Process Evaluation[6]

Requirements for Recommendations from a Process Evaluation
▪ Realistic, appropriate to Entergy New Orleans' structure, constructive, and achievable using available resources ▪ Linked to specific conclusions ▪ Adequately supported. Each recommendation should be included in the Executive Summary of the report and then presented in the findings text along with analysis conducted and theoretical basis for making recommendation. Findings section should include a description of how recommendation is expected to help the program, including the expected effect implementing the change will have on the operations of the program. ▪ Focused on ways to increase overall program effectiveness and be linked to researchable issues addressed in process evaluation (e.g., ways to improve program design, approach, operations, marketing, or address issues related to program under-performance) ▪ Providing specific steps / tasks for implementation (to extent possible) ▪ Compared across program evaluations to identify areas for portfolio-level improvements

[6] CA Evaluation Protocols, 2006

2.5 Protocols for Evaluation of New Construction Projects

2.5.1 DESCRIPTION

This protocol is intended to describe the recommended method when evaluating the whole building performance of new construction projects in the commercial sector. The protocol focuses on energy conservation measures (ECMs) or packages of measures where evaluators can analyze impacts using building simulation. These ECMs typically require the use of calibrated building simulations under Option D of the International Performance Measurement and Verification Protocol (IPMVP).[7]

Examples of such measures include Leadership in Energy & Environmental Design (LEED) building certification, novel and/or efficient heating, ventilation, and air conditioning (HVAC) system designs, and extensive building controls systems. In general, it is best to evaluate any ECM expected to significantly interact with other systems within the building and with savings sensitive to seasonal variations in weather.[8] The protocol classifies commercial new construction projects as:

- Newly constructed buildings: The design and construction of an entirely new structure on a greenfield site or wholesale replacement of a structure torn down to the ground.
- Addition (expansion) to existing buildings: Significant extensions to an existing structure that requires building permits and triggers compliance with current codes.
- Major renovations or tenant improvements of existing buildings: Significant reconstruction or “gut rehab” of an existing structure that requires building permits and triggers compliance with current codes.

Evaluators may need to apply the evaluation methods described here for new construction projects for some projects in the retrofit programs. While some retrofit projects have much in common with new construction projects, their scope does not uniformly fall under the new construction categories previously described. Evaluators should assess these projects according to the guidelines described for retrofit equipment (described in separate protocols).

EM&V of new construction programs involves unique challenges, particularly when defining baseline energy performance. An agreed-upon building energy code or industry standard defines the baseline equipment evaluators use to measure energy impacts for new construction measures. As the baseline equipment for new construction measures does not physically exist and cannot be measured or monitored, evaluators typically employ a simulation approach. Due to the nuances involved in appropriately determining baseline equipment/performance evaluations, experienced professionals with a good understanding of building construction practices, simulation code limitations, and the relevant building codes should oversee these types of projects.

Further, evaluators typically assess new construction measures within the first few years of construction. During this period, there is often considerable change in building occupancy and operation before the measures design

---

[7] As discussed in the section “Considering Resource Constraints” of the Introduction chapter to this report, small utilities (as defined under U.S. Small Business Administration regulations) may face additional constraints in undertaking this protocol. Therefore, alternative methodologies should be considered for such utilities.

[8] Note the term whole-building modeling does not necessitate use of sophisticated stand-alone simulation software (e.g., eQuest, EnergyPlus). It is acceptable to employ engineering models using spreadsheet calculations, provided they meet the guidelines set forth in Section 4

intent becomes realized. This results in additional challenges for evaluators using monitored data and/or facility utility billing or energy consumption history to define as-built building performance.

2.5.2 APPLICATION CONDITIONS OF PROTOCOL

Use the algorithms and protocols described here to evaluate new construction whole-building performance ECM installed in commercial facilities. When new construction ECM do not directly impact HVAC energy use, it is often possible to use spot measurements and engineering calculations to evaluate savings with sufficient rigor (ASHRAE, 2002). This is usually the case, for example, with lighting and domestic hot water retrofits.[9] This protocol does not cover the guidelines for selecting the appropriate M&V rigor for such measures. Consult the IPMVP or measure-specific protocols within the Uniform Methods Project protocols to review evaluation guidelines for measures that do not require calibrated building simulation.

2.5.2.1 Incentive Types

Program administrators typically classify new construction program incentives as being either component-based or performance-based and design the program to offer one or both types of incentives.

2.5.2.1.1 Component-Based Incentives

Component-based (or "prescriptive") incentives tend to involve individual technologies and equipment. Examples of prescriptive incentives may include lighting fixtures, occupancy sensors, motors, and small packaged (unitary) HVAC units. Evaluators often determine rebate amounts and claimed savings estimates based on stipulated per-unit estimates.[10] Evaluators will sometimes assess component-based rebates according to measure-specific protocols using partial or complete retrofit isolation evaluation strategies (IPMVP Option A or Option B).

2.5.2.1.2 Performance-Based Incentives

Performance-based incentives tend to target more complex projects involving improvements to the overall building energy performance. Whole-building performance incentives can:

- Encompass various specific (above-code) upgrades
- Fund design, analysis, equipment, and/or installation (labor) costs.[11]

An example of a performance-based project is LEED certification. Buildings that are LEED certified often encompass ECM that range from envelope improvements to high-efficiency equipment installations (often going beyond just HVAC) and complicated controls algorithms.

The complex interactions between these ECM can only be reliably determined through the use of calibrated building simulation models.

[9] While the general magnitude of the secondary impacts imparted by lighting measures on HVAC equipment are well-established for various building types, take care to estimate these impacts appropriately in new construction building stock. New buildings typically have more efficient HVAC equipment, which reduces the magnitude of heating and cooling interactive effects. Secondary impacts can be estimated using prototypical building models, representative of the physical facility. See the Uniform Method Project's Chapter 2: Commercial and Industrial Lighting Evaluation Protocol or CPUC 2004 for guidelines regarding HVAC interactive factors.

[10] Units used do not necessarily represent quantity. Frequently applied units include installed horsepower, tons of refrigeration, and square footage.

[11] Some new construction programs have been successfully implemented without direct financial incentives (e.g., design assistance, financing, etc.).

Performance-based incentive amounts are typically determined by the expected annual energy and/or demand impacts (e.g., per kilowatt-hour, therm, kilowatt).12 Annual energy-savings estimates for performance-based projects (and programs) require evaluators to use custom calculations via whole-building simulation modeling tools. Therefore, highly skilled technical labor is required to successfully implement and evaluate these programs.[13]

2.5.3 SAVINGS CALCULATIONS

Use the following algorithm to calculate energy savings for new construction measures. Note that evaluators can calculate demand savings using the same algorithms by simply substituting "demand" for "energy use"[14].

$$Energy\ Savings = Projected\ Baseline\ Energy\ Use - Post\ Construction\ Energy\ Use$$

Where:

$Projected\ Baseline\ Energy\ Use =$ Projected energy use of baseline system at full designed occupancy and typical building operating conditions.

$Post\ Construction\ Energy\ Use =$ Energy use of measure systems at full design occupancy and typical building operating conditions.

As described in Section 4, Measurement and Verification Plan, calculate projected baseline energy use and post-construction energy use using a whole-building simulation model that is calibrated to monthly (or hourly) utility energy consumption histories. Evaluators can use four components to report savings for the new construction ECM:

- Expected (planned) measure savings
- Rebated measure savings
- Non-rebated measure savings
- Total achieved savings

Section 4 discusses each component.

2.5.4 MEASUREMENT AND VERIFICATION PLAN

2.5.4.1 IPMVP Option

The preferred approach to calculate savings for whole-building performance new construction projects is calibrated building simulation models according to IPMVP Option D (IPMVP, 2006). The recommended approach requires sufficient resources be allocated to the project to allow for detailed onsite data collection, preparation of the simulation models, and careful calibration. The method is less costly when a functioning ex-ante model is available to the evaluator, though obtaining the ex-ante model is not a prerequisite to its application.

Determine the appropriate modeling software by the specifics of the evaluated buildings (e.g., HVAC system and zoning complexity, building constructions, complexity of the ECM); there is no single software (currently

[12] Depending on program design, the "expected" energy impacts can be either *ex ante* or *ex post*.

[13] Johnson & Nadel, 2000

[14] When calculating the coincident peak demand savings, average the hourly demand savings over the "peak demand window" period, as defined by the utility.

available) that can simulate all variations of HVAC system type, building construction, and ECM. Thus, it may be necessary to use multiple tools to evaluate building performance accurately.

In general, the appropriate software for modeling building systems and energy performance must[15]:

- Create outputs that comply with American National Standards Institute (ANSI)/ASHRAE Standard 140-2011[16]
- Accurately simulate the building's systems and controls
- Use an hourly or sub-hourly time step to perform simulation[17]
- Simulate building performance using user-defined weather data at hourly intervals

The DOE Energy Efficiency and Renewable Energy website[18] contains a list of building energy simulation software. Although some tools listed are proprietary, the website also lists public-domain DOE-sponsored tools. Summary comparisons and descriptions of commonly used software can be found in Crawley, 2005.

The preferred full Option D approach will in some cases be intractable due to limited data availability or evaluation budgetary limitations. In such cases, alternate methodologies are acceptable, but the following guidelines should be followed:

- Onsite verification and review of as-built drawings and commissioning reports (as available) should be performed to verify which energy saving features were installed and are functioning
- *Ex ante* savings calculations should be based in a whole building simulation model of the building or of a building that is representative of the actual facility
- Results should be compared with billing data (when available), engineering rules of thumb, and/or secondary literature to review reasonability

2.5.4.1.1 Verification Process

Figure 2-1 depicts the overall process to verify savings under Option D, from the California Evaluation Framework (CPUC, 2004). The process starts by specifying which site data collection and equipment monitoring requirements are in an M&V plan. Additionally, the M&V plan should specify:

- The applicable version of the building codes and equipment standards that determine the baseline (or applicable 'practice' that may determine baseline); discussed in 2.5.4.3 *Baseline Considerations*
- The above-code technologies present in the building (claimed as ECMs)
- The software for modeling building performance
- Appropriate data for calibrating the simulations
- How to address modeling uncertainties
- Against what statistical indices calibration will be measured.

[15] For more information on specific requirements for simulation software, see pp. 133 in The California Evaluation Framework (CPUC, 2004) and pp. 26-27 in Appendix J – Quality Assurance for Statistical, Engineering, and Self-Report for Estimating DSM Program Impacts (CADMAC, 1998).

[16] ANSI/ASHRAE Standard 140-2011 establishes test procedures validating software used to evaluate thermal performance of buildings (and applicable HVAC equipment).

[17] It is preferable the software use unique time steps for each interval (e.g., 8,760 hours).

[18] This website can be found here: https://www.energy.gov/eere/office-energy-efficiency-renewable-energy

While reviewing the energy consumption data can be useful in developing data collection needs, it is not a prerequisite to creating and implementing the M&V plan. However, when developing the M&V plan, evaluators should consider how long a building has been occupied because that will determine amount and granularity of energy consumption data available. Fewer months of consumption data, or the availability of only monthly data, usually means there will be a greater emphasis on metering specific pieces of equipment. Conversely, the presence of a building automation system, energy monitoring system, lighting control panels, (collectively referred to here as building automation system) or other devices to control and/or store data about the operational characteristics of the building will allow for a lesser dependence upon utility usage data.

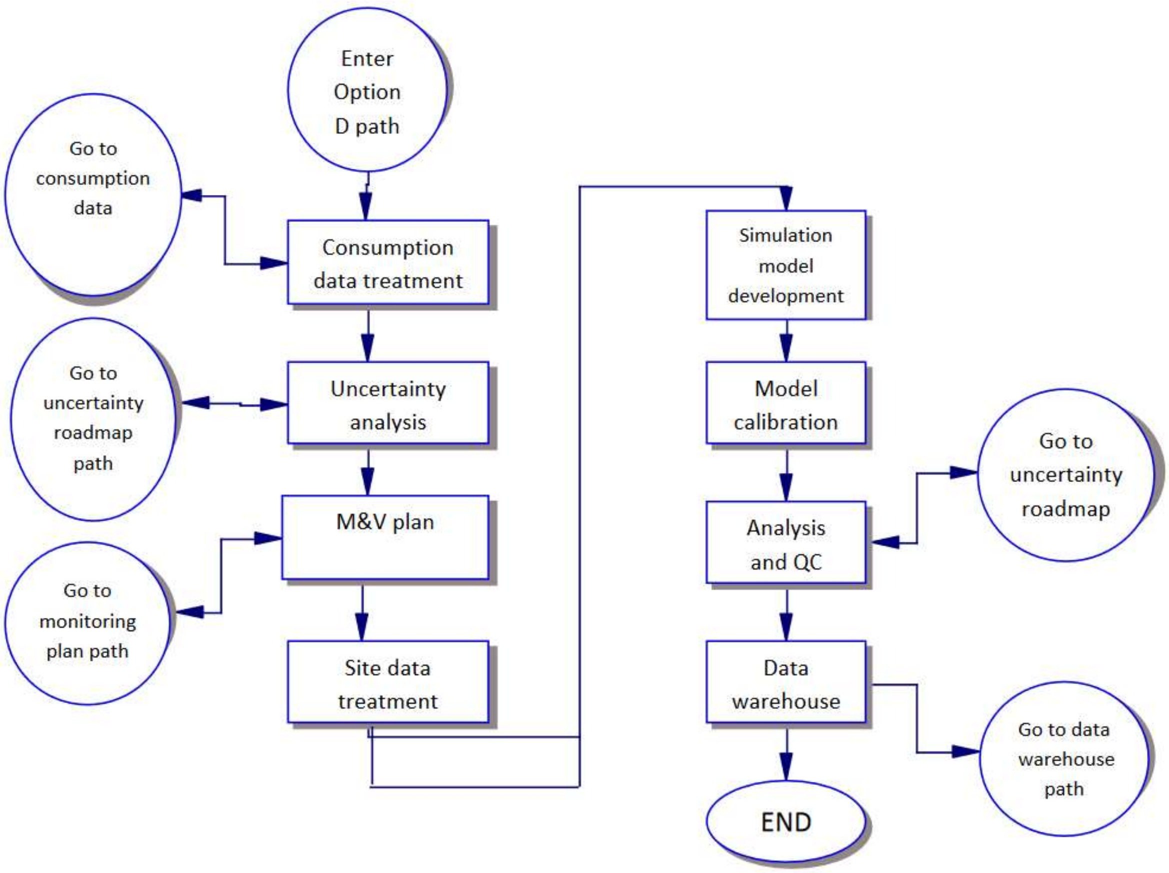

Figure 2-5 Roadmap for IPMVP Option D

2.5.4.1.2 Data Requirements and Collection Methods

Data collected during this step includes all of the information required to define and calibrate the building simulation model. Due to the unique nature of each new construction project, it is impractical to prescribe a comprehensive list of specific parameters evaluators should collect on site. Instead, use the following guidelines to identify key data points and minimize the uncertainty in the final calibrated simulations. After identifying specific parameters, refer to the Uniform Methods Project's (UMP) *Metering Cross-Cutting Protocols* for instructions regarding the methods to submeter the physical parameters. The data used to define building simulation models come from stipulated and physical sources. Furthermore, these data can be static or dynamic in nature, as described here:

- Static data points. These are essentially constant values that describe physical properties of the equipment and the building surfaces or the set point and operational range controlling the building equipment.[19] Examples of static data points are window glazing, motor efficiencies, and thermostat set points.
- Dynamic data. These are time-dependent variables that describe building and equipment operations. These data capture the behavioral and operational details (e.g., weather, motor loading, and building occupancy) needed to establish a building's energy-use characteristics. Dynamic data, which are often the most difficult to collect, represent the greatest source of uncertainty in a building simulation.

IPMVP Option D (IPMVP, 2006) allows use of stipulated data, although it is important to minimize the number of these inputs, as they represent degrees of freedom (and, therefore, additional uncertainty) in the model. Sources for such data include peer-reviewed research, engineering references, simulation program defaults, manufacturers' specifications, and/or survey information from on-site visits (e.g., mechanical, and architectural drawings and visual inspection of nameplate information).

The following are convenient categories of important physical data to collect on site (ASHRAE, 2002):

- Lighting systems
- Plug loads
- HVAC systems
- Building envelope and thermal mass
- Building occupants
- Other major energy-using loads[20]

Another important element of the data collection process entails the use of submetering to define behavioral and dynamic aspects of a building and its subsystems. In this protocol, the term submetering encompasses both direct placement of monitoring equipment by evaluation personnel and collecting data from the building automation systems (also known as trend data) when available. Even when the absolute accuracy of the collected data is unknown, sub-metered data is useful for informing operational schedules (e.g., lighting and ventilation) and calibrating the model.

The degree of submetering required is largely dependent upon the quality and resolution of the facility's energy consumption history. The following descriptions of submetering represent the minimum amount of data collected for calibrating simulation models. Additional submetering may be necessary to verify complex control schemes and/or set points. Perform additional submetering as budget and time permit. Use such data to inform model inputs rather than to function as a calibration target.

[19] 3 set points can refer to a control zone, thermostat, control valve, flow rate, voltage, photocell, or other parameter that is designed to maintain optimal environmental conditions within the building. Some set points are "dynamic" in that they may change according to the time of day.

[20] This category is particularly important in buildings such as grocery stores, refrigerated warehouses, and some retail.

(i) Submetering With Monthly Bills

When only a monthly utility billing history is available for a facility, it is important to submeter both HVAC fan schedules[21] and interior lighting fixtures. Also, if the facility has unique or considerable equipment loads (e.g., data centers), meter these as well.

When monitoring unitary HVAC equipment, isolate the power used by fans from that used by compressors. This ensures evaluators can use the resulting data when calibrating time-of-use and magnitude of fan power.

If, due to site or budget limitations, the electrical monitoring must comprise the unitary system as a whole, use motor nameplate information and fan curves in conjunction with local weather data to disaggregate the fan and compressor power.[22]

Alternatively, use one-time power measurements to establish a unit's demand for each operation mode. Combine these measurements with time-series data to identify time spent in each operation mode and, thereby, determine the fan schedules.

(ii) Submetering With Hourly Bills

Hourly (or sub-hourly) energy consumption histories contain much more information for model calibration than monthly usage alone. While this additional information reduces submetering requirements, it does not eliminate the need to submeter HVAC fan schedules as they are important for disaggregating base loads from ventilation. As described for monthly billing data, consider submetering other large energy-using features (e.g., pool-heating and space-cooling equipment, atria lighting, and internet technology loads) if possible given evaluation budgets.

2.5.4.2 Simulation Model Development

It is important to model several iterations of the simulated building to fully capture the various aspects of the savings for new construction ECM. Table 2-12 lists this iterative process, which entails three versions of the as-built building and two versions of the baseline building, including:

- As-built physical
- As-built design
- As-built expected design
- Whole-building reference
- Measure building reference

Table 2-12 does not include intermediate modeling of individual ECM. Intermediate modeling can be used to disaggregate individual measure impacts and interactive effects. If measure-level savings estimates (and therefore, intermediate modeling of measures) are required, work with the governing jurisdiction for the evaluation process to establish an appropriate hierarchy to govern the order in which measures are stacked and individual measure savings assessed.

---

[21] It is important to capture a building's ventilation schedule when HVAC systems are used to supply outside air to maintain required fresh requirements. If performing submetering on a sample of HVAC fans, place priority on accurately capturing when (and how much) outside air is introduced into the building.

[22] To employ this method, the modeler must have the requisite expertise to apply appropriate statistical and engineering modeling techniques to perform this analysis. For further information on energy consumption analysis, see the Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol.

Table 2-12 List of models used to simulate savings for NC ECM

Model	Model Name and Purpose	Model Description
1	As-Built Physical To calibrate simulations and assess uncertainty	Model and simulate, as found during site visit. Use the occupancy and building operation, as reflected in billed energy history and sub-metered data. Simulate using actual local weather observations matching the consumption history period.
2	As-Built Design To estimate typical usage at full occupancy	Base on as-built physical model. Use full design occupancy and expected typical building schedules. Use construction and equipment efficiencies, as found during site visits. Simulate using normalized weather data (e.g., typical meteorological year [TMY] datasets).[23]
3	As-Built Expected Design To estimate difference between original and as-built models	Base on as-built design model. Use full design occupancy and expected typical building schedules. Use assumed constructions and equipment efficiencies. Simulate using normalized weather data (e.g., TMY datasets).
4	Whole-Building Reference To estimate savings for the ECM	Base on as-built design model. Use full design occupancy and expected typical building schedules. Apply baseline requirements defined by reference codes or standards. Simulate using normalized weather data (e.g., TMY).
5	Measure Building Reference To isolate savings claimed by the participant	Base on whole-building reference model. Use full design occupancy and expected typical building schedules. Apply baseline requirements defined by reference codes or standards. Include ECM not incentivized by program. Simulate using normalized weather data (e.g., TMY).

Begin the development of the model by generating a model of the building as it was built and is operating during the site visit—and as reflected by utility energy consumption data. Use this initial model, the as-built physical model, to calibrate the modeled building to available physical data. This ensures evaluators can use successive iterations in a predictive capacity. A detailed discussion of the calibration process falls outside the scope of this protocol; however, for detailed calibration procedures and guidelines see Section 6.3.3.4 in ASHRAE Guideline 14- 2002 (ASHRAE, 2002).

Once calibrated, use the as-built physical model to generate the as-built design model, which should reflect the building at full-design occupancy and operation according to expected typical schedules. The only differences between these models are building occupancy, operational schedules, and any modeling guidelines incorporated from codes or standards used to define baseline performance. For buildings currently operating at full occupancy, there may be very little difference between these models. Refer to Table 11.3.1 and G3.1 in ASHRAE Standard 90.1-2007 (ASHRAE, 2007) for examples of modeling requirements specified by codes and standards.

[23] Note the TMY are referenced here as an example series of normalized weather data. When incorporating TMY weather data, use TMY3 weather data when available. While TMY weather represents a common standard, review the reporting needs of the project, as other normalized weather datasets may be more appropriate (e.g., Weather year for Energy Calculations [WYEC] or California Thermal Zones [CTZ]).

Then, use the as-built design model to generate the as-built expected design model. While this model simulates the building's operation according to its design intent, it also includes claimed assumptions regarding envelope constructions and equipment efficiencies. Review the model for discrepancies between claimed assumptions and the physical building; if no discrepancies exist, this model will be identical to the as-built design.

After developing as-built models, evaluators can model baseline building performance, which results in the whole-building reference model; to generate this model, apply the appropriate codes and standards used to define baseline building performance to the as-built design model. The M&V plan should identify such standards before modeling begins. The following section, *Baseline Considerations*, discusses additional considerations for baseline selection. Similar to the as-built design model, the whole-building reference model should reflect the building's operation according to its expected long-term patterns while using equipment and construction that minimally complies with the reference code or standard.

Finally, start with the whole-building reference model to generate the measure building reference model—this model will include ECM not incentivized by the program. It is likely all the implemented ECM are included in the whole-building performance incentives; therefore, both the baseline models may be identical. However, as incentives often are applied for during the building's design and construction process, additional above-code equipment or construction may be implemented that were not included in the final incentive.

2.5.4.3 Baseline Considerations

Defining baseline building physical characteristics and equipment performance is one of the most important (and difficult) tasks in evaluating savings for new construction ECM. This is for several reasons. As noted, new construction ECM do not have a physical baseline to observe, measure, or document. Rather, evaluators must define the baseline "hypothetically" through an appropriate interpretation of the applicable energy codes and standards. It is typically complicated to establish an appropriate interpretation due to the overlapping scope of federal, state, and local codes. Conversely, some states do not have a building energy-efficiency standard separate from the federal standards. Typically, evaluators determine baseline building characteristics and equipment performance requirements by locally adopted building energy codes. In some cases, however, applying a more rigorous, above-code baseline may better reflect standard local construction or industry-standard practices. Thus, in addition to a good understanding of the relationship between federal, state, and local standards, evaluators may need to consult with program guidelines (which often specify greater than code stringency or other technical specifications) or statewide evaluation frameworks. Enforcement of the state codes is the responsibility of the local building officials. The EM&V of programs is usually carried out by utility or other program administrators or by a public utilities commission. Whereas the public utilities commission usually has no enforcement responsibility for the codes and standards, they often point to the official state standards as the governing document regardless of the degree of enforcement of those codes at the local level.

In general, the baseline must satisfy the following criteria (IPMVP, 2006):

- It must appropriately reflect how a contemporary, nonparticipant building would be built in the program's absence.

- Evaluators must rigorously define it with sufficient detail to prescribe baseline conditions for each individual ECM and for the building components simulated.[24]
- Evaluators must develop it with sufficient clarity and documentation to be repeatable.

The BCAP-OCEAN website[25] can be a useful resource in identifying locally adopted energy codes and standards when starting the evaluation of a whole building or new construction project.

2.5.4.4 Calculating Savings

To calculate savings, apply simulation outputs (from models 2 through 5 in Table 2-13) to the formulas described in section 2.5.4.3 *Baseline Considerations*. In all cases except as-built physical, simulate the postconstruction energy use and the projected baseline energy use using normalized weather data (TMY).

As discussed in section 2.5.4.3 *Baseline Considerations*, there are four components that comprise calculated energy savings (defined in Table 2-13 and shown in Figure 2-6). Determine the final reported (verified) savings values in the context of M&V objectives.

Table 2-13 Comparison of Savings Components for NC ECM

Savings Component	Model Subtraction	Description
Expected Measure Savings	N/A	Energy savings expected by the building designers and/or the program application (also known as the project's planned savings)
Rebated Measure Savings	5 - 2	Evaluated (or realized) energy savings for incentivized ECM, often determined by the TPE. Calculate these savings by subtracting the difference in simulated energy use of the as-built design from the measure building reference (the result is also known as the project's *ex post* savings).
Non-Rebated Measure Savings	4 - 5	Energy savings resulting from ECM implemented in the final building design, but not rebated by the program. Calculate these savings by subtracting the difference in simulated energy use of the measure building reference from the whole-building reference (the result is also known as the spillover savings).
Total Achieved Savings	4 - 2	Evaluated (or realized) energy savings for all implemented ECM, whether rebated or not. These are often determined using the TPE and calculated by subtracting the difference in simulated energy use of the as-built design from the whole-building reference. Some programs report this (rather than rebated measure savings) as the project's *ex post* savings.

[24] Locally adopted building codes will define gross savings of new construction programs. Only consider standard construction practices of nonparticipant buildings when performing a net-to-gross analysis. One notable exception is when the evaluated program defines its own baseline, according to an above-code standard (for example, ASHRAE Standard 189.1-2011).

[25] This website can be found here: http://energycodesocean.org

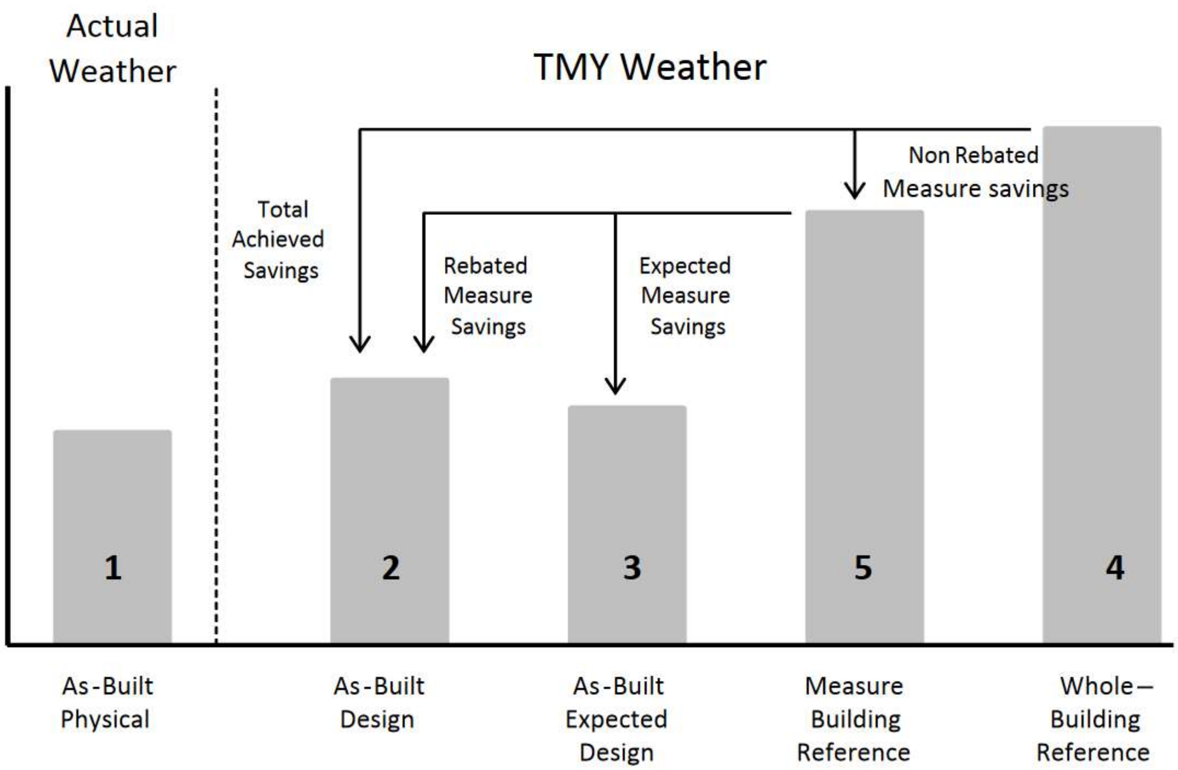

Figure 2-6 Illustration of savings components for NC ECM

2.5.4.4.1 Quantify and Locate Modeling Uncertainty

Due to the complex set of physical, thermodynamic, and behavioral processes simulated, it is difficult to fully characterize the uncertainty in modeled outputs without multiple statistical and analytical tools. Additionally, practical limitations on budgets and time allotted for M&V activities frequently result in qualifying uncertainty in final simulated savings by reporting uncertainty in the model's calibration to energy consumption history. Quantify calibration uncertainty using the Normalized Mean Bias Error (NMBE) and Coefficient of Variation of the Root Mean Square Error (CVRMSE)[26]. Pages 13-16 of ASHRAE Guideline 14-2002 (ASHAE, 2002), provides detailed descriptions of these calculations and their applications.

Determine calibration uncertainty by comparing outputs from the calibrated as-built physical model with the facility's consumption history. Table 2-14 shows calibration uncertainty targets for monthly and hourly consumption history resolutions (ASHRAE, 2002).

[26] These two statistical measurements provide an assessment of the variance between the simulated and measured (by the utility meter) energy use and electric demand. This protocol considers modeling uncertainty acceptable when this variance is below the thresholds suggested in Table 3

Table 2-14 Acceptable Tolerances for Uncertainty in Calibrated Building Simulations

Resolution of Energy Consumption History	NMBE Tolerance	CVRSME Tolerance
Monthly	±5%	±15%
Hourly	±10%	±30%

As newly constructed buildings have a short energy consumption history, it is important to consider how many monthly observations are required to attain a suitably calibrated model. The amount of consumption history required for calibration depends on building type and occupancy. Buildings with little seasonal variations in energy use[27] and short ramp-up periods may need as little as three or four months of consumption history, assuming building occupancy and usage are well-defined and stable. Typically, buildings in this category include grocery stores, restaurants, and data centers.

Conversely, buildings that experience significant seasonal variation, or that are not fully occupied for extended periods, may require a complete year (or more) of consumption history before modelers can determine a reliable calibration. For these buildings, occupancy and usage must be well-defined and stable during all observations used for calibration. Typical buildings of this type include offices, schools, and malls (both strip and enclosed).

Mandating definitive requirements for the minimum number of observations required to sufficiently calibrate a simulation would unduly constrain modelers and could place impractical limitations on EM&V efforts. However, this protocol recommends the following as guidelines:

- Observations should sufficiently characterize a building’s energy use, so modelers can extrapolate reliable annual energy-use values.
- Observations should sufficiently describe expected seasonal variations in building operations.
- Building occupancy and operating conditions must be known for the set of observations.
- Building occupancy and operating conditions must remain stable for the duration of observations used for calibration.

While NMBE and CVRSME may prove useful in describing uncertainty in final savings, it is important to minimize the uncertainty in the simulation inputs. These metrics will not completely capture uncertainty in the inputs.

All software packages acceptable for use in Option D require modelers specify a significant number of physical parameters before simulating a building. Often, many of these parameters have default settings in the software package; however, evaluators can base the parameter inputs on experience or standard practices.

[27] Although energy used by HVAC systems can vary seasonally, such usage generally correlates well with outside weather. Thus, the energy simulation model can sufficiently extrapolate such seasonality (when simulated using the appropriate weather data), reducing the number of billed observations required to calibrate buildings having HVAC use that is dominated by weather.

Lack of a unique calibration point can cause misleading results for NMBE and CVRSME. Furthermore, the resultant calibrations respond differently to changes in other parameters, which can lead to significantly divergent savings estimates. Therefore, it is very important modelers minimize calibration uncertainty and they accomplish the calibration for the correct reasons. Modelers should not unreasonably alter inputs simply to reduce NMBE or CVRSME.

The following guidelines minimize uncertainty in the calibration process:

- Experienced simulators (or modelers directly supervised by an experienced simulator must perform the modeling.
- Modelers must document each simulation process step, so reviewers can audit the model, its outputs, and its assumptions.
- Simulators and auditors should determine the most influential default model parameters and confirm their appropriateness.[29]
- Simulated equipment (e.g., HVAC coils, chillers, pumps) should not "auto size" in final simulations.
- Simulators should identify the parameters to which the simulation outputs are most sensitive.[30]

In addition to quantifying NMBE and CVRSME errors, modelers should analyze the sensitivity of final savings to variations in key model inputs. Modelers should also report such parameters (including their effects on simulated energy savings and the uncertainty in their values) with calibration uncertainty.

2.5.4.5 Sample Design

Use sampling under the following conditions:

- When performing submetering on building equipment
- When performing a detailed survey of an entire building proves impractical.

Evaluators determine the specific targets for sampling certainty and relative precision in the context of the evaluation. For detailed information regarding sample design and for calculating certainty and precision, see the UMP *Sample Design Cross-Cutting Protocol*.

2.5.4.5.1 Sampling for Submetering

Perform submetering to collect information regarding a building's operational schedules. Monitored systems include lighting, ventilation, large equipment (e.g., data centers), and HVAC zone temperatures. Generally, it is acceptable to assume a coefficient of variation (CV) of 0.5 for most submetering; however, while many of these schedules are a function of the overall building type, significant variation in schedules can occur from space to space within a facility. Therefore, interview site personnel to identify any operational differences (and the magnitude of such differences) within the facility before creating a sample design. Account for variations in operating schedules and usage patterns by using a larger CV or by stratifying unique usage groups. See the

[29] When specific data are unavailable, auto-sizing can be helpful in determining appropriate coil capacities, fan speeds, etc. However, only use it for initial equipment sizing. Once equipment sizes have been determined, input them directly. Often, modelers must use auto-sizing to define baseline equipment, as the measures impact building loads. In such cases, calculate an oversize ratio for as-built equipment and apply it to the baseline simulation

[30] Further discussion regarding sensitivity analysis of simulation parameters falls outside this chapter's scope. For additional material on this topic, see Spitler, Fisher, & Zietlow 1989.

Uniform Method Project's Metering Cross-Cutting Protocols for additional considerations for commonly monitored equipment.

(i) Example: Monitoring the Lighting Schedule in a Two-Story Office Building

A two-story commercial office building receives a whole-building performance rebate for LEED certification. For the certification process, a DOE2.2 model is built, for which evaluators develop lighting loads and schedules. During the on-site visit, evaluators note the same tenant occupies both floors, and the building remains open from 6:30 a.m. to 10:00 p.m. The evaluators also identify two unique lighting usage patterns:

- Enclosed offices are located on the building's perimeter
- Open office space is located in the building's core.

As the evaluators identified two distinct usage patterns, they should design the sampling to capture the variability within the schedules for both space types.

- As the open office space is located in the building's core, lighting fixtures likely operate continuously during the building's open hours. Additionally, lighting is commonly shared by all workspaces in the building's core. Therefore, a CV of 0.5 is justified and may prove conservative in determining how many fixtures to monitor.
- Lighting fixtures located in enclosed office spaces typically experience significantly more usage variation due to exaggerated behavioral and external influences. Also, the enclosed office space fixtures receive additional light from perimeter windows, thereby reducing the need for interior lighting during daytime hours. These impacts can be exaggerated (or diminished), depending on fixture control types, building aspects, weather, and times of year. Such additional variability would necessitate a higher assumed CV and additional monitoring points.

2.5.4.5.2 Sampling for Building Surveys

The on-site data collection encompasses a detailed survey of building systems, such as:

- Lighting fixtures
- Plug loads
- HVAC equipment and controls
- Elevator and auxiliary equipment
- Fenestration
- Envelope constructions

For many buildings, surveyors can perform a complete walk-through and can install monitoring equipment within a single day. However, larger buildings (such as high-rise office buildings, hotel casinos, and hospitals) present logistical and budgetary complexities that make it impractical (and often impossible) to perform a complete facility walk-through. In these cases, it is permissible to perform a walkthrough of a representative sample of building areas and extrapolate the findings to the rest of the building. Evaluators can apply the findings to individual spaces or to entire floors (the exact sample design depends on the facility design, including any considerations, such as access to space).

as office space, including a Department of Agriculture office. Evaluators collect data during the on-site visit to build a DOE2.2 model; however, the building owner will only provide evaluation personnel access to the building for a single day.

The building is too large to conduct a thorough walk-through in one day. Additionally, it is expected at least one tenant will have areas within its occupied space that evaluators will not be allowed to access. Therefore, evaluators will have to perform sampling for both floors and space types. Evaluators should audit enough floor space to sufficiently characterize internal loads and usage patterns for each tenant and for the building as a whole. The exact number of floors visited will depend on the number of tenants and on the homogeneity between spaces/floors.

The evaluators should:

- Identify unique operating conditions, such as occupancy schedules, lighting power density (and schedules), and equipment power density (and schedules)
- Identify currently vacant areas (or floors)
- Interview facility staff to:
- Identify differences in space temperatures or ventilation requirements for each tenant
- Determine variations in building occupancy (by month or as appropriate) since its opening
- Audit all central plant equipment
- Sample air distribution system equipment using sampling criteria described in the Uniform Method Project's Chapter 11 *Sample Design Cross-Cutting Protocol*

2.5.5 PROGRAM EVALUATION ELEMENTS

These elements differentiate evaluations of new construction programs from those of other programs:

- Evaluators need significantly more resources to define and justify a hypothetical baseline
- Evaluators have a limited selection of methods for determining site-level savings
- Buildings rarely operate at a "steady state" at the time of evaluation

While this is not a comprehensive list, it specifies critical factors that evaluators must consider in developing an evaluation plan—particularly regarding budget resources for defining and justifying the baselines used to determine energy savings.

Commonly applied codes (such as ASHRAE 90.1) provide multiple compliance pathways but leave room for local jurisdictions to maintain their own interpretations. Therefore, evaluators should work with local jurisdictions, program implementers, and evaluation managers and oversight agencies to identify the most appropriate baseline for a building. Further, local jurisdictions may adopt an updated building code during implementation of a program, so the evaluator may have to develop baselines from multiple building codes for a given program year.

Given the limited information available to assess new construction ECMs, using calibrated building simulations is often the only option for determining energy savings. Significant planning ensures:

- Evaluators develop detailed M&V plans each project site
- The evaluation allows sufficient time to perform the analyses

Evaluators often collect additional information using submetering and/or consumption data analysis. As this information is important for model calibration, the M&V plan should allot sufficient time for a thorough analysis of all sub-metered data and consumption data.

For programs offering incentives, evaluators usually assess energy efficiency measure performance during the first few years of their operation. During this period, building systems and controls typically require troubleshooting[31] and buildings have low, but growing, occupancy rates.

Evaluators should also keep in mind that owners (or tenants) may use building spaces differently than as originally designed. Thus, the specific codes or standards governing the originally permitted building drawings may not be appropriate for assessing actual energy use or energy savings. This protocol strongly recommends evaluators consider these and other such factors when calibrating models and simulating annual energy savings.

[31] Troubleshooting is formally done through a commissioning process; however, not all buildings are professionally commissioned. In many facilities, facility management must dial in building controls.

2.6 Protocols for Evaluation of Retrocommissioning Projects

2.6.1 MEASURE DESCRIPTION

Retrocommissioning (RCx) is a systematic process for optimizing energy performance in existing buildings. It specifically focuses on improving the control of energy-using equipment (e.g., HVAC equipment and lighting) and typically does not involve equipment replacement. Field results have shown proper RCx can achieve energy savings ranging from 5% to 20%, with a typical payback of two years or less (Thorne, 2003). [32]

The method presented in this protocol provides direction regarding: (1) how to account for each measure's specific characteristics and (2) how to choose the most appropriate savings verification approach.

A study conducted on behalf of LBNL analyzed data from 11 utilities operating RCx programs across the United States. The dataset included 122 RCx projects and more than 950 RCx measures (PECI, 2009). Table 2-15 lists a summary of the most common RCx measures, highlighting the nine measures that represent the majority of the analyzed project savings.

Table 2-15 Common RCx Measures

RCx Measure	Percentage of Total Savings
Revise control sequence	21%
Reduce equipment size	15%
Optimize airside economizer	12%
Add/optimize supply air temperature reset	8%
Add variable frequency drive to pump	6%
Reduce coil leakage	4%
Reduce/reset duct static pressure set point	4%
Ad/optimize optimum start/stop	3%
Add/optimize condenser water supply temperature reset	2%

As shown in Table 2-15 (PECI, 2010), RCx measures vary, depending on types of equipment and control mechanisms introduced or optimized. For example, some RCx measures control HVAC equipment according to a predefined schedule, while some measures introduce outdoor air temperature (OAT) dependent controls.

[32] As discussed in the section "Considering Resource Constraints" of the Introduction chapter to this report, small utilities (as defined under U.S. Small Business Administration regulations) may face additional constraints in undertaking this protocol. Therefore, alternative methodologies should be considered for such utilities.

Table 2-16 Categorization of RCx Measures

Control Mechanism	Equipment Type		
	HVAC Airside	HVAC	Lighting
Scheduled	Matching supply fan schedule to occupancy schedule	Adding/optimizing space setback temperatures	Matching lighting schedule to occupancy schedule
Variable	Optimizing airside economizer	Adding chilled water supply temperature set point reset strategy	Optimizing daylighting control

The classic RCx process helps identify, implement, and maintain improvements to building systems and operations via the following five phases (BPA, 2011a).

- Planning. This phase involves screening buildings to determine whether they provide a good fit for RCx by assessing indicators such as equipment age and condition, building energy performance and size, and type of control system. Ideally, facilities should have an existing building automation system (BAS) in good working order, as well as HVAC equipment that is in relatively good condition. A facility without a BAS can install the system; however, the project would then become an HVAC controls and commissioning project rather than an RCx project. When a facility's HVAC equipment nears the end of its useful life, undertaking RCx may not be appropriate because control measures could become obsolete with replaced equipment.
- Investigation. The investigation phase involves analyzing facility performance by reviewing building documentation; performing diagnostic monitoring and functional tests; interviewing staff; identifying a list of recommended improvements; and estimating savings and costs. Evaluators should clearly differentiate valid RCx measures that meet program eligibility guidelines from retrofit measures and/or operation and maintenance (O&M) activities at this phase.
- Implementation. The implementation phase involves prioritizing recommended measures and developing an implementation plan; implementing the measures; and testing to ensure proper operation. Implementation often entails an iterative approach, as the evaluator may need to determine the final control set points through several stages of modification and assessment. These stages ensure building equipment continues to operate properly and maintains the occupants' comfort. Typically, evaluators will review a facility's BAS to assess how effectively RCx measures operate.
- Turnover. The turnover phase involves updating building documentation (e.g., system operation manuals); developing and presenting a final report; and training building operators on proper O&M.
- Persistence. The persistence phase involves monitoring and tracking energy use over time; continually implementing persistence strategies (e.g., refining control measures or enhancing O&M procedures) to sustain savings; and documenting ongoing changes. Depending on the availability of resources and the timeline, program stakeholders may not always actively support this phase.

2.6.2 APPLICATION CONDITIONS OF PROTOCOL

The RCx program design includes activities intended to overcome several market barriers, as listed in Table 2-17.

Table 2-17 RCx Market Barriers

Market Segment	Barrier	Opportunities
Supply-Side Actors, End Users	No tangible examples of RCx performance in situ	Undertaking pilot opportunities
Supply-Side Actors	Lack of service provider capacity for undertaking the RCx investigation and implementation phases	Training for service providers
End Users	Lack of awareness and understanding of the RCx benefits	Education to increase building owner and operator awareness
End Users	Cost of undertaking RCx	Incentives

Ideally, programs overcome these barriers through various activities that address available opportunities. Retrocommissioning programs may include some or all the following activities:

- Pilot projects. Program administrators sometimes fund pilot projects to demonstrate the benefits of RCx to end users in their target markets. Evaluators can verify pilot savings using the methods presented later in this protocol and, in theory, these savings will attract participants to the program.
- Training. Program administrators sometimes fund or develop training for service providers. In some jurisdictions, service providers do not routinely provide RCx services to their customer base. Thus, to develop RCx capacity in the market, program administrators might offer training to service providers on how to provide common practice RCx investigation and implementation services. Service providers may also require training on how to sell these services to their clients.
- Education. Program administrators sometimes develop educational materials and hold events or workshops for end users. Prior to making a decision to undertake RCx activities in their facilities, building management and building operators need to understand the business case for RCx. Detailed case studies showcasing project savings are an example of education tools program staff can use to facilitate this decision-making process.
- Incentives. Program administrators often provide incentives to undertake the RCx investigation, implementation, and persistence phases. Even though the payback for RCx measures is typically low, end users often require incentives to encourage them to move forward with projects.[33] Incentives may also encourage end users to undertake projects sooner—or with a greater scope—than they would have without market intervention.

This protocol provides structured methods for determining energy savings resulting from the implementation of RCx measures. The approaches described here provide direction on how to verify savings consistently from pilot projects, as well as from projects implemented by program participants. It does not address savings achieved through training or through market transformation activities.

[33] Some programs may impose a penalty rather than an incentive. For example, if participants fail to implement the measures that fell below a certain payback threshold identified during the investigation phase, they may not be eligible for the full investigation phase incentive.

2.6.3 SAVINGS CALCULATIONS

Specific savings calculations[34] for RCx measures inherently vary, due to the breadth of possible RCx measures, which can differ by type of equipment or control mechanism. This section presents a high-level gross energy savings equation that is applicable to all RCx measures. Section 2.5.4 *Measurement and Verification Plan*, includes detailed directions for calculating savings for specific measure categories.

Use the following general equation (EVO, 2012) to determine energy savings:

$$Energy\ Savings = (Baseline\ Energy - Reporting\ Period\ Energy) \pm Routine\ Adjustments \pm Nonroutine\ Adjustments$$

Where:

$Energy\ Savings =$ First-year energy consumption savings

$Baseline\ Savings =$ Pre-Implementation consumption

$Reporting\ Period\ Savings =$ Post-Implementation consumption

$Routine\ Adjustments =$ Adjustments made to account for routinely changing independent variables (variables that drive energy consumption). If applicable, normalize savings to typical meteorological year (TMY35) weather data, as well as other significant independent variables (e.g., occupancy, production data).

$Nonroutine\ Adjustments =$ Adjustments made to account for parameters typically not expected to change during the implementation period. Account for these parameters if they change and this change influences the reporting period energy use (e.g., changes to a facility's building envelope during implementation of and RCx HVAC measure). Evaluators only need to consider nonroutine adjustments if verifying savings using Option C of the International Performance Measurement and Verification Protocol (IPMVP).[36]

Determining RCx demand savings is not a straightforward extension of verified consumption savings (unlike lighting retrofits, where evaluators can easily apply established load savings profiles to consumption savings data). For RCx projects, load savings profiles vary depending on the type of measures implemented and the distribution of these measures. If applicable, evaluators should produce load savings profiles on a measure-by-measure basis, aggregate these profiles, and then apply site-specific coincidence factors to determine coincident peak demand savings at the project level.

[34] As presented in the Introduction, the protocols focus on *ex post* gross energy savings and do not include other parameter assessments, such as net-to-gross, peak coincidence factors, or cost-effectiveness.

[35] Evaluators should use the most recent typical meteorological year dataset. As of January 2014, the most comprehensive national typical meteorological year dataset is TMY3. Evaluators should confer with the local jurisdiction to see if they should use a different regional dataset.

[36] Option C is the "whole-facility approach" to verifying savings

2.6.4 MEASUREMENT AND VERIFICATION PLAN

This section outlines the recommended approaches to determining RCx energy savings and provides directions on how to use the approaches under the following headings:

- Measurement and verification (M&V) method
- Data collection
- Interactive effects
- Specific savings equations
- Regression model direction
- Deemed spreadsheet tool functionality requirements

2.6.4.1 Measurement and Verification Method

There is a structured method for determining the most appropriate approach to verifying RCx energy savings. This method balances the need for accurate energy-savings estimates with the need to keep M&V costs in check, relative to project costs and anticipated energy savings. Depending on which measures are implemented, different approaches to estimating the savings are appropriate. Following the IPMVP, the options are:

- Option A—Retrofit Isolation: Key Parameter Measurement
- Option B—Retrofit Isolation: All Parameter Measurement
- Option C—Whole Facility
- Option D—Calibrated Simulation

Measurement is inherent with most RCx projects because RCx measures typically involve modifications made through a facility's BAS. As mentioned, RCx implementation (an iterative process) often leverages metered data to evaluate and optimize changes throughout the process. Therefore, in many cases, a retrofit isolation approach adhering to Option A or Option B of the IPMVP proves most logical. That said, scenarios exist where Option C, Option D, or even a deemed approach may be more appropriate. Figure 2-7 presents a decision flow chart for determining the approaches to follow.

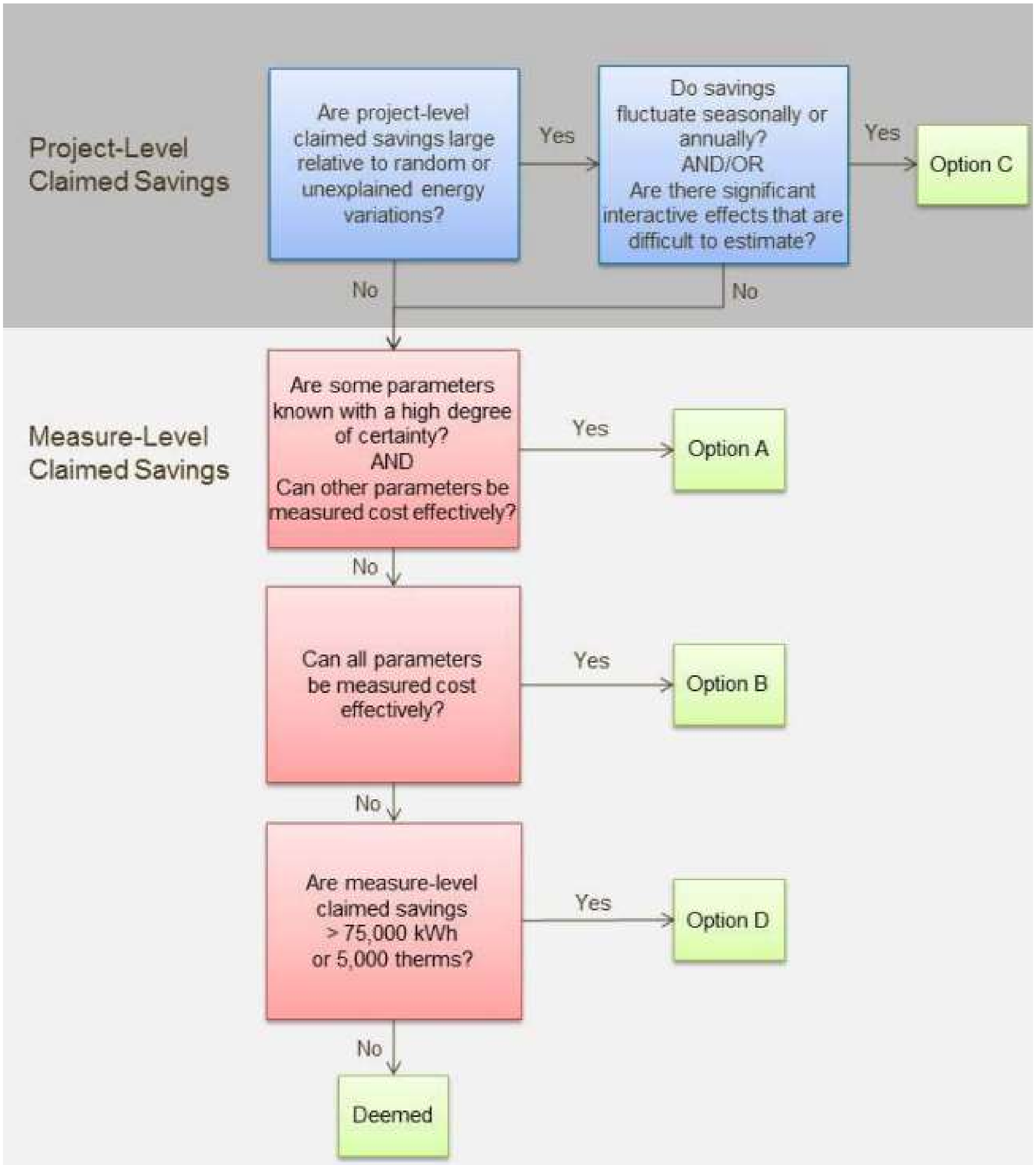

Figure 2-7 RCx Decision Flow Chart[37]

The decision flow chart accounts for factors such as the magnitude of estimated savings and the measurement's cost-effectiveness. Begin the process by considering project-level savings:

- Option C. Use a whole-facility approach—adhering with Option C of the IPMVP—if estimated project-level savings are large compared to the random or unexplained energy variations that occur at the whole-

[37] NREL RCx Evaluation Protocol https://www.nrel.gov/docs/fy17osti/68572.pdf

facility level[38] and if savings fluctuate over a seasonal or annual cycle (e.g., savings that fluctuate depending on OAT). This approach is likely the most cost-effective approach for verifying savings. The whole-facility approach is relatively inexpensive because evaluators can use utility billing data for the analysis. The downside of the approach is that evaluators cannot perform verification until after collecting a full season or year of reporting period data and monitoring and documenting any changes to the facility's static factors[39] over the course of the measurement period. Even if savings remain consistent month to month, Option C may provide the best approach if project measures cause complex, significant interactive effects. Such interactive effects are, by nature, difficult to estimate accurately. Also, if the effects are significant (large, relative to direct-measure savings), evaluators will be required to use a whole-facility approach to measure impacts accurately. The reduced heating and cooling energy resulting from schedule changes to an air-handling unit, when control modifications have also been undertaken for both the heating and cooling systems, is an example of a complex significant interactive effect warranting Option C.

If Option C is ruled out, consider performing verification on a measure-by-measure basis:

- Option A. If measures involve some parameters known with a high degree of certainty and other parameters can be measured cost-effectively, use a retrofit isolation approach adhering to Option A of the IPMVP. In many cases, evaluators can collect metered data directly from the facility's BAS. If required, the facility can add control points to the BAS, either as part of the implementation process or specifically for M&V purposes. Where the BAS cannot provide the information, use temporary meters to collect data (provided that costs are not prohibitive).
- Option B. If a given measure's parameters are uncertain but can be measured cost-effectively, use a retrofit isolation approach, adhering to Option B of the IPMVP. Again, collect metered data (similar to Option A) either through the BAS or by using temporary meters.
- Option D. For measures where it is prohibitive to meter all required parameters, use a calibrated simulation approach adhering to Option D of the IPMVP. Undertake calibrations in two ways: (1) calibrate the simulation to the actual baseline or reporting consumption data and (2) confirm the reporting period inputs via the BAS front-end system, when possible.[40] [41]
- Deemed. Finally, if a measure is relatively common[42] and its estimated savings are small, evaluators can deem savings rather than simulate them. Use this approach for common measures with savings less than 75,000 kilowatt-hour (kWh) or 5,000 therms[43] (PECI, 2010). Use a spreadsheet tool to calculate savings, adhering to functionality requirements presented later in the Protocol.

---

[38] Typically, savings should exceed 10% of the baseline energy for a particular meter (e.g., electricity meter) to confidently discriminate the savings from the baseline data when the reporting period is shorter than two years (EVO, 2012).

[39] Many factors can affect a facility's energy consumption, even though evaluators do not expect them to change. These factors are known as "static factors" and include the complete collection of facility parameters that are generally expected to remain constant between the baseline and reporting periods. Examples include building envelope insulation, space use within a facility, and facility square footage.

[40] In many cases, the simulation should represent the entire facility; however, in some cases, depending on the facility's wiring structure, a similar approach could be applied to building submeters, such as distribution panels that include the affected systems.

[41] See the Uniform Method Project's Commercial New Construction Protocol for more information on using Option D.

[42] If regulators are involved, going through the effort of deeming savings for a rare measure can be burdensome.

[43] Program administrators and evaluators may wish to customize these thresholds for particular programs and/or jurisdictions.

2.6.5 DATA COLLECTION

Depending on the approach followed, these M&V elements will require consideration:

- The measurement boundary
- The measurement period and frequency
- The functionality of measurement equipment being used
- The savings uncertainty

2.6.5.1 Measurement Boundary

For measures evaluators assess using Option A or Option B and that require metering external to the BAS, it will be important to define the measurement boundary. When determining boundaries—the location and number of measurement points required—consider the project's complexity and expected savings:

- While a narrow boundary simplifies data measurement (e.g., a single piece of equipment), variables driving energy use outside the boundary (i.e., interactive effects) still need to be considered.
- A wide boundary will minimize interactive effects and increase accuracy (e.g., systems of equipment like chilled water plants and air-handling units). However, as M&V costs may also increase, it is important to ensure the expected project savings justify the increased M&V costs.

2.6.5.2 Measurement Period and Frequency

For all measures assessed with Option A or Option B, consider two important timing metrics:

- The measurement period (the length of the baseline and reporting periods)
- The measurement frequency (how regularly to take measurements during the measurement period)

As a general rule, choose the measurement period to capture a full cycle of each operating mode. For example, if there is a control modification to heating equipment, collect data over the winter and shoulder seasons.

Choose the measurement frequency by assessing the type of load measured:

- Spot measurement: For constant loads, measure power briefly, preferably over two or more intervals.
- Short-term measurement: For loads predictably influenced by independent variables (e.g., HVAC equipment influenced by OAT), take short-term consumption measurements over the fullest range of possible independent variable conditions, given M&V project cost and time limitations.[44] For systems expected to have nonlinear dependence (such as air handling units with outside air economizers), measurements should incorporate sufficient range to characterize the full breadth of conditions.
- Continuous measurement: For variable loads, measure consumption data continuously, or at appropriate discrete intervals, over the entire measurement period.

See Section 4.4, Specific Saving Equations, for direction regarding measurement periods and frequency for specific measure types.

[44] For example, if a chiller plant undergoes control modifications, the measurement frequency should be long enough to capture the full OAT operating range. In a temperate climate zone, evaluators can accomplish this by taking measurements over a four-week period in the shoulder season and another four-week period during the summer season.

2.6.5.3 Measurement Equipment

When meters external to the BAS are required, follow these guidelines to select a meter[45]:

- Size the meter for the range of values expected most of the time.
- Select the meter repeatability and accuracy that fits the budget and intended use of the data.
- Install the meter as recommended by the manufacturer.
- Calibrate the meter before it goes into the field, and maintain calibration as recommended by the manufacturer. If possible, select a meter with a recommended calibration interval that is longer than the anticipated measurement period.

If BAS data is used, evaluators should exercise due diligence by determining when the BAS was last calibrated and by checking the accuracy of the BAS measurement points.

2.6.5.4 Savings Uncertainty

If possible, quantify the accuracy of measured data[46] and, if practical, conduct an error propagation analysis to determine overall impacts on the savings estimate.

2.6.6 INTERACTIVE EFFECTS

For projects following Option A, Option B, or deemed approaches, consider, and estimate interactive effects if they are significant. For example, if a facility reduces an air-handling unit supply fan schedule, not only will direct fan savings be achieved, but significant cooling and heating energy savings may be realized due to decreases in conditioned ventilation air supplied to the space. Estimate interactive effects using equations that apply the appropriate engineering principles. Ideally, use a spreadsheet tool adhering to the same functionality requirements discussed in 2.6.8 *Deemed Spreadsheet Tool Functionality Requirements* for the deemed spreadsheet tool to conduct these analyses. When interactive effects are large, it may be possible to measure them rather than apply engineering estimates. In the “supply fan” example discussed in the paragraph above, an evaluator can meter the chilled water plant to determine the cooling load reduction.

Interactive effects for projects being verified using Option C or Option D are typically included in facility-level savings estimates.

2.6.6.1 Specific Savings Equations

If following Option A or Option B, verify savings using equations matching a given measure’s characteristics—specifically, whether savings are dependent on independent variables (such as OAT) and the control mechanism for affected equipment.

Figure 2-8 shows the three categories of savings equations, with further explanations following the flow chart.

[45] For more information on selecting measurement equipment, see the Uniform Methods Project’s Metering Cross Cutting Protocols.

[46] Metering accuracy is only one element of savings uncertainty. Inaccuracies also result from modeling, sampling, interactive effects, estimated parameters, data loss, and measurements being taken outside of a meter’s intended range.

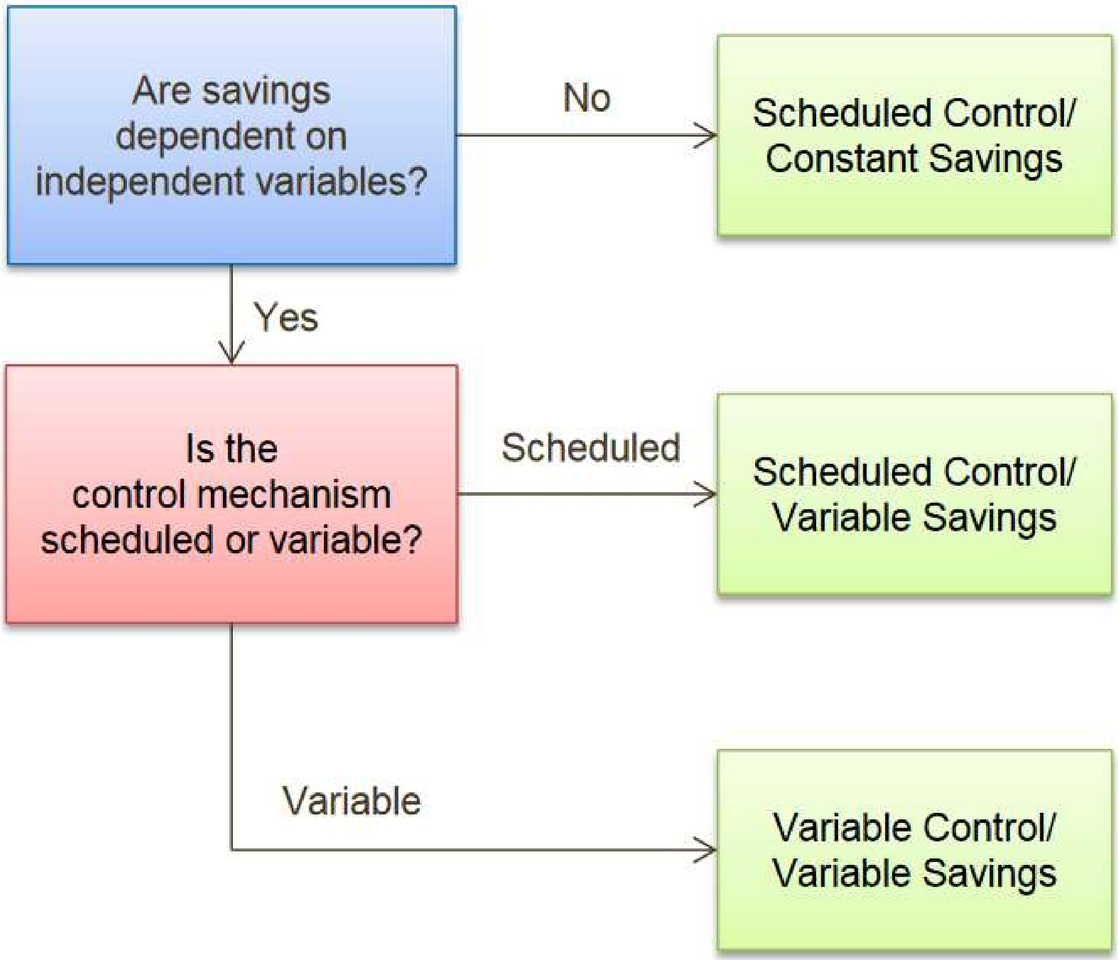

Figure 2-8 Savings Equation Categories[47]

2.6.6.2 Scheduled Control/Constant Savings

This savings equation category encompasses scheduled control measures on equipment not influenced by independent variables (such as OAT); therefore, this is the most straightforward equation category.

Lighting schedule optimization is an example of a measure verified using this savings equation category. In this example, lighting is turned off according to a schedule (scheduled control), and constant savings is achieved while it is off (constant savings).

$$Scheduled\frac{Control}{Savings} = Baseline\ Energy\ - Reporting\ Period\ Energy$$

[47] *Ibid.*

Where:

$Scheduled\ Control/Savings =$ First-year energy consumption savings resulting from a scheduled control measure with constant savings.

$Baseline\ Energy =$ $HRS_{baseline}$ x $kW_{controlled}$

$Reporting\ Period\ Energy =$ $HRS_{reporting}$ x $kW_{controlled}$

$HRS_{baseline}$ = Annual operating hours during the baseline: if this parameter is not known with a high degree of certainty, take short-term measurements for the duration of each existing schedule type.

$HRS_{reporting}$ = Annual operating hours during the reporting period: take short-term measurements for the duration of each new schedule type.

$kW_{controlled}$ = Electric demand controlled by scheduling measure: if this parameter is not known with a high degree of certainty, take spot measurements during the baseline or reporting period.

2.6.6.3 Scheduled Control/Variable Savings

This savings equation category encompasses scheduled control measures on equipment influenced by independent variables (such as OAT). Space setback temperature optimization provides an example of a measure verified using this savings equation category. In this example, the heating space temperature set point is lowered according to a schedule during unoccupied hours (scheduled control), and the savings achieved will vary, depending on OAT (variable savings).

Following the equation below, Table 2-18 lists the five-step process for determining adjusted baseline and reporting period energy consumption.

$$Scheduled\ Control/Variable\ Savings = Baseline\ Energy\ - Adjusted\ Reporting\ Period\ Energy$$

Where:

$Scheduled\ Control/Variable\ Savings =$ First-year energy consumption savings resulting from a scheduled control measure with variable savings.

$Adjusted\ Baseline\ Energy = \sum_{All\ Schedule\ Types} Adj\ Baseline\ Consumption_{Schedule\ Type}$ and determined through the five-step process listed in Table 2-18.

$Adjusted\ Reporting\ Period\ Energy =$ $\sum_{All\ Schedule\ Types} Adj\ Reporting\ Period\ Consumption_{Schedule\ Type}$ and determined through the five-step process listed in Table 2-18.

Table 2-18 Adjusted Consumption for Scheduled Control/Variable Savings Measures[48]

<table>
<tr><th>Steps</th><th colspan="3">Details</th></tr>
<tr><td>Develop baseline/reporting regression model(s) by measuring equipment operation and independent variables.</td><td colspan="3">Take short-term measurements at representative load levels for the affected equipment for each schedule type.
Take coincident measurements of the independent variable(s).
Do a regression analysis to determine the relationship between independent variables and equipment load. This relationship should be expressed in terms of an equation (baseline/reporting period model).
If there are schedules for occupied and unoccupied times during the reporting period, evaluators will need two regression models, one for each set of data.</td></tr>
<tr><td rowspan="3">Develop a bin operating profile[49] a by normalized independent variable data.</td><td colspan="3">Develop bin data tables presenting the following data:</td></tr>
<tr><td>Independent Variable</td><td>Load</td><td>Annual Hours</td></tr>
<tr><td>Create approximately 10 bins over the normalized independent variable data range (if the equipment’s energy consumption varies depending on weather, use TMY data).</td><td>Calculate the normalized load by applying the baseline/ reporting period regression model to the midpoint of each bin.</td><td>Use short-term measured data to estimate hours of operation within each bin or base this on TMY data and the equipment operating schedule.</td></tr>
<tr><td>Calculate the baseline/ reporting period consumption at each load bin.</td><td colspan="3">Adjusted Consumption:
$Load, Schedule\ Type = Load_{Schedule/Type} \times Annual\ Hrs_{Schedule\ Type}$</td></tr>
<tr><td>Sum the consumption savings across bins for each schedule type.</td><td colspan="3">$\sum_{All\ Load\ Bins_{Schedule\ Type}} Adj\ Consumption_{Load,Schedule\ Type}$</td></tr>
<tr><td>Sum the consumption savings across schedule types.</td><td colspan="3">$\sum_{All\ Schedule\ Types} Adj\ Consumption_{Schedule\ Type}$</td></tr>
</table>

2.6.6.4 *Variable Control/Variable Savings*

This savings equation category encompasses variable control measures on equipment influenced by independent variables, such as OAT. Introducing a chilled water supply temperature set point reset strategy serves as an example of a measure verified through this savings equation category. In this example, the chilled water supply temperature set point is determined depending on OAT (variable control), and the savings achieved will vary depending on OAT (variable savings).

Following the equation below, Table 2-19 lists the four-step process for determining the adjusted baseline and reporting period energy consumption.

---

[48] Alternatively, of the independent variable is OAT, TPE can develop an hourly profile over the operating schedule of the affected equipment.

[49] Alternatively, if the independent variable is OAT, evaluators can develop an hourly profile over the full operating schedule of the affected equipment.

$$Variable\ Control/Variable\ Savings$$
$$= Adjusted\ Baseline\ Energy\ - Adjusted\ Reporting\ Period\ Energy$$

Where:

$Variable\ Control/Variable\ Savings =$ First-year energy consumption savings resulting from a variable control measure with variable savings.

$Adjusted\ Baseline\ Energy = \sum_{All\ Load\ Bons} Adj\ Baseline\ Consumption_{Load}$ and determined through the five-step process listed in Table 2-19.

$Adjusted\ Reporting\ Period\ Energy =$ and determined through the five-step process listed in Table 2-19.

Table 2-19 Adjusted Consumption for Variable Control/Variable Savings Measures

Steps	Details		
Develop baseline/reporting regression model(s) by measuring equipment operation and independent variables.	Take short-term measurements at representative load levels for the affected equipment for each schedule type. Take coincident measurements of the independent variable(s). Do a regression analysis to determine the relationship between independent variables and equipment load. This relationship should be expressed in terms of an equation (baseline/reporting period model). If there are schedules for occupied and unoccupied times during the reporting period, evaluators will need two regression models, one for each set of data.		
Develop a bin operating profile[50] a by normalized independent variable data.	Develop bin data tables presenting the following data:		
	Independent Variable	**Load**	**Annual Hours**
	Create approximately 10 bins over the normalized independent variable data range (if the equipment's energy consumption varies depending on weather, use TMY data).	Calculate the normalized load by applying the baseline/ reporting period regression model to the midpoint of each bin.	Use short-term measured data to estimate hours of operation within each bin or base this on TMY data and the equipment operating schedule.
Calculate the baseline/ reporting period consumption at each load bin for each schedule type.	Adjusted Consumption: Load Schedule Type = $Load_{Schedule/Type}$ x Annual $Hrs_{Schedule\ Type}$		
Sum the consumption savings across bins for each schedule type.	$\sum_{All\ Load\ Bins_{Schedule\ Type}} Adj\ Consumption_{Load,Schedule\ Type}$		

[50] *Ibid.*

Sum the consumption savings across schedule types.	$\sum_{All\ Schedule\ Types} Adj\ Consumption_{Schedule\ Type}$

2.6.7 REGRESSION MODELING DIRECTION

Calculating normalized savings for most projects—whether following the IPMVP's Option A, Option B, or Option C— will require the development of a baseline and reporting period regression model.[51] Use one of the following three types of analysis methods to create the model:

- Linear Regression: For one routinely varying significant parameter (e.g., OAT).[52]
- Multivariable Linear Regression: For more than one routinely varying significant parameter (e.g., OAT, occupancy).
- Advanced Regression: For a multivariable, nonlinear fit requiring a polynomial or exponential model.[53]

Develop all models in accordance with best practices and only use them when they are statistically valid (see Section 2.6.7.2 *Testing Model Validity*). If no significant independent variables arise (as with a lighting schedule measure), evaluators are not required to use a model because calculated savings will be inherently normalized.

2.6.7.1 Recommended Methods for Model Development

Use energy and independent variable data that is representative of a full cycle of operation. For example, if facility staff implement a heating space temperature setback measure, collect energy data across the full range of OAT for each of the operating schedules (occupied and unoccupied) for each season, as shown in Table 2-20.

Table 2-20 Example of Data Required for Model Development

	Shoulder Season	**Winter Season**
Occupied Hours	Short-term energy measurements during occupied hours. Measurements should be representative of the full range of shoulder season OAT (approximately 10 OAT bins).	Short-term energy measurements during occupied hours. Measurements should be representative of the full range of winter-season OAT (approximately 10 OAT bins).
Unoccupied Hours	Short-term energy measurements during unoccupied hours. Measurements should be representative of the full range of shoulder-season OAT (approximately 10 OAT bins).	Short-term energy measurements during unoccupied hours. Measurements should be representative of the full range of winter-season OAT (approximately 10 OAT bins).

Analyze the data collected to identify outliers. Only remove outliers when there is a tangible explanation to support the erratic data points. Discussion of how to identify outliers is outside the scope of this Protocol.

---

51 This could either be a single regression model that uses a dummy variable to differentiate the baseline/reporting period data or two independent models for the baseline and reporting period, respectively.

52 One of the most common linear regression models is the three-parameter change point model. For example, a model that represents cooling electricity consumption will have one regression coefficient that describes nonweather-dependent electricity use, a second regression coefficient that describes the rate of increase of electricity use with increasing temperature, and a third parameter that describes the change point temperature, also known as the balance point temperature, where weather-dependent electricity use begins.

53 Evaluators may need to use advanced regression methods if RCx activities impact manufacturing or industrial process equipment.

2.6.7.2 Testing Model Validity

To assess the model's accuracy, begin by reviewing the parameters in the table below.

Table 2-21 Model Statistical Validity Guide[54]

Parameter Evaluated	Description	Suggested Acceptable Values
Coefficient of determination (R^2)	A measure of the extent that the regression model explains variations in the dependent variable from its mean value.	> 0.75
T-statistic (absolute value)	An indication of whether regression model coefficients are statistically significant.	> 2a[55]
Mean bias error	An indication of whether the regression model overstates or understates actual energy consumption.	Will depend on the measure, but generally: < ±5%

A model outside the suggested range indicates parameter coefficients that are relatively poorly determined, with the result that normalized consumption will have relatively high statistical prediction error. Ordinarily, evaluators should not use such a model for normalization, unless the analysis includes appropriate statistical treatment of this prediction error. Discussion of how to proceed in such circumstances is outside the scope of this protocol.

When possible, attempt to enhance the regression model by:

- Increasing or shifting the measurement period
- Incorporating more data points
- Including independent variables previously unidentified
- Eliminating statistically insignificant independent variables

Also, when assessing model validity, consider coefficient of variation of the root mean squared error, fractional savings uncertainty, and residual plots. Refer to ASHRAE Guideline 14-2002 and Bonneville Power Administration's Regression for M&V: Reference Guide for direction on how to assess these additional parameters.

2.6.8 DEEMED SPREADSHEET TOOL FUNCTIONALITY REQUIREMENTS

When collecting measured energy data is not cost-effective and claimed (*ex ante*) savings estimates for a given measure are sufficiently small (75,000 kWh or 5,000 therms), use a deemed approach to calculate savings. In this scenario, the protocol recommends using a spreadsheet tool to calculate savings, and this tool should meet these general requirements:

[54] EVO, 2012

[55] Determine the t-statistic threshold based on the evaluator's chosen confidence level; a 95% confidence level requires a t-statistic of 1.96. Evaluators should determine an acceptable confidence level depending on project risk (i.e., savings risk), budget, and other considerations.

- Ensure model transparency. A third party should be able to review the spreadsheet tool and clearly understand how the evaluator derived all savings outputs. To this end, clearly explain and reference all inputs and calculation algorithms within the spreadsheet. Do not lock or hide cells or sheets and check to ensure all links work properly.
- Use relevant secondary data. When using secondary data as inputs to savings algorithms, ensure they are relevant to the project's region or jurisdiction. Substantiate input relevancy within the spreadsheet. For example, if using assumed values for hours of operation for heating equipment, take these secondary data from a regional resource (e.g., a technical resource manual from the most applicable demand-side management authority).
- Verify input elements. Either on site or through the BAS front-end system. Even when using a deemed approach, verify and update some inputs with actual site observations (rather than solely relying on secondary data). For example, confirm a new lighting schedule through the BAS front-end system and note it in the spreadsheet tool.
- Establish default values for unverifiable parameters. Use default values for parameters that cannot be verified. For example, clearly state assumed values for motor efficiencies and load factors.

The Building Optimization Analysis Tool developed by PECI, now CLEAResult, provides an example of benchmark for RCx spreadsheet tools. Although the protocol does not require the following level of rigor, ideally, a best-practice spreadsheet tool should:

- Incorporate regional TMY data.
- Incorporate regional building archetype templates.
- Undergo a calibration process by using measured data from previous regional projects to test algorithms.

2.6.9 SAMPLE DESIGN

Consult the Uniform Methods Project's *Sample Design Cross-Cutting Protocols*[56] for general sampling procedures if the RCx program project population is sufficiently large or if the evaluation budget is constrained. Ideally, use stratified sampling to partition RCx projects by measure type, facility type, and/or project size. Stratification ensures evaluators can confidently extrapolate sample findings to the remaining project population. Regulatory or program administrator specifications typically govern the confidence and precision-level targets that influence sample size.

2.6.10 OTHER EVALUATION ISSUES

When claiming lifetime and net program RCx impacts, evaluators should consider persistence and net-to-gross in addition to first-year gross impact findings.

2.6.11 PERSISTENCE

Persistence of savings encompasses both the retention and the performance degradation of measures. Evaluators should consider persistence on a program-by-program basis because the persistence of RCx projects can vary widely depending on the distribution of measure types implemented and, perhaps more significantly,

[56] https://www.nrel.gov/docs/fy17osti/68567.pdf

on how well facility staff maintains the modifications. Consult the Uniform Methods Project's *Assessing Persistence and Other Evaluation Issues Cross-Cutting Protocols*[57] for more information.

2.6.11.1 Estimating EUL in RCx Projects

For cases where unable to determine measure persistence, the TPE has conducted an analysis of persistence for common measure types in RCx projects and extrapolated EULs suitable for cost-benefit analysis. The analysis was based on findings from a field study of persistence in RCx projects[58].

2.6.11.1.1 Methodology

The TPE calculated EUL for a group of measures using savings persistence estimates. The savings persistence estimates were calculated relative to a baseline program year when measures were implemented. As such, they represent both measure life and savings persistence. Savings persistence accounts for changes in equipment life (the amount of time before equipment fails), measure persistence (i.e., equipment failure or business turnover), and true savings persistence as defined by the UMP (i.e., changes in operating hours, process operations, or performance degradation of the equipment relative to the baseline option).

Savings persistence values were obtained for each measure at various dates following measure installation (e.g., every three years). Savings persistence in years not measured was interpolated between years in which saving persistence was known. This creates a step-like function with different slopes for each measured interval. For years that exceed the last measured persistence, the TPE extrapolated persistence using the slope from the prior measured interval. The EUL was capped at 7 years to account for a lack of savings persistence estimates after year 6.

The equation below shows how the EUL for each measure was calculated from the predicted savings persistence values.

$$Effective\ Useful\ Life = \sum_{t=0}^{n} Savings\ Persistence_t$$

Table 2-22 EUL by Measure

Measure Type	EUL Capped (Yr 7)
Air distribution	4.00
Plant optimization	5.00
Ventilation	5.00
Scheduling	5.50
Filters	5.50
General	5.50

[57] https://www.nrel.gov/docs/fy17osti/68569.pdf

[58] Seventhwave Field Study for ComEd, as referenced in the UMP chapter.

2.6.12 NET TO GROSS

Consult the UMP's *Estimating Net Energy Savings: Common Practices*[59] for a discussion about determining net program impacts at a general level, including direction on how to assess free-ridership. Supplementary to that chapter, however, evaluators may consider assessing participant spillover if evidence emerges of participants implementing no-cost measures. This would specifically apply to no-cost measures identified during the investigation phase, but not explicitly included under the scope of program-funded RCx implementation activities.

If no-cost measures exist and there are no savings claims, the attribution evaluation may involve interviews with building operators and their service providers to obtain estimates of the savings magnitude resulting from these measures. Participant spillover would positively influence the program's overall NTG.

[59] https://www.nrel.gov/docs/fy17osti/68578.pdf

2.7 Protocols for Evaluating Behavioral Programs

The table below outlines common initialisms made in this chapter.

Initialism	What it stands for
BB	Behavior-Based
DiD	Difference in Differences
IPMVP	International Performance Measurement and Verification Protocol
ITT	Intent To Treat
IV	Instrumental Variable
LATE	Local Average Treatment Effect
NREL	National Renewable Energy Laboratory
OLS	Ordinary Least Squares
PG&E	Pacific Gas & Electric
RCT	Randomized Control Trial
RED	Randomized Encouragement Design
SEE Action	State and Local Energy Efficiency Action
TOT	Treatment effect On the Treated
UMP	Uniform Methods Project

2.7.1 DESCRIPTION

Residential BB programs use strategies grounded in the behavioral and social sciences to influence household energy consumption. These may include providing households with feedback about their real-time or historical energy consumption; reframing of consumption information in different ways; supplying energy efficiency education and tips; rewarding households for reducing their energy use; comparing households to their peers; and establishing games, tournaments, and competitions.[60] BB programs often target multiple energy end uses and encourage energy savings, demand savings, or both. Savings from BB programs are usually a small percentage of energy use, typically less than 3%.[61] Utilities introduced the first large-scale residential BB programs in 2008. Since then, scores of utilities have offered these programs to their customers. Although program designs differ, many share these features:

- They are implemented as randomized experiments wherein eligible homes are randomly assigned to treatment or control groups.
- They are large scale by energy efficiency program standards, targeting thousands of utility customers.
- They provide customers with analyses of their historical consumption, energy savings tips, and energy efficiency comparisons to neighboring homes, either in personalized home reports or through a web portal or offer incentives for savings energy.
- They are typically implemented by outside vendors.

[60] See Ignelzi et al. (2013) for a classification and descriptions of different BB intervention strategies and Mazur Stommen and Farley (2013) for a survey and classification of current BB programs. Also, a Minnesota Department of Commerce, Division of Energy Resources white paper (2015) defines, classifies, and benchmarks behavioral intervention strategies.

[61] See Allcott (2011), Davis (2011), and Rosenberg et al. (2013) for savings estimates from residential BB programs.

Utilities will continue to implement residential BB programs as large-scale, randomized control trials (RCTs); however, some are now experimenting with alternative program designs that are smaller scale; involve new communication channels such as the web, social media, and text messaging; or that employ novel strategies for encouraging behavior change (for example, Facebook or other social network competitions to reduce consumption).6 These programs will create new evaluation challenges and may require different evaluation methods than those presented in this protocol. Quasi-experimental methods require stronger assumptions to yield valid savings estimates and may not measure savings with the same degree of validity and accuracy as randomized experiments.

2.7.2 APPLICATION CONDITIONS OF PROTOCOL

This protocol recommends the use of RCTs or randomized encouragement designs (REDs) for estimating savings from BB programs. A significant body of research indicates that randomized experiments result in unbiased and robust estimates of program energy and demand savings. Moreover, recently evaluators have conducted studies comparing the accuracy of savings estimates from randomized experiments and quasi-experiments or observational studies. These comparisons suggest that randomized experiments produce the most accurate savings estimates.[62] This protocol applies to BB programs that satisfy the following conditions[63]:

- Residential utility customers are the target.
- Energy or demand savings are the objective.
- An appropriately sized analysis sample can be constructed.
- Treated customers can be identified and accurate energy use measurements for sampled customers are available.
- It must be possible to isolate the treatment effect when measuring savings.

This Protocol applies only to residential BB programs.[64] This Protocol addresses best practices for estimating energy and demand savings. There are no significant conceptual differences between measuring energy savings and measuring demand savings when interval data are available; thus, evaluators can apply the algorithms in this protocol for calculating BB program savings to either. The Protocol does not directly address the evaluation of other BB program objectives, such as increasing utility customer satisfaction and engagement, educating customers about their energy use, or increasing awareness of energy efficiency.[65] But these program outcomes could be studied in a complementary fashion alongside the energy savings.

[62] Allcott (2011) compares RCT difference in differences (DiD) savings estimates with quasi-experimental simple differences and DiD savings estimates for several home energy reports programs. He found large differences between the RCT and quasi-experimental estimates. Also, Baylis et al. (2016) analyzed data from a California utility time-of-use and critical peak pricing pilot program and found that RCT produced more accurate savings estimates than quasi-experimental methods such as DiD and propensity score matching that relied on partly random but uncontrolled variation in participation.

[63] As discussed in the "Considering Resource Constraints" section of the UMP Chapter 1: Introduction, small utilities (as defined under U.S. Small Business Administration regulations) may face additional constraints in undertaking this protocol. Therefore, alternative methodologies should be considered for such utilities.

[64] Evaluators may be able to apply the methods recommended in this protocol to the evaluation of some nonresidential BB programs. For example, Pacific Gas and Electric (PG&E) offers a Business Energy Reports Program, which it implemented as an RCT (Seelig, 2013). Also, Xcel Energy implemented a business energy reports program as an RCT (Stewart, 2013b). Other nonresidential BB programs may not lend themselves to evaluation by randomized experiment. For example, many strategic energy management programs enroll large industrial customers with unique production and energy consumption characteristics for which a randomized experiment would not be feasible (NREL, 2017).

[65] Process evaluation objectives may be important, and omission of them from this protocol should not be interpreted as a statement that these objectives should not be considered by program administrators.

This Protocol also requires that the analysis sample be large enough to detect the expected savings with high probability. Because most BB programs result in small percentage savings, a large sample size (often in the thousands or tens of thousands of customers) is required to detect savings. This protocol does not address evaluations of BB programs with a small number of participants.

Finally, this Protocol requires that the energy use of participants or households affected by the program (for the treatment and control groups) can be clearly identified and measured. Typically, the analysis unit is the household; in this case, treatment group households must be identifiable and individual household energy consumption must be metered. However, depending on the BB program, the analysis units may not be households. For example, for a BB program that generates an energy competition between housing floors or residential buildings at a university, the analysis unit may be floors or buildings; in this case, the energy consumption of these units must be metered.

The characteristics of BB programs that do not determine the applicability of the evaluation protocol include:

- Whether the program is opt-in or opt-out[66]
- The specific behavior-modification theory or strategy
- The channel(s) through which program information is communicated.

Although this protocol strongly recommends RCTs or REDs, it also recognizes that implementing these methods may not always be feasible. Government regulations or program designs may prevent the utilization of randomized experiments for evaluating BB programs. In these cases, evaluators must employ quasi-experimental methods, which require stronger assumptions than do randomized experiments to yield valid savings estimates.[67] If these assumptions are violated, quasi-experimental methods may produce biased results. The extent of the biases in the estimates is not knowable *ex ante*, so results will be less reliable. As more evidence accumulates about the efficacy of quasi-experiments, NREL may update this protocol as appropriate.

2.7.2.1 Examples of Protocol Applicability

Examples of residential BB programs for which the evaluation protocol applies follow:

- Example 1. A utility sends energy reports encouraging conservation to thousands of randomly selected residential customers.
- Example 2. A utility sends email or text alerts to residential customers with tips about reducing energy consumption when their energy consumption is on track to exceed normal levels for the billing period.
- Example 3. A utility invites thousands of residential customers to use its web portal to track their energy consumption in real time, set goals for energy saving, find ideas about how to reduce their energy consumption, and receive points or rewards for saving energy.
- Example 4. A utility sends voice, text, and email messages to thousands of residential utility customers encouraging—and providing tips for— reducing energy consumption during an impending peak demand event.

[66] In opt-in programs, customers enroll or select to participate. In opt-out programs, the utility enrolls the customers, and the customers remain in the program until they opt out. An example opt-in program is having a utility web portal with home energy use information and energy efficiency tips that residential customers can use if they choose. An example opt-out program is sending energy reports to utility selected customers.

[67] For example, Harding and Hsiaw (2012) use variation in timing of adoption of an online goal-setting tool to estimate savings from the tool.

- Example 5. A utility uses a mass-media advertising campaign that relies on radio and other broadcast media to encourage residential customers to conserve energy.
- Example 6. A utility initiates a social media campaign (for example, using Facebook or Twitter) to encourage energy conservation.
- Example 7. A utility runs a pilot program to test the savings from in-home energy-use displays and enrolls too few customers to detect the expected savings.
- Example 8. A utility runs a BB program in a large, master-metered college dormitory to change student attitudes about energy use. The utility randomly assigns some rooms to the treatment group and other rooms to the control group.

The Protocol does not apply to Example 5 or Example 6 because the evaluator cannot identify who received the messages. This does not apply to Example 7 because too few customers are in the pilot to accurately detect energy savings. This does not apply to Example 8 because energy-use data are not available for the specific rooms assigned to the treatment and control groups.

2.7.3 SAVINGS CONCEPTS

The protocol recommends RCTs and REDs to develop unbiased and robust estimates of energy or demand savings from BB programs that satisfy the applicability conditions described in Section 2.7.2 Unless otherwise noted, all references in this protocol to savings are to net energy or demand savings.

Section 2.7.3.1 defines some key concepts and 2.7.3.2 describes specific evaluation methods.

2.7.3.1 Definitions

The following key concepts are used throughout this protocol.

- Control group. In an experiment, the control group comprises subjects (for example, utility customers) who do not receive the program intervention or treatment.
- Experimental design. Randomized experiments rely on observing the energy use of subjects who were randomly assigned to program treatments or interventions in a controlled process.
- External validity. Savings estimates are externally valid if evaluators can apply them to different populations or time periods from those studied.
- Internal validity. Savings estimates are internally valid if the savings estimator is expected to yield an estimate of the causal effect of the program on consumption.
- Opt-in program. A program in which customers must enroll themselves. Utilities use opt-in BB programs if the customers must agree to participate, and the utility cannot administer treatment without consent.
- Opt-out program. A program in which a utility can automatically enroll customers. Utilities use opt-out BB programs if the utility does not need prior agreement from the customer to participate. The utility can administer treatment without the customer’s consent, and customers remain enrolled until they ask the utility to stop the treatment.
- Quasi-experimental design. Quasi-experimental designs rely on a comparison group that is not obtained via random assignment. Such designs observe energy use and determine program treatments or interventions based on factors that may be partly random but not controlled.
- Randomized control trial. An RCT uses random variation in which subjects are exposed to the program treatment to obtain an estimate of the program treatment effect. By randomly assigning subjects to

treatment, an RCT controls for all factors that could confound measurement of the treatment effect. An RCT is expected to yield an unbiased estimate of program savings. Evaluators randomly assign subjects from a study population to a treatment group or a control group. Subjects in the treatment group receive one program treatment (there could be multiple treatments and treatment groups), whereas subjects in the control group receive no treatment. The RCT ensures that receiving the treatment is uncorrelated with the subjects' pretreatment 6 energy use, and that evaluators can attribute any difference in energy use between the groups to the treatment.

- Randomized encouragement design. In a RED, evaluators randomly assign subjects to a treatment group that receives encouragement to participate in a program or to a control group that does not receive encouragement. The RED yields unbiased estimates of the effect on energy consumption from the encouragement and the effect on energy consumption from participating in the program for subjects who participated because of the encouragement. This latter estimate is known as the local average treatment effect (LATE).
- Treatment. A treatment is an intervention administered through the BB program to subjects in the treatment group. Depending on the research design, the treatment may be a program intervention or encouragement to accept an intervention.
- Treatment effect. This is the effect of the BB program intervention(s) on energy consumption for a specific population and time period. The treatment effect may persist after the period in which the intervention is administered. This means that for long-running programs, some savings may be attributable to treatments administered in previous periods. Section B.6.6.1 of this protocol addresses BB program savings persistence and measure life.
- Treatment group. The experimental group of subjects who received the treatment.

2.7.3.2 Randomized Experimental Research Designs

This section outlines the application of randomized experiments for evaluating BB programs. The most important benefit of an RCT or RED is that, if carried out correctly, the experiment results in an unbiased estimate of the program's causal impact. Unbiased savings estimates have internal validity. A result is internally valid if the evaluator can expect the value of the estimator to equal the savings caused by the program intervention. The principal threat to internal validity in BB program evaluation derives from potential selection bias about who receives a program intervention. RCTs and REDs yield unbiased savings estimates because they ensure that receiving the program intervention is uncorrelated with the subjects' energy consumption.

Randomized experiments may yield savings estimates that are applicable to other populations or time periods, making them externally valid. Whether savings have external validity will depend on the specific research design, the study population, and other program features. Program administrators should exercise caution in applying BB program savings estimates for one population to another or to the same population at a later time, because differences in population characteristics, weather, or naturally occurring efficiency can cause savings to change.

A benefit of randomized field experiments is their versatility: evaluators can apply them to a wide range of BB programs regardless of whether they are opt-in or opt-out programs. Evaluators can apply randomized experiments to any program where the objective is to achieve energy or demand savings; evaluators can construct an appropriately sized analysis sample; and accurate measurements of the energy consumption of sampled units can be obtained.

Randomized experiments, particularly those with large sample sizes, yield highly robust savings estimates that are not model dependent; that is, they do not depend on the specification of the model used for estimation.

The choice of whether to use an RCT or RED to evaluate program savings should depend on several factors, including whether it is an opt-in or opt-out program, the expected number of program participants, and the utility's tolerance for subjecting customers to the requirements of an experiment. For example, using an RCT for an opt-in program might require delaying or denying participation for some customers. A utility may prefer to use a RED to accommodate all the customers who want to participate.

Implementing an RCT or RED design requires upfront planning. Program evaluation must be an integral part of the program planning process, as described in the randomized experiment research design descriptions in Section 2.7.3.3 *Basic Features.*

2.7.3.3 Basic Features

This section outlines several types of RCT research designs, which are simple but very powerful research tools. The core feature of RCT is the random assignment of study subjects (for example, utility customers, floors of a college dormitory) to a treatment group that receives or experiences an intervention or to a control group that does not receive the intervention.

Section 2.7.3.3.1 *Common Features of Randomized Control Trial Designs* outlines some common features of RCTs and discusses specific cases.

2.7.3.3.1 Common Features of Randomized Control Trial Designs

The key requirements of an RCT are incorporated into the following steps:

1. Identify the study population. The program administrator screens the utility population if the program intervention is only offered to certain customer segments, such as single family homes. Programs designers can base eligibility on dwelling type (for example, single family, multifamily), geographic location, completeness of recent billing history, heating fuel type, utility rate class, or other energy use characteristics.
2. Identify the treatment effect the experiment will measure and the measurement approach. Is the BB treatment designed to reduce peak demand, energy consumption, or both? For what periods will savings be measured? A year? Each month of the year or the sample? Hour of the day?
3. Determine sample sizes. The numbers of subjects to assign to the treatment and control groups should depend on the type of randomized experiment (for example, REDs and opt-out RCTs generally require more customers), the hypothesized savings, the variance of consumption, and tolerance for error. The number of subjects assigned to the treatment versus control groups should be large enough to detect the hypothesized program effect with the required probability (the statistical power of the experiment), though it is not necessary for the treatment and control groups to be equally sized. Furthermore, some jurisdictions or program administrators may require savings estimates to achieve certain levels of confidence and precision such as 90% confidence with +/-10% precision. An experiment may have sufficient statistical power, but not yield estimates that meet the required confidence and precision.

obtain minimum sample sizes for the treatment and control groups that meet the desired level of statistical power and confidence and precision. Program administrators and regulators should specify requirements for statistical significance and/or confidence and precision before a program is designed so evaluators can size the experiment appropriately. It is not uncommon for BB programs with expected savings of less than 3% to require thousands of subjects in the treatment and control groups.

4. Randomly assign subjects to treatments and control. Study subjects should be randomly assigned to treatment and control groups. To maximize the credibility and acceptance of BB program evaluations by regulators and program administrators, this protocol recommends that a qualified independent third party perform the random assignment. Also, to preserve the integrity of the experiment, customers must not choose their assignments. The procedure for randomly assigning subjects to treatment and control groups should be transparent and well documented.
5. Verify equivalence. An important component of the random assignment process is to verify that the treatment and control groups are statistically equivalent or balanced in their observed covariates. At a minimum, evaluators should verify that before the intervention there are no statistically significant differences between treatment and control homes in average pretreatment energy consumption and in the distribution of pretreatment energy use. Evaluators should conduct analogous tests using customer demographic and housing characteristics data if such data are available.
6. Administer the treatment. The intervention must be administered to the treatment group and withheld from the control group. To avoid a Hawthorne effect, in which subjects change their energy use in response to observation, control group subjects should receive minimal information about the study. Depending on the research subject and intervention type, the utility may administer treatment once or repeatedly and for different durations. However, the treatment period should be long enough for evaluators to observe any effects of the intervention.
7. Collect data. Data must be collected from all randomized study subjects, not only from those who chose to participate or only from those who participated for the whole study or experiment. Preferably, evaluators should collect multiple pre and post-treatment energy consumption measurements. Such data enable the evaluator to control for time-invariant differences in average energy use between the treatment and control groups to obtain more precise savings estimates. Step 8 discusses this in further detail.
8. Estimate savings. Evaluators should calculate savings as the difference in energy consumption or difference in differences (DiD) of energy consumption between the subjects who were initially assigned to the treatment and those assigned to the control group. To obtain an unbiased savings estimate, evaluators must compare the energy consumption from the entire group of subjects who were originally randomly assigned to the treatment group to the entire group of subjects who were originally randomly assigned to the control group. For example, the savings estimate would be biased if evaluators used only data from utility customers in the treatment group who chose to participate in the study. The difference in energy consumption between the treatment and control groups, usually called an intent-to-treat (ITT) effect, is an unbiased estimate of savings because subjects were randomly assigned to the treatment and control groups. The effect is an ITT because, in contrast to many randomized clinical medical trials, ensuring that treatment group subjects in most BB programs comply with the treatment is impossible. For example, some households may opt out of an energy reports program, or they may fail to notice or open the energy reports. Thus, the effect is ITT, and the evaluator should base the results on the initial assignment of subjects to the treatment group, whether or not subjects actually complied with the

treatment. The savings estimation should be well documented, transparent, and performed by an independent third party.

2.7.3.4 Common Designs

This section describes some of the RCT designs commonly used in BB programs.

2.7.3.4.1 Randomized Control Trial with Opt-Out Program Design

One common type of RCT includes the option for treated subjects to opt out of receiving the program treatment. This design reflects the most realistic description of how most BB programs work. For example, in energy reports programs, some treated customers may ask the utility to stop sending them reports.

Figure 2-9 depicts the process flow of an RCT in which treated customers can opt out of the program. In this illustration, the utility initially screened its customers to refine the study population.

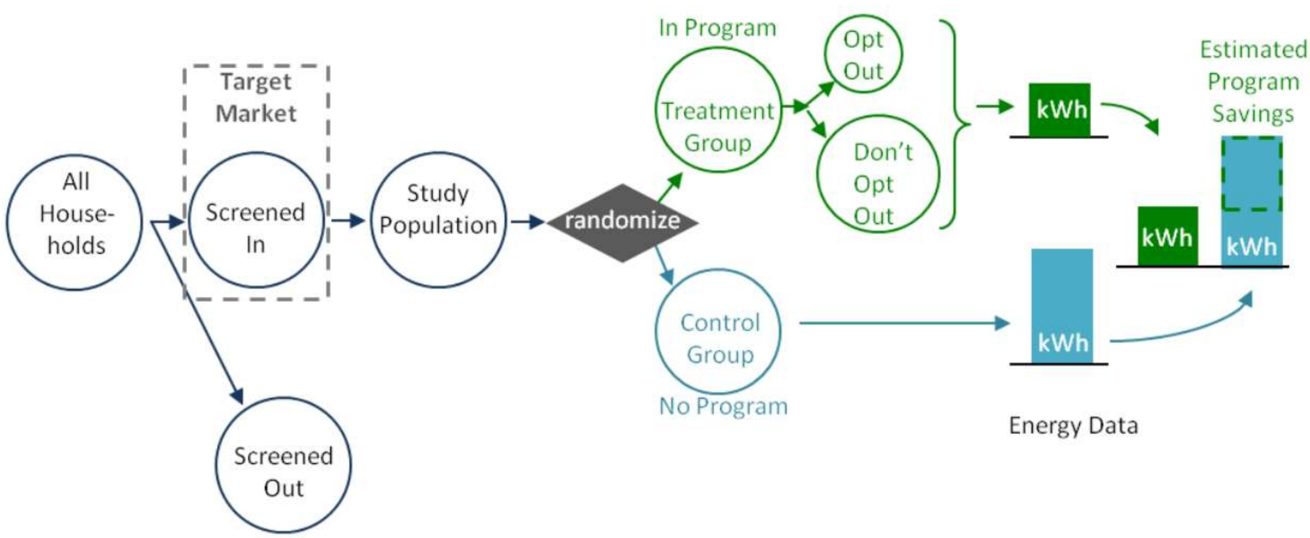

Figure 2-9 Illustration of RCT with Opt-Out Program Design[68]

Customers who pass the screening comprise the study population or sample frame. The ITT savings estimate will apply to this population. Alternatively, the utility may want to study only a sample of the screened population, in which case customers from the population should be sampled randomly. The analysis sample must be large enough to meet the minimum required sizes for the treatment and control groups.

The next steps in an opt-out RCT are to (1) randomly assign subjects in the study population to the program treatment and control groups, (2) administer the program treatments, and (3) collect energy use data.

The distinguishing feature of this randomized experimental design is that customers can opt out of the program. As Figure 2-9 shows, evaluators should include opt-out subjects in the energy savings analysis to obtain unbiased savings estimates. Evaluators can then calculate savings as the difference in average energy consumption between treatment group customers, including optout subjects and control group customers. Removing opt-out subjects from the analysis would bias the savings estimate because certain subjects in the control group would

[68] UMP Chapter 17, 2014

have opted out if they had been treated but it is impossible to know who that might be in the control group. The resulting savings estimate is therefore an average of the savings of treated customers who remain in the program and of customers who opted out.

Depending on the type of BB program, the percentage of customers who opt out may be small and opt outs may not affect the savings estimates significantly (for example, few customers opt out of energy reports programs).

2.7.3.4.2 Randomized Control Trial with Opt-In Program Design

Some BB programs require utility customers to enroll before they can be treated. Examples include web-based home audit or energy consumption tools; online courses about energy rates and home efficiency; or in-home displays. These interventions contrast with interventions such as home energy reports that can be administered to subjects without having their prior agreement.

An opt-in RCT (Figure 2-10) can accommodate the necessity for customers to opt into some BB programs. This design results in an unbiased estimate of the ITT effect for customers who opt into the program. The estimate of savings will have internal validity; however, it will not necessarily have external validity because it will not apply to subjects who do not opt in.

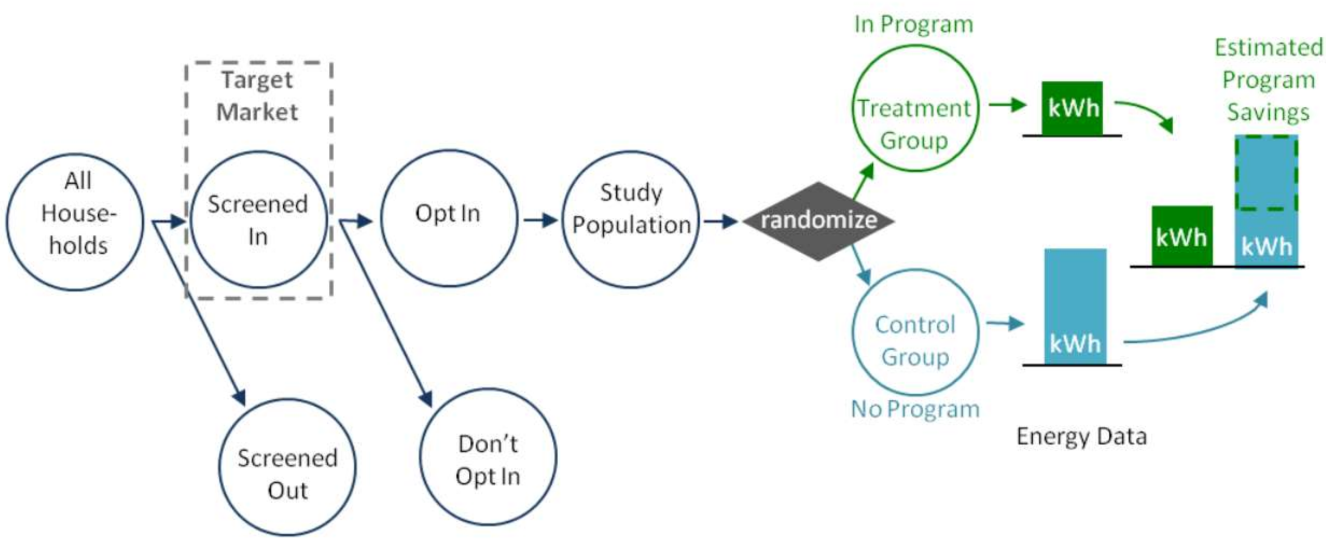

Figure 2-10 Illustration of RCT with Opt-In Program Design[69]

Implementing opt-in RCTs is very similar to implementing opt-out RCTs. The first step, screening utility customers for eligibility to determine the study population, is the same. The next step is to market the program to eligible customers. Some eligible customers may then agree to participate. Then, an independent third party randomly assigns these customers to either a treatment group that receives the intervention or a control group that does not. The utility delays or denies participation in the program to customers assigned to the control group. Thus, only customers who opted in and were assigned to the treatment group will receive the treatment.

[69] *Ibid.*

Randomizing only opt-in customers ensures that the treatment and control groups are equivalent in their energy use characteristics. In contrast, other quasi-experimental approaches, such as matching participants to nonparticipants, cannot guarantee either this equivalence or the internal validity of the savings estimates.

After the random assignment, the opt-in RCT proceeds the same as an RCT with opt-out subjects: the utility administers the intervention to the treatment group. The evaluator collects energy consumption data from the treatment and control groups, then estimates energy savings as the difference in energy consumption between the groups. The evaluator does not collect energy consumption data for customers who do not opt into the program.

An important difference between the opt-in RCT and opt-out RCT is how to interpret the savings estimates. In the opt-out RCT, the evaluator bases the savings estimate on a comparison of the energy consumption between the randomized treatment and control groups, which pertains to the entire study population. In contrast, in the opt-in RCT, the savings estimate pertains to the subset of customers in the study population who opted into the program, and the difference in energy consumption represents the treatment effect for customers who opted into the program. Opt-in RCT savings estimates have internal validity; however, they do not apply to customers who did not opt into the program.

2.7.3.4.3 Randomized Encouragement Design

Some BB interventions require participants to opt into treatment but delaying or denying participation to some customers may be undesirable. In this case, neither the opt-out nor the opt-in RCT design would be appropriate, and this protocol recommends an RED. Instead of randomly assigning subjects to receive or not receive the intervention, a third party randomly assigns them to a treatment group that is encouraged to accept the intervention (that is, to participate in a program or adopt a measure), or to a control group that does not receive encouragement. Examples of common kinds of encouragement include direct paper mail or email informing customers about the opportunity to participate in a BB program. Customers who receive the encouragement can refuse to participate, and, depending on the program design, control group customers who learn about the program may be able to participate.

The RED yields an unbiased estimate of the effect of encouragement on energy consumption and, depending on the program design, can also provide an unbiased estimate of either the effect of the intervention on customers who accept the intervention because of the encouragement or the effect of the intervention on all customers who accept it. Necessary conditions for a RED to produce an unbiased estimate of savings from the BB intervention is that the encouragement only influence the decision to accept the BB intervention and not energy consumption. For example, the RED must be such that customers who receive a direct mailing encouraging them to log into a website with personalized energy efficiency recommendations only save energy if they decide to log into the site; the mailing itself must not cause the customer to save energy if the customer never logs on. If the encouragement causes customers to save energy, it may be impossible to isolate the savings from the intervention. Programs designed as REDs should design and distribute encouragement materials that will not affect consumption. If evaluators expect that the encouragement will cause energy savings, they could send the similar messaging that excludes the program enrollment option to the control group or to a second randomized treatment group. Evaluators could use the second randomized treatment group to test whether the encouragement produces savings.

Figure 2-11 illustrates the process flow for a RED program evaluation. As with the RCT with optout and opt-in RCT, the first two steps are to identify the sample frame and select a study population. Next, like the RCT with opt out, a third party randomly assigns subjects to a treatment group, which receives encouragement, or to a control group, which does not. For example, a utility might employ a direct mail campaign that encourages treatment group customers to use an online audit tool. The utility would administer the intervention to treatment group customers who opt in. Although customers in the control group did not receive encouragement, some may learn about the program and decide to sign up. The program design shown in Figure 2-11 allows for control group customers to receive the behavioral intervention.

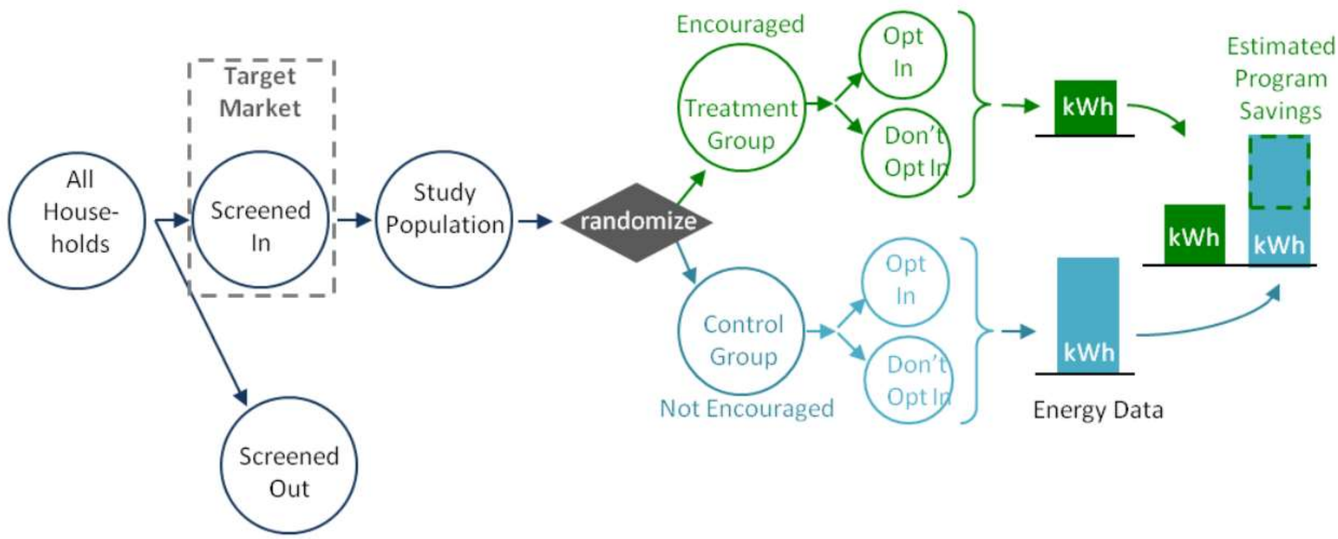

Figure 2-11 Illustration of RED Program Design[70]

In Figure 2-11, the difference in energy consumption between homes in the treatment and control groups is an estimate of savings from the encouragement, not from the intervention. However, evaluators can also use the difference in energy consumption to estimate savings for customers who accept the intervention because of the encouragement. To see this, consider that the study population comprises three types of subjects: (1) always takers, or those who would accept the intervention whether encouraged or not; (2) never takers, or those who would never accept the intervention even if encouraged; and (3) compliers, or those who would accept the intervention only if encouraged. Compliers participate only after receiving the encouragement.

Because eligible subjects are randomly assigned to groups depending on whether they receive encouragement, the treatment and control groups are expected to have equal frequencies of always takers, never takers, and compliers. After treatment, the only difference between the treatment and control groups is that compliers in the treatment group accept the treatment and compliers in the control group do not. In both groups, always takers accept the treatment and never takers always refuse the treatment. Therefore, the difference in energy consumption between the groups reflects the treatment effect of encouragement on compliers (known as LATE).

Furthermore, for the study to have enough statistical power to detect the expected effect, there must be very large encouraged and non-encouraged groups relative to a RCT or quasi-experimental design and/or a high proportion of compliers in the treatment group; a power calculation should be done to ensure that there are

[70] *Ibid.*

enough customers in the encouraged and non-encouraged groups to produce significant savings estimates for the expected take-up rate.[71]

To estimate the effect of the intervention on compliers, evaluators can either employ instrumental variables (IV), using the random assignment of customers to receive encouragement as an instrument for the customer's decision to accept the intervention (that is, to participate).

The IV approach is presented in Section 2.7.4.4.9 *Randomized Encouragement Design*. Another option is that evaluators can scale the treatment effect of the encouragement by the difference between treatment and control groups in the percentage of customers who receive the intervention (note that in this equation, if the non-encouraged customers are not allowed to take up the treatment, the second term in the denominator will be zero)[72].

$$\frac{1}{(\%\text{ of encouraged customers who accepted} - \%\text{ of nonencouraged customer who accepted})}$$

If customers in the control group are permitted to participate if they find out about the treatment even though they did not receive encouragement, the LATE does not capture the program effect on always takers. (Note, however, in most programs, the control group is not permitted to take up the treatment). If customers in the control group are permitted to participate, the LATE may differ from the average treatment effect unless the savings from the intervention is the same for compliers and always takers. However, the LATE will be equal to the average treatment effect if the control group customers (non-encouraged customers) are not permitted to take up the treatment.

For BB programs with REDs that do not permit control group customers to participate, evaluators can estimate the treatment effect on the treated (TOT). The TOT is the effect of the program intervention on all customers who accept the intervention. In this case, the difference in energy use between the treatment and control groups reflects the impact of the encouragement on the always takers and compliers in the treatment group. Scaling the difference by the inverse of the percentage of customers who accepted the intervention yields an estimate of the TOT impact.[73]

Successful application of a RED requires that compliers comprise a percentage of the encouraged population that is sufficiently large given the number of encouraged customers. If the RED generates too few compliers, the effects of the encouragement and receiving the intervention will not be precisely estimated. Therefore, before employing a RED, evaluators should ensure that the sample size is sufficiently large, and that the encouragement will result in the required number of compliers. If the risk of a RED generating too few compliers is significant, evaluators may want to consider alternative approaches, including quasi-experimental methods.

[71] For an example of a power calculation for REDs, see Fowlie (2010).

[72] This approach of estimating savings from the intervention because of encouragement assumes zero savings for customers who received encouragement but did not accept the intervention. If encouraged customers who did not accept the intervention reduced their energy use in response to the encouragement, the savings estimate for compliers will be biased upward.

[73] If the effect of program participation is the same for compliers as for others, those who would have participated without encouragement (always takers) and those who do not participate (never takers), then the RED will yield an unbiased estimate of the population average treatment effect.

2.7.3.5 Randomized Experiments Implementation Requirements and Evaluation Guidance

This protocol strongly recommends the use of randomized field experiments (RCTs or REDs) for evaluating residential BB programs. Table 2-23 summarizes the benefits and requirements of evaluating BB programs using RCTs and REDs, as described in Sections 2.7.3.1 *Definitions* through 2.7.3.4 *Common Designs*.

Table 2-23 Benefits and Implementation Requirements of Randomized Experiments[74]

Evaluation Benefits	Implementation Requirements
▪ Yield unbiased, valid estimates of causal program impacts, resulting in a high degree of confidence in the savings ▪ Yield savings estimates that are robust to changes in model specification ▪ Are versatile, and can be applied to opt-out and opt-in BB programs ▪ Are widely accepted as the "gold standard" of good program evaluations ▪ Result in transparent and straightforward analysis and evaluation ▪ Can be designed to test specific research questions such as persistence of savings after treatment ends	▪ An appropriately sized analysis sample ▪ Accurate energy use measurements for sampled units ▪ Advance planning and early evaluator involvement in program design ▪ Restricted participation or program marketing to randomly selected customers

The principal benefit of randomized experiments is that they yield unbiased and robust estimates of program savings or other treatment effects. They are also versatile, widely accepted, and straightforward to analyze. The principal requirements for implementing randomized experiments include the availability of accurate energy consumption measurements and a sufficiently large study population.

Also, this protocol specifically recommends REDs or RCTs for estimating BB program savings as both designs yield unbiased savings estimates. The choice of RED or RCT will depend primarily on program design and implementation considerations whether the program has an opt-in or opt-out design. RCTs work well with opt-out programs such as residential energy reports programs. Customers who do not want to receive reports can opt out without adversely affecting the evaluation. RCTs also work well with opt-in programs, for which customer participation can be delayed (for example, customers are put on a "waiting list") or denied. For situations in which delaying or denying a certain subset of customers is impossible or costly, REDs may be more appropriate. REDs can accommodate all interested customers, but have the disadvantages of requiring larger analysis samples, two analysis steps to yield a direct estimate of the behavioral intervention's effect on energy use, and a high proportion of compliers among encouraged customers.

Table 2-24 lists some issues to consider when choosing a RCT or RED.

[74] *Ibid.*

Table 2-24 Considerations in Selecting a Randomized Experimental Design

Experimental Design	Evaluation Benefits	Implementation and Evaluation Requirements
RCT	▪ Yields unbiased, robust, and valid estimates of causal program impacts, resulting in a high degree of confidence in the savings ▪ Simple to understand ▪ Works well with opt-out programs ▪ Works well with opt-in programs if customers can be delayed or denied	▪ May require delaying or denying participation of some customers if program requires customers to opt in
RED	▪ Yields unbiased, robust, and valid estimates of causal program impacts, resulting in a high degree of confidence in the savings ▪ Can accommodate all customers interested in participating ▪ Works well with opt-in and opt-out programs	▪ More complex design and harder to understand ▪ Requires a more complex analysis ▪ Requires larger analysis sample

2.7.3.6 Quasi-Experimental Methods

There are other evaluation design methods that use non-randomized control groups, called quasi-experimental methods. With these methods, the control group is not randomly assigned. Thus, quasi-experimental methods often suffer from selection bias and may produce biased estimates of energy savings. However, in specific cases in which RCTs are not feasible, quasi-experimental approaches can still meet the acceptability criteria recommended in this report, although the results they generate will be less reliable. These methods are discussed in Section 2.7.4.4.10 *Quasi-Experimental Methods*.

2.7.4 SAVINGS ESTIMATION

Energy savings for a household in a BB program is the difference between the energy the household consumed and the energy the household would have consumed if it had not participated. However, the energy consumption of a household cannot be observed under two different states. Instead, to estimate savings, evaluators should compare the energy consumption of households in the treatment group to that of a group of households that are statistically the same but did not receive the treatment. In a randomized experiment, assignment to the treatment is random; thus, evaluators can expect control group subjects to consume, on average, the same amount of energy that the treatment group would have consumed without the treatment. The difference in their energy consumption will therefore be an unbiased estimate of energy savings.

Savings can be estimated using energy consumption data from the treatment period only or from before and during the treatment. If energy consumption data from only the treatment period are used, evaluators estimate the savings as a simple difference. If data on energy consumption before treatment is administered are available, evaluators can estimate the savings as a DiD or a simple difference that controls for pretreatment energy consumption. The approach that estimates savings conditional on pretreatment consumption is sometimes referred to as a "post-only model with pre-period controls."[75] The availability of energy consumption data for the period before the treatment will determine the approach but incorporating pretreatment consumption data in the analysis is strongly advised when such data are available.

Both approaches result in unbiased estimates of savings (that is, in expectation, the two methods are expected to yield an estimate equal to the true savings). However, estimators using pretreatment data generally result in more precise savings estimates (that is, the estimators using pretreatment data will have a smaller standard error) as they account for time-invariant energy use that contributes significantly to the variance of energy consumption between utility customers.[76]

Evaluators should collect at least one full year of historical energy use data (the 12 months immediately before the program start date) to ensure baseline data fully reflect seasonal energy use effects.

Regulators usually determine the frequency of program evaluation. Although requirements vary between jurisdictions, most BB programs are evaluated once per year. Annual evaluation will likely be necessary for the first several years of many BB programs such as HER programs because savings tend to increase for several years before leveling off. However, some program administrators may desire measurement or evaluation more frequently than annually to closely track program performance and optimize the program delivery.

2.7.4.1 IPMVP Option

This protocol's recommended evaluation approach aligns best with IPMVP Option C, which recommends statistical analysis of data from utility meters for whole buildings or facilities to estimate savings. Option C is intended for projects with expected savings that are large relative to consumption. This protocol recommends regression analysis of residential customer consumption and statistical power analysis to determine the analysis sample size necessary to detect the expected savings.

2.7.4.2 Sample Design

Utilities should integrate the design of the analysis sample with program planning, because numerous considerations, including the size of the analysis sample, the method of recruiting customers to the program, and the type of randomized experiment, must be addressed before the program begins.

2.7.4.2.1 Sample Size

The analysis sample should be large enough to detect the minimum hypothesized program effect with desired probability.[77] If the sample is too small, evaluators risk being unable to detect the program's effect and possibly wrongly accepting a hypothesis of no effect or there may be substantial uncertainty about the program's effect

---

[75] The model with pretreatment consumption control variables is a more efficient estimator (that is, it is expected to have smaller variance) than the DiD estimator when the model errors are independent and identically distributed or when serial correlation of consumption is low (Burlig, Preonas, and Woerman 2017). This model is more efficient because it uses one degree of freedom rather than multiple degrees of freedom—one for each study subject—to account for between-subject differences in consumption. However, when serial correlation of customer consumption is high, there is little or no gain in efficiency over the fixed effects in the DiD approach.

[76] Postonly or DiD estimation with customer fixed effects also accounts for differences in mean energy use between treatment and control group subjects that are introduced when subjects are randomly assigned to the treatment or control group. Evaluators may not expect such differences with random assignment; however, these differences may nevertheless arise.

[77] A program can comprise a collection of randomized cohorts or waves in which the treatment effect of interest is at the program level and not at the level of individual cohorts. In this case, power calculations and tests of statistical significance can be applied to the collection of cohorts. Examples of this design include behavioral programs that consist of several waves launched over time or rolling enrollment waves.

the sample size is too large, researchers may risk wasting scarce program resources[78]. Oversizing the sample is primarily a concern for pilot programs, for which determining the savings is often a primary objective.

To determine the minimum number of subjects required and the number of subjects to be assigned to the treatment and control groups, researchers should employ a statistical power analysis. Statistical power is the likelihood of detecting a program impact of minimum size (the minimum detectable effect). Typically, researchers design studies to achieve statistical power of 80% or 90%. A study with 80% statistical power has an 80% probability of detecting the hypothesized treatment effect.

Statistical power analysis can be conducted in two ways. First, if data on consumption or another outcome of interest before treatment are available for the study population, researchers can use simulation to estimate the probability of detecting an effect of a certain size (for example, 1%) for possible treatment and control groups sizes, NT and NC.

Simulation follows these steps:

- Researchers should divide the pretreatment sample period into two parts, corresponding to a simulation pretreatment and post-treatment period. For example, an evaluator with monthly billing consumption data for 24 pretreatment months could divide the pretreatment period into months 1 to 12 and months 13 to 24 and designate the first 12 months as the simulation pretreatment period.
- From the eligible program population, researchers should randomly assign NT subjects to the treatment group and NC subjects to the control group.
- Researchers should decide upon the minimum detectable treatment effect (for example, 2 kWh/period/subject), and a distribution of treatment effects (for example, normal distribution with mean 2 and standard deviation 1). For each treatment customer, the researcher should simulate the program treatment effect, taken randomly from the distribution of treatment effects, during the simulation treatment period. (One could also assume the treatment effect is the same for all customers and merely apply the same effect to all households; however, the power calculation is likely to underestimate the number of households needed because it assumes zero variance for the treatment effect).
- Researchers should randomly sample with replacement NT customers from the treatment group and NC subjects from the control group.
- Researchers should estimate the program treatment effect for the sample only using data from the simulation pretreatment and simulation post-treatment periods and record the estimate and whether the estimate was statistically significant for a given Type 1 error.
- Researchers should repeat steps 4 and 5 many times (for example, >250), and calculate the percentage of iterations when the estimated treatment effect was statistically different than zero. This is the statistical power of the study, the probability of detecting savings of x with treatment group size NT and control group size NC.

It is important that the estimation method used in the statistical power simulation adheres as closely as possible to the method evaluators plan to use for the actual savings estimation. Otherwise, the statistical power analysis may be misleading about the likelihood of detecting the savings.

[78] The utility may also base the number of subjects in the treatment group on the total savings it desires to achieve.

The second approach to calculating statistical power uses analytic formulas. Researchers employing panel data methods and using statistical power formulas are advised to use the formulas in Burlig et al. (2017). Though more demanding to implement than those in Frison and Pocock (1992), the statistical power formulas in Burlig et al. (2017) are more accurate because they account for both intracluster correlations and arbitrary serial correlations of customer consumption over time. The required inputs for the power calculation are:

- The minimum detectable treatment effect
- The coefficient of variation of energy use, taken from a sample of customers
- The specific analysis approach to be used (for example, simple differences of means or a repeated measure analysis)
- The numbers of pretreatment and post-treatment observations per subject
- The tolerances for Type I and Type II statistical errors (as discussed in Section 2.7.4.3.3 *Other Data Requirements*)
- The intracluster correlation of an individual subject's energy use or error term covariances for pretreatment and post-treatment periods and between periods.

Many statistical software applications, including SAS, STATA, and R, include packages for performing statistical power analyses.

Researchers conducting statistical power analyses should keep in mind the following:

- For a given program population, statistical power will be maximized if 50% of subjects are assigned to the treatment group and 50% are assigned to the control group. However, especially for large programs, researchers may obtain acceptable levels of statistical power with unbalanced treatment and control groups. The principal benefit of a smaller control group is that more customers are available to participate in the program.
- If the BB program will operate for more than several months and repeated measurements are planned, researchers should adjust the required sample sizes to account for attrition (the loss of some subjects from the analysis sample because of account closures or withdrawal from the study).

Finally, many studies will not estimate statistically significant savings. This null result could mean that the program did not save energy or that the evaluation did not detect the savings. During the program and evaluation design phase, if clear guidelines are not already available, program administrators, regulators, and evaluators should reach agreement about how statistically insignificant savings estimates should be treated and reported and whether all or some savings based on such estimates can be claimed.

2.7.4.2.2 Random Assignment to Treatment and Control Groups by an Independent Third Party

After determining the appropriate sizes of the treatment and control group samples, researchers should randomly assign subjects to the treatment and control groups. For the study to have maximum credibility and acceptance, this protocol recommends that an independent and experienced third party, such as an independent evaluator, perform the randomization. If there is a significant risk that the random assignment will result in unbalanced treatment and control groups with statistically different consumption, this protocol recommends that evaluators first stratify the study population by pretreatment energy consumption levels and then randomly assign subjects in each stratum to treatment and control groups. Stratifying the sample will increase the likelihood that treatment and control group subjects have similar pretreatment means and variances.

This protocol also recommends that the unit of analysis (for example, a household) should be the basis for random assignment to treatment or control group. For example, in an analysis of individual customer consumption, it is better to randomly assign individual customers instead of all customers in the same neighborhood (for example, in a zip code or census block) to receive the treatment. However, for some BB programs, it may not be feasible to randomize the unit of analysis. For example, in some multifamily housing BB programs, the unit of analysis may be individual customers but all customers in the same multifamily building may receive the treatment. In this case, it will be necessary to randomly assign multifamily buildings to the treatment or control group. In this case, researchers will need to account for correlations in consumption between customers in the same housing units.

Although this protocol recommends that an independent and experienced third party perform the random assignment, circumstances sometimes make this impossible. In such cases, a third-party evaluator should verify that the assignment of treatment and control group subjects was done correctly and did not introduce bias into the selection process.

2.7.4.2.3 Equivalency Check

The third party performing the random assignment must verify that the characteristics of subjects in the treatment group, including pretreatment energy consumption, are balanced with those in the control group. If subjects in the groups are not equivalent, the energy savings estimates may be biased. Evaluators should perform two equivalency checks: (1) for all customers who were randomly assigned to the treatment and control groups; and (2) for all randomized customers who remain in the analysis sample after data cleaning and preparation are completed. Ideally, the consumption data used for the equivalency checks should cover 12 months preceding the start treatment and equivalency should be checked for each month of the year.

To verify the equivalence of energy consumption, this protocol recommends that the third-party test for differences between treatment and control group subjects in both the mean pretreatment period energy consumption and in the distribution of pretreatment energy consumption. Evaluators should attempt to verify equivalence of energy consumption using the same frequency of data to be used in the savings analysis. For example, evaluators should use hour interval consumption data to verify equivalence if the study objective is to estimate peak hour energy savings. Evaluators should also test for differences in other available covariates, such as energy efficiency program participation, home floor area, heating fuel type, and customer demographics. These tests can be used to further demonstrate that the treatment and control groups are well-balanced, as would be expected if assignment to treatment or control group was random. Evaluators can use t-tests or the following regression equation of energy consumption to verify the randomization.

Suppose the evaluator has monthly billing consumption data for all treatment and control group customers for the 12 months, m, m=1, 2, … 12, before treatment began.

$$y_{im} = \sum_{m=1}^{12} \beta_{1m} \times Tr_i + \mu_m + \varepsilon_{it}$$

Where:

y_{im} = The metered energy consumption of subject i in month m

β_{1m} = The average difference in daily energy consumption between the treatment and control groups in month m of the pretreatment period (2)

Tr_i = An indicator for whether subject i was randomly assigned to receive the treatment; the variable equals 1 for subjects in the treatment group and equals 0 for subjects in the control group

μ_m = A month-year fixed effect; the model controls for the month-year fixed effects with a separate intercept for each month, which represents the average daily consumption of the control group in month m

ε_{it} = The model error term, representing random influences on the energy use of customer i in month m.

In this simple model, the coefficient β1m provides an estimate of the difference in average daily consumption between the treatment and control group in month m of the pretreatment period. Because of the random assignment to treatment, it is expected that the differences will be close to zero and statistically insignificant. Ordinary least squares (OLS) estimation of this model will result in an unbiased estimate of β1m. The standard errors should be clustered on the customer or subject.[79]

Evaluators can check for differences in time-invariant (e.g., demographic or home) characteristics between treatment and control group customers by replacing the dependent variable with the time-invariant characteristics and replacing the month-year fixed effects with a constant β_0 and $\sum_{m=1}^{12}\beta_{1m} \times Tr_i$ with $\beta_{1m} \times Tr_i$. The coefficient β1 will measure the average difference between the treatment and control groups.

If significant differences are found, and it is possible to perform the random assignment again before treatment starts, the third party should consider doing so. Ideally, random assignment should not result in differences; however, differences occasionally appear, and it is better to redo the random assignment than to proceed with unbalanced treatment and control groups, which may lead to biased savings estimates.[80] As noted in Section 2.7.4.2.2 *Random Assignment to Treatment and Control Groups by an Independent Third Party*, stratifying the study population by pretreatment energy use will increase the probability that the groups are balanced.

If the evaluator is not the third party who performed the random assignment, they should perform an equivalency check before estimating the savings. The evaluator may be able to use statistical methods to control for differences in pretreatment energy consumption found after the program is underway.[81] This should be done whether the program is designed as an RCT or a quasi-experiment.

2.7.4.3 Data Requirements and Collection

2.7.4.3.1 Energy Use Data

Estimating BB program impacts using a field experiment requires collecting energy consumption data from subjects in the analysis sample. This protocol recommends that evaluators collect multiple energy consumption measurements for each sampled unit for the periods before and during the treatment.[82]

---

[79] Although the methods recommended in this protocol minimize the potential for violations of the assumptions of the classical linear regression model, evaluators should be aware of and take steps to minimize—potential violations. The clustering of standard errors accounts for the correlation of individual customer consumption across time periods. In general, it is incorrect to treat observations of a customer's consumption readings as being independent of one another.

[80] Evaluators should keep in mind that at a statistical significance level of 10%, it is expected that statistically significant differences from random assignment will be found 10% of the time as a result of random chance.

[81] If energy use data are available for the periods before and during the treatment, it is possible to control for timeinvariant differences between sampled treatment and control group subjects using subject fixed effects

[82] A single measurement of energy use for each sampled unit during the treatment period also results in an unbiased estimate of program savings. The statistical significance of the savings estimate depends on the variation of the true but unknown savings and the number of sampled units.

These data are known as a panel. Panels can comprise multiple hourly, daily, or monthly energy use observations for each sampled unit. In this protocol, a panel refers to a data set that includes energy measurements for each sampled unit either for the pretreatment and treatment periods or for the treatment period only. The time period for panel data collection will depend on the program timeline, the frequency of the energy consumption data, and the amount of such data collected.

Panel data have several advantages for use in measuring BB program savings:

- Relative ease of collection. Collecting multiple energy consumption measurements for each sampled unit from utility billing systems is usually easy and inexpensive.
- Can estimate savings during specific times. If the panel collects enough energy consumption observations per sampled unit, estimating savings at specific times during the treatment period may be possible. For example, hourly energy consumption data may enable the estimation of precise savings during utility system peak hours. Monthly energy consumption data may enable the development of precise savings estimates for each month of the year.
- Savings estimates are more precise. Evaluators can more precisely estimate energy savings with a panel because they may be able to control for the time-invariant differences in energy consumption between subjects that contribute to higher variance.
- Allows for smaller analysis samples. All else being equal, fewer units are required to detect a minimum level of savings in a panel study than in a cross-section analysis. Thus, collecting panel data may enable studies with smaller analysis samples and data collection costs.

Using panel data has some disadvantages relative to a single measurement per household in a cross-sectional analysis. First, evaluators must correctly cluster the standard errors within each household or unit (as described in the following section). Second, panel data generally require statistical software to analyze, whereas estimating savings using single measurements in a basic spreadsheet software program may be possible.

For the analyses of savings, we recommend using a panel data model that compares the change in energy use for the treatment group to the change in energy use for the control group, especially if the evaluation design is quasi-experimental.

This Protocol also recommends that evaluators collect energy consumption data for the duration of the treatment to ensure they can observe the treatment effect for the entire study period. Ideally, an energy efficiency BB program will last for a year or more because the energy end uses affected by BB programs may vary seasonally. For example, these programs may influence weather-sensitive energy end uses, such as space heating or cooling, so collecting less than one year of data may yield incomplete results. With these evaluation designs, failure to collect one year (twelve months) of historical data can result in severely biased estimates of energy savings that are imprecise and thus not advised. Quasi-experimental analysis specifications that use at least a year of baseline data are typically less biased because they control for pre-existing differences between the control and treatment groups. Below, Table 2-25 provides rule-of-thumb guidelines for length of baseline data collection for RCT and quasi-experimental design.

Table 2-25 Length of Baseline Period Recommendation[83]

If RCT	If Quasi-Experimental	Condition
Good	Good	12 months or more of historical data collected
Reasonable	Not Advisable	Less than 12 months of historical data collected
	Not Advisable	No historical data collected

2.7.4.3.2 Makeup of Analysis Sample

Evaluators must collect energy consumption measurements for every household or unit that is initially assigned to a control or treatment group, whether or not the household or unit later opts out. Not collecting energy consumption data for opt-out households will result in imbalanced treatment and control groups and could bias the savings estimates.

2.7.4.3.3 Other Data Requirements

Program information about each participant must also be collected. Evaluators will need to collect data on customer assignments to the treatment or control group and when the treatments began. Evaluators must have this information to accurately construct regression analysis model variables and to estimate savings. Also, depending on the research design and evaluation objectives, evaluators may also want to collect data on how many and when individual treatments were administered, if and when customers opted out, or details about the specific information included in the treatment. For example, evaluators will need information about the number of reports delivered to customers to estimate the impact of varying the number of delivered reports. Information about how many and which customers opted out may be helpful for evaluating opt-out behavior programs when the opt-out rate is high. The treatment effect for customers who received treatment (LATE) may be different than the ITT effect.

Temperature and other weather data may allow for more precise savings estimates but are often not necessary for estimating savings. Typically, researchers can use dummy variables for individual time periods to account for the effect of weather on household energy consumption. In a regression with time period fixed effects, weather data will improve the precision of the savings estimates only if there is significant variance between customers in weather. If weather data will be collected, evaluators should obtain them from the weather station nearest to each household.

2.7.4.3.4 Data Collection Method

Energy use measurements used in the savings estimation should be collected directly from the utility, not from the program implementer, at the end of the program evaluation period. Depending on the program type, utility billing system, and evaluation objectives, the data frequency can be at 15-minute, 1-hour, daily, or monthly intervals.

[83] If efficiency programs are designed to reduce usage only during a specific season (e.g., summer), then only historical and program year data from that season is necessary. However, comparing summer season measurements with winter season measurements of electricity load creates a situation where an incomplete year may produce significantly biased results or at least results that are difficult to interpret.

2.7.4.4 Analysis Methods

This protocol recommends using panel regression analysis to estimate savings from BB field experiments where subjects were randomly assigned to either treatment or control groups. Panel regression analysis is preferred to calculating savings differences of unconditional mean energy use, because regression results in more precise savings estimates. A significant benefit of randomized field experiments is that regression-based savings estimates are usually quite insensitive to the type of model specification.

Section 2.7.4.3.1 *Energy Use Data* addresses issues in panel regression estimation of BB program savings, including model specification and estimation, standard errors estimation, robustness checks, and savings estimation. It illustrates some specifications as well as the application of energy-savings estimation.

2.7.4.4.1 Panel Regression Analysis

In panel regressions, the dependent variable is usually the energy use of a subject (a utility customer home, apartment, or dormitory) per unit of time such a month, day, or hour. The right side of the equation includes an independent variable to indicate whether the subject was assigned to the treatment or control group. This variable can enter the model singularly or be interacted with another independent variable, depending on the analysis goals and the availability of energy use data from before treatment. The coefficient on the term with the treatment indicator is the energy savings per subject per unit of time. DiD models of energy savings must also include an indicator for whether the period occurred before or during the treatment period.

Many panel regressions also include fixed effects. Subject fixed effects capture unobservable energy consumption specific to a subject that does not vary over time. For example, home fixed effects may capture variation in energy consumption that is caused by differences such as home sizes or makeup of a home's appliance stock. Time-period fixed effects capture unobservable energy consumption specific to a time period that does not vary between subjects. Including time or subject fixed effects in a regression of energy consumption of subjects randomly assigned to the treatment or control group will increase the precision but not the expected unbiasedness of the savings estimates.[84]

Fixed effects can be incorporated into panel regression in several ways, as follows:

- Include a separate dummy variable or intercept for each subject in the model. The estimated coefficient on a subject's dummy variable represents the subject's time-invariant average energy use. This approach, known as least squares dummy variables, may, however, not be practical for evaluations with a large

[84] Standard econometric formulations assume that fixed effects account for unobservable factors that are correlated with one or more independent variables in the model. This correlation assumption distinguishes fixed-effects panel model estimation from other types of panel models. Fixed effects eliminate bias that would result from omitting unobserved time-invariant characteristics from the model. In general, fixed effects must be included to avoid omitted variable bias. In an RCT, however, fixed effects are unnecessary to the claim that the estimate of the treatment effect is unbiased because fixed effects are uncorrelated with the treatment by design. Although fixed effects regression is unnecessary, it will increase precision by reducing model variance.

Some evaluators may be tempted to use random-effects estimation, which assumes time- or subject-invariant factors are uncorrelated with other variables in the model. However, fixed-effects estimation has important advantages over random-effects estimation: (1) it is robust to the omission of any time-invariant regressors. If the evaluator has doubts about whether the assumptions of the random-effects model are satisfied, the fixed-effects estimator is better; and (2) it yields consistent savings estimates when the assumptions of the random-effects model hold. The converse is not true, making the fixed-effects approach more robust.

Because weaker assumptions are required for the fixed-effects model to yield unbiased estimates, this protocol generally recommends the fixed-effects estimation approach. The remainder of this protocol presents panel regression models that satisfy the fixed-effects assumptions.

number of subjects, because the model requires thousands of dummy variables that may overwhelm available computing resources.

- Transform the dependent variable and all independent variables (except for the fixed effects) by subtracting the subject-specific mean of each variable from the variable and then running OLS on the transformed data. This approach is equivalent to least squares dummy variables.[85]
- Estimate a first difference or annual difference of the model. Differencing removes the subject fixed effect and is equivalent to the dummy variable approach if the fixed-effects model is correctly specified.

2.7.4.4.2 Panel Regression Model Specifications

This section outlines common regression approaches for estimating treatment effects from residential BB programs. Unless otherwise stated, assume that the BB program was implemented as an RCT or RED field experiment.

2.7.4.4.3 Simple Differences Regression Model of Energy Use

Consider a BB program in which the evaluator has energy consumption data for the treatment period only and wishes to estimate the average energy savings per period from the treatment. Let t, t = 1, 2, …, T, denote the time periods during treatment for which data are available, and let i, i = 1, 2, …, N, denote the treatment and control group subjects in the analysis sample. For simplicity, assume that all treated subjects started the treatment at the same time.

A basic specification to estimate the average energy savings per treated customer per period is:

$$y_{it} = \beta_0 + \beta_1 \times Tr_i + \varepsilon_{it}$$

Where:

y_{it} = The metered energy consumption of subject i in period t

β_0 = The average energy consumption per unit of time for subjects in the control group

β_1= The average treatment effect of the program; the energy savings per subject per period equals -β1

Tr_i = An indicator for whether subject i received the treatment; the variable equals 1 for subjects in the treatment group and equals 0 for subjects in the control group

ε_{it} = The model error term, representing random influences on the energy consumption of customer i in period t.

In this simple model, the error term ε_{it} is uncorrelated with Tr_i because subjects were randomly assigned to the treatment or control group. The OLS estimation of this model will result in an unbiased estimate of β1. The standard errors should be clustered on the subject (customer). [86]

This specification does not include subject fixed effects. Because the available energy consumption data only apply to the treatment period, it is not possible to identify the program treatment effect and to incorporate

[85] Greene (2011) Chapter 11 provides more details.

[86] Although the methods recommended in this protocol minimize the potential for violations of the assumptions of the classical linear regression model, evaluators should be aware of and take steps to minimize—potential violations

subject fixed effects into the model. However, as previously noted, because of the random assignment of subjects to the treatment group, any time-invariant characteristics affecting energy use will be uncorrelated with the treatment, so omitting that type of fixed effects will not bias the savings estimates.

However, in the equation above in Section 2.7.4.2.3 *Equivalency Check*, more precise estimates of savings could be obtained by replacing the coefficient β0 with time-period fixed effects. The model would capture more of the variation in energy consumption over time, resulting in greater precision in the savings estimate. The interpretation of β1, the average treatment effect per home per time period, is unchanged.

2.7.4.4.4 Simple Differences Regression Estimate of Heterogeneous Savings Impacts

Suppose that the evaluator still has energy consumption data that apply to the treatment period only but wishes to obtain an estimate of savings from the treatment as a function of some exogenous variable, such as preprogram energy consumption, temperature, home floor space, or pretreatment efficiency program participation (to determine, for example, whether high energy users save more or less energy than low energy users). If data for treatment and control group subjects on the exogenous variable of interest are available, the evaluator may be able to estimate the treatment effect as a function of this variable.

Let m_{ij} be an indicator that subject i belongs to a group j, j = 1, 2, …, J, where membership in group j is exogenous to receiving the treatment. Then the average treatment effect per subject for subjects in group j can be estimated using the following regression equation:

$$y_{it} = \beta_0 + \sum_{j=1}^{J} \beta_{1j} \times Tr_i \times m_{ij} + \sum_{j=1}^{J-1} \gamma_{jmij}\, \mu_m + \varepsilon_{it}$$

Where:

m_{ij} = An indicator for membership of subject i in group j; it equals 1 if customer i belongs to group j and equals 0, otherwise

β_{1j} = The average treatment effect for subjects in group j; energy savings per subject per period j equals -β1j

γ_j = The average energy consumption per period for subjects in group j, j = 1, 2, …J-1.

All of the other variables are defined as in Section 2.7.4.2.3 *Equivalency Check.*

This specification includes a separate intercept for each group indicated by γ_j and the treatment indicator Tr_i interacted with each of the m_{ij} indicators. The coefficients on the interaction variables β_{1j} show average savings for group j relative to baseline average energy use for group j. It is important that the equation include the uninteracted indicator variables for the groups if average energy consumption varies between groups; otherwise, the treatment effect for group j will be incorrectly estimated relative to the average consumption of all control subjects rather than control subjects in group j.

2.7.4.4.5 Simple Differences Regression Estimate of Savings During Each Time Period

To estimate the average energy savings from the treatment during each period, the evaluator can interact the treatment indicator with indicator variables for the time periods as in the following equation.[87]

$$y_{it} = \sum_{j=1}^{J} \beta_{1j} \times Tr_i \times d_{ij} + \sum_{j=1}^{T} \theta_j d_{jt} + \varepsilon_{it}$$

Where:

> β_t = The average savings per subject for period j (for example, the average savings per subject during month 4 or during hour 6)
>
> d_{jt} = An indicator variable for period j, j = 1, 2, …,T. djt equals 1 if j = t (that is, the period is the tth) and equals 0 if j ≠ t (that is, the period is not the tth)
>
> θ_t = The average effect on consumption per subject specific to period j.

Equation 4 can be estimated by including a separate dummy variable and an interaction between the dummy variable and Tri for each time period t, where t = 1, 2, ..., T. When the time period is in months, the time-period variables are referred to as month-by-year fixed effects. The coefficient on the interaction variable for period t, βt, is the average savings per subject for period j. Again, because ε_{it} is uncorrelated with the treatment after accounting for the average energy consumption in period t, the OLS estimation of the equation in Section 2.7.4.4.4 *Simple Differences Regression Estimate of Heterogeneous Savings Impacts* (with standard errors clustered on subjects) results in an unbiased estimate of the average treatment effect for each period.

Evaluators with smart meter data can use this specification to estimate BB program demand savings during specific hours of the analysis period. The coefficient βj would indicate the demand savings from the treatment during hour j. Examples of research that estimates savings during hours of peak usage include Stewart (2013a), Todd (2014), and Brandon et al. (2019).

2.7.4.4.6 Difference-in-Differences Regression Model of Energy Use

This section outlines a DiD approach to estimating savings from BB field experiments. This protocol recommends DiD estimation to the simple differences approach but DiD requires information about the energy use of treatment and control group subjects during the pretreatment and treatment periods. These energy use data enable the evaluator to:

- Include subject fixed effects to account for differences between subjects in time-invariant energy use
- Obtain more precise savings estimates
- Test identifying assumptions of the model

Assume there are N subjects and T +1 periods, T > 0, in the pretreatment period denoted by t = - T, -T+1, …, -1, 0, and T periods in the treatment period, denoted by t = 1, 2, …, T. A basic DiD panel regression with subject fixed effects could be specified as:

[87] If the number of time periods is very large, the number of time period indicator variables in the regression may overwhelm the capabilities of the available statistical software. Another option for estimation is to transform the dependent variable and all of the independent variables by subtracting time-period-specific means and then running the OLS on the transformed data.

$$y_{it} = \alpha_i \times \beta_1 P_t + \beta_2 P_t \times Tr_i + \varepsilon_{it}$$

Where:

α_i = Unobservable, time-invariant energy use for subject i; these effects are controlled for with subject fixed effects

β_1 = The average energy savings per subject during the treatment period that was not caused by the treatment

P_t= An indicator variable for whether time period t occurs during the treatment; it equals 1 if treatment group subjects received the treatment during period t, and equals 0 otherwise

β_2= The average energy savings resulting from the treatment per subject per unit of time.

The model includes fixed effects to account for differences in average energy consumption between subjects. Including subject fixed effects would likely explain a significant amount of the variation in energy consumption between subjects and result in more precise savings estimates. The interaction of Pt and Tri equals one for subjects in the treatment group during periods when the treatment is in effect, and 0 for other periods and all control subjects.

The equation in Section 2.7.4.4.5 *Simple Differences Regression Estimate of Savings During Each Time Period* is a DiD specification. For control group subject i, the expected energy use is αi during the pretreatment period and αi + β1 during the treatment period. The difference in expected energy use between pretreatment and treatment periods, also known as naturally occurring savings, is β1. If that same subject i had been in the treatment group, the expected energy use would have been αi during the pretreatment period and αi + β1 + β2 during the treatment period. The expected savings would have been β1 + β2, which is the sum of naturally occurring savings and savings from the BB program. Taking the difference yields β2, a DiD estimate of program savings. The OLS estimation in Section 2.7.4.4.5 *Simple Differences Regression Estimate of Savings During Each Time Period* results in an unbiased estimate of β2.

A more general form of Section 2.7.4.4.5 *Simple Differences Regression Estimate of Savings During Each Time Period* would allow the treatment period to vary for each subject and substitute time-period fixed effects (such as a separate indicator variable for each day or month of the analysis period) for the stand-alone post period variable. The specification with time-period fixed effects in Eq. 6 can be handy when subjects begin the treatment at different times, such as with rolling program enrollments or if it is difficult to define when treatment would have begun for a control group subject.

$$y_{it} = \alpha_i \times \tau_t + \beta_2 P_{it} \times Tr_i + \varepsilon_{it}$$

Where:

τ_t = The time-period fixed effect (an unobservable that affects the consumption of all subjects during time period t); the time period effect can be estimated by including a separate dummy variable for each of T-1 time periods t, where t = -T, -T+1, …, -1, 0, 1, 2, ..., T; one time period dummy variable must be dropped to avoid collinearity

P_{it} = An indicator variable for whether time period t occurs during the treatment for subject i; it equals 1 if treatment group subject i received the treatment during period t, and equals 0 otherwise.

As in Section 2.7.4.4.4 *Simple Differences Regression Estimate of Heterogeneous Savings Impacts*, the coefficient β2 represents the average savings per treated customer per time period. The interpretations of the other variables and coefficients in the model remain unchanged.

2.7.4.4.7 Difference-in-Differences Estimate of Savings for Each Time Period

By re-specifying the equation in Section 2.7.4.4.5*Simple Differences Regression Estimate of Savings During Each Time Period* with time-period fixed effects, savings can be estimated during each period and the identifying assumption tested to determine that assignment to the treatment was random. Consider the following DiD regression specification:

$$y_{it} = \alpha_i + \sum_{j=-T}^{T} \theta_j d_{jt} + \sum_{j=-T}^{-1} \beta_j Tr_i \times d_{ij} + \sum_{j=1}^{J} \beta_j Tr_i \times d_{ij} + \varepsilon_{it}$$

Savings in each period are estimated by including a separate dummy variable and an interaction between the dummy variable and Tri for each time period t, where t = -T, -T+1, ..., - 1, 0, 1, 2, ..., T. The coefficient on the interaction variable for period t, βt T, is the DiD savings for period t.

Unlike the simple differences regression model, this model yields an estimate of BB program savings during all periods except one, which must be excluded to avoid collinearity, for a total of 2T-1 period savings estimates. Figure 2-12 shows an example of savings estimates obtained from such a model. The dotted lines show the 95% confidence interval for the savings estimates using standard errors clustered on utility customers.

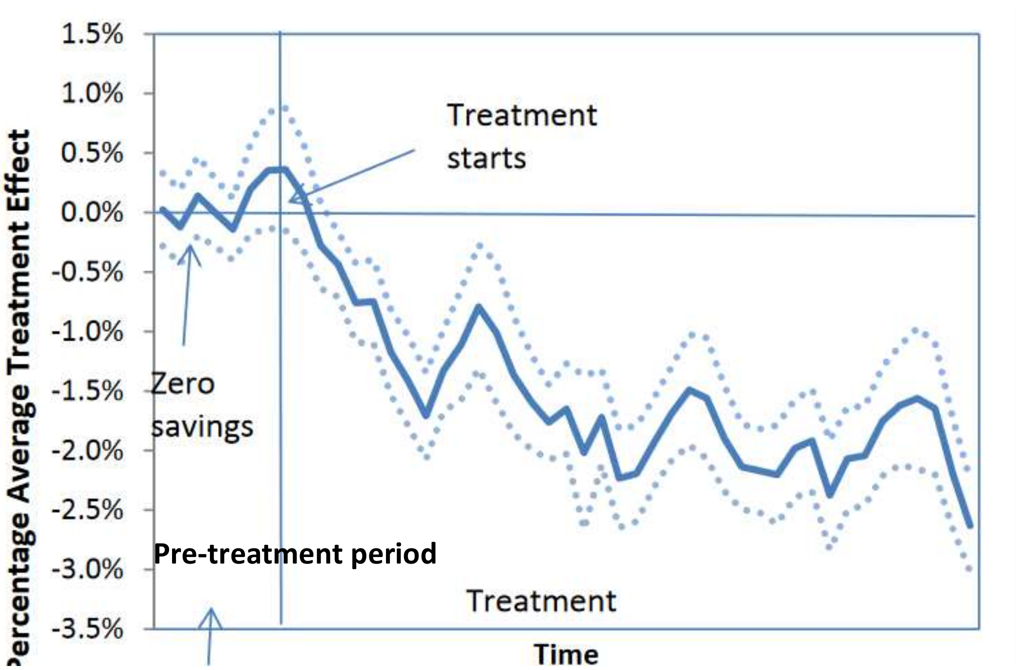

Figure 2-12 Example of DiD Regression Savings Estimates[88]

Estimates of pretreatment savings can be used to test the assumption of random assignment to the treatment. Before utilities administer the treatment, it should not be possible to reject the hypothesis of statistically significant differences in energy consumption between treatment and control group subjects, that is, the confidence intervals should contain the x axis. BB program pretreatment saving estimates that were statistically different from zero might suggest a flaw in the experiment design or implementation or the evaluator's understanding of the experiment.

As with the equation in Section 2.7.4.4.3 *Simple Differences Regression Model of Energy Use* this specification can be used to estimate demand savings during specific hours. Energy consumption data for hours before the treatment are required, however.

2.7.4.4.8 Simple Differences Regression Model with Pretreatment Energy Consumption

In addition to estimating energy savings as a DiD, evaluators can estimate savings as a simple difference conditional on average pretreatment energy consumption. This estimator, often referred to as a post-only model with pre-period controls or lagged dependent variable, includes pretreatment energy consumption as an independent variable in the regression to account for differences between subjects in their post-treatment consumption, serving a purpose similar to that of customer fixed effects in the DiD model. However, many researchers favor the post-only estimator because it usually has smaller variance than the standard fixed effects DiD estimator when energy consumption is uncorrelated or weakly correlated over time.[89] However, evaluators can estimate both specifications and compare results. In large samples, the models should produce very similar estimates.

Consider the following post-only with pre-period controls regression specification:

$$y_{it} = \tau_t + \beta_1 \times Tr_{it} + \rho \overline{y_{\varphi}^{pre}} + \varepsilon_{it}$$

Where:

> τ_t= The time-period fixed effect (an unobservable that affects consumption of all subjects during time period t); the time period effect can be estimated by including a separate dummy variable for each time period t, where t = -T, - T+1, …, -1, 0, 1, 2, ..., T
>
> β_1 = Coefficient for the average treatment effect of the program; the energy savings per subject per period equals -β_1
>
> Tr_{it}= An indicator variable for whether subject i received the treatment in period t; the variable equals 1 for subjects who receive the treatment in period t and equals 0 otherwise

[88] *Ibid.*

[89] Some researchers refer to this model as a "post-only" model; however, this name is misleading because the model uses pretreatment consumption as an explanatory variable. In a personal correspondence with the authors, Hunt Allcott, who introduced this method in evaluation of Home Energy Reports, points out that if seasonal effects are being estimated, this model "has slightly smaller standard errors and can be better at addressing naturally occurring randomization imbalances that may result in the baseline pretreatment energy usage differing between the control and treatment group."

ρ = Coefficient indicating the effect of average pretreatment consumption on consumption during the treatment period

$\rho\overline{y_{\varphi}^{pre}}$ = Average consumption during the corresponding pretreatment period for subject i; for example, if the dependent variable was a customer's average daily consumption in July during the treatment period, $\rho\overline{y_{\varphi}^{pre}}$ would equal the customer's average daily consumption for July in the pretreatment period

ε_{it} = The model error term, representing random influences on the energy consumption of customer i in period t.

With random assignment of subjects to treatment and control groups, the OLS estimation of 2.7.4.4.8 *Simple Differences Regression Model with Pretreatment Energy Consumption* is expected to produce an unbiased estimate of the average savings per subject per period.

Evaluators can estimate slightly different versions of this model:

- Savings for each treatment period. Evaluators can include a treatment indicator variable for each period instead of a treatment indicator variable for the entire treatment period. This specification will produce an estimate of average savings per subject for each treatment period.
- Additional pretreatment consumption control variables. Instead of one pretreatment consumption variable, evaluators can include multiple pretreatment consumption variables, such as pretreatment consumption for different seasons or months of a year, days of the week, or hours of the day.
- Additional control variables. Evaluators can add other variables such as weather to the model. The addition of such variables might help to improve the precision of the savings estimates.

2.7.4.4.9 Randomized Encouragement Design

Some field experiments involve a RED in which subjects are only encouraged to accept a BB measure, in contrast to RCTs in which a program administers a BB intervention. This section outlines the types of regression models that are appropriate for estimating savings from REDs, how to interpret the coefficients, and how to estimate savings from RED programs.

Evaluators can apply the model specifications previously described for RCTs to REDs. The model coefficients and savings are interpreted differently; however, an additional step is required to estimate average savings for utility customers who accept the behavioral intervention. Treatment in a RED is defined as receiving encouragement to adopt the BB intervention, rather than actually receiving the intervention, as with RCTs.

Consider a field experiment with a RED that has energy consumption data for treatment and control group subjects available for the pretreatment and treatment periods. Equations 1 through 4 can be used to estimate the treatment effect, or the average energy consumption effect on those receiving encouragement. If control group customers can participate, the estimate only captures savings from compliers, because, as discussed previously, never takers never accept the intervention, and always takers accept the intervention with or without encouragement.

To recover an estimate of the LATE—the savings from subjects who accept the treatment because of the encouragement—evaluators can scale the estimate of β2 by the inverse of the difference between the percentage of subjects in the treatment group who accept the intervention and the percentage of subjects in the control group who accept the intervention (which is zero if control group subjects are prohibited from accepting the intervention). Estimate this as:

$$LATE = \frac{\beta_2}{(\pi_T - \pi_C)}$$

Where:

π_T= The percentage of treatment group subjects who accept the intervention

π_C= The percentage of control group subjects who accept the intervention.

A related approach for obtaining an estimate of savings for the BB intervention in a RED study is instrumental variables, two-stage least squares (IV-2SLS). This approach uses the random assignment of subjects to the treatment as an instrumental variable for the decision by encouraged customers to participate in the program. The instrumental variable provides the exogenous variation necessary to identify the effect of endogenous participation on energy consumption. Participation is endogenous because the encouraged customers' decisions to participate is not random and depends on unobserved characteristics that may be correlated with energy consumption. For encouragement to be a valid instrument, it must be that encouragement affects only energy consumption through its impact on BB program participation.

In the first stage, the evaluator regresses a binary program participation decision variable on an indicator for whether the customer was randomly assigned to receive encouragement and other exogenous independent variables from the second-stage energy consumption equation. The evaluator then uses the regression to predict the likelihood of participation for each subject and time period. In the second stage, the evaluator estimates the energy consumption equation, substituting the first-stage predicted likelihood of participation for the variable indicating actual program participation. The estimated coefficient on the predicted likelihood of participation is the LATE for the BB intervention.

For a detailed method of using an IV approach, see Cappers et al. (2013) and for a real-world example of the IV-2SLS approach applied to a home weatherization program implemented as a RED, see Fowlie et al. (2018).

2.7.4.4.10 Quasi-Experimental Methods

(i) Regression Discontinuity Method

Among the quasi-experimental methods, regression discontinuity typically yields the most unbiased estimate of energy savings. However, it is also the most complicated method: it requires knowledge of econometric models and often requires field conditions that allow the evaluator to utilize this analytic technique and is therefore not always practical. This method works if the eligibility requirement for households to participate in a program is a cutoff value of a characteristic that varies within the population. For example, households at or above a cutoff energy consumption value of 900 kWh per month might be eligible to participate in a behavior-based efficiency program, while those below 900 kWh are ineligible. In this case, the households that are just below 900 kWh per month are probably very similar to those that are just above 900 kWh per month. Thus, the idea is to use a group of households right below the usage cutoff level as the control group and compare changes in their energy use to households in right above the usage cutoff level as the treatment group. This method assumes that the program impact is constant over all ranges of the eligibility requirement variable that are used in the estimation (e.g., that the impact is the same for households at all levels of energy usage), although there are more complex methods that can be used if this assumption is not true.[90] In addition, regression discontinuity relies on the eligibility requirement being strictly enforced.[91]

(ii) Matched Control Group Method

If it is not possible to create a randomized control group, then savings estimates could be calculated by constructing a non-random control group made up of households that are as similar to the treatment group as possible. The challenge with a matched control group method is that households have both observable characteristics (e.g., level of energy use, zip code, presence of central air conditioning) that could potentially be matched, and unobservable characteristics (e.g., energy attitudes, or propensity to opt in to an energy efficiency program) that are harder or impossible to match.

(iii) Match on Observables

A matched control group or post-matched control group is a non-random control group where the observable characteristics of the households in the program are known or measured, and then a control group that best matches those characteristics is constructed. The idea is to create a control group that is as similar as possible to the treatment group. For example, with an opt-in program, it may be true that all households that opted in lived in a rural area and had high energy use. In this case, a matched control group might include households in the same rural area with high energy use that did not opt in to the program. This control group is matched on two observable characteristics (energy use and location). However, it is not matched on the unobserved variable of propensity to opt in: it ignores the fact that households that opt in to a program are fundamentally different than those that do not opt in. For example, these households may be more inclined to conserve energy than those that are not interested in participating in the program. In the case of an opt-out program, the households that could be used in the matched control group are either those that were screened out or those that opted out.

(iv) Propensity Score Matching

Propensity score matching attempts to match households on both observable and unobservable characteristics for the case of an opt-in program. This method uses observable characteristics to predict the probability that a household will decide to opt in to a program, and then chooses households that had a high probability of opting in to the program but did not actually opt in to be in the control group. While this method is better than a matching method without propensity scores, it still assumes that whatever observable characteristics of the households were used to calculate the propensity score are sufficient to explain any unobservable differences between the treatment and non-random control group. This method is more credible if accurate detailed household demographic information is obtainable, rather than generic categories (e.g., broad census demographics or categories such as “rural youth”). However, in cases for which RCTs and regression discontinuity methods are impractical, propensity score matching is an acceptable method.

(v) Variation in Adoption (With a Test of Assumptions)

This variation in adoption approach takes advantage of variation in the timing of program adoption. This allows for the comparison of the energy usage of households that opt in to the energy usage of households that have not yet opted in but will ultimately opt in at a later point. It relies on the assumption that in any given month,

[90] See Imbens and Lemieux (2008).

[91] In addition, the eligibility requirements cannot be endogenously determined; that is, if there is prior knowledge that households above 900 kWh have a strong response to the program while those below 900 kWh do not, then regression discontinuity will yield biased estimates.

households that have already opted in and households that will opt in soon are the same types of households. For this assumption to be valid, households must decide to opt in to the program at different times, and the decision of each household to opt in during any particular month should be essentially random, and only influenced by marketing exposure and awareness of the program (this is different than an RCT with a recruit-and-delay design, in which households do not decide when to opt in but rather are randomly assigned different times to opt in). The decision to opt in should not be related to observable or unobservable household characteristics (e.g., energy conservation attitudes). Because the validity of the estimated program impact depends upon this assumption, it should be tested to the extent possible with a test of assumptions.[92]

In addition, if the energy savings due to the program do not persist over time, then the estimated program impact will be biased and thus require corrections.[93] If the assumption that the timing of household program adoption is essentially random is valid, then this method is as good as a regression discontinuity method. However, although the assumption can be tested and found to not hold, it cannot be found to hold with certainty (e.g., household adoption may correspond to unobservable characteristics, such as willingness to opt in during a specific season).

(vi) Pre-Post Energy Use Method

Another quasi-experimental method is to compare the energy use of households in the treatment group after they were enrolled in the program to the same households' historical energy use prior to program enrollment. In effect, this means that each household in the treatment group is its own non-random control group. This is called a pre-post; within subjects; or interrupted time series design analysis. The challenge in using this method is that there are many other factors (independent variables) that may influence energy use before, during, and after the program that are not captured with this method. Some of these factors, such as differences in weather or number of occupants, may be reliably accounted for in the analysis. However, other factors are less easily observed and/or accounted for. For example, the economy could have worsened, leading households to decrease energy (even if there were no program), or a pop culture icon could suddenly decide to advocate for energy efficiency. With a pre-post analysis, there is no way to discern and separate the impact of other influences (e.g., economic recession) that may affect energy use over time compared to the impact of the behavior-based efficiency program leading to an estimate of energy savings that could be biased.[94]

2.7.4.4.11 Standard Errors

Panel data have multiple energy consumption observations for each subject; thus, the energy consumption data are very likely to exhibit within-subject correlations. Many factors affecting energy consumption persist over time, and the strength of within-subject correlations usually increases with the frequency of the data. When standard errors for panel regression model coefficients are calculated, the within-subject correlations must be accounted for. Failing to do so will lead to savings estimates with standard errors that are biased.

---

[92] One way to test it is by conducting a duration analysis, which tests whether household adoption in any particular month is driven by marketing activity, as opposed to observed household characteristics or unobserved heterogeneity. Another test is to determine if the energy usage of households before they opt in differs between households that opt in during one particular month as opposed to another month. In addition, propensity score matching can be used to further verify the assumption by accounting for potentially varying demographics of the households over time as they opt in to the program.

[93] For a detailed description of a robust variation in adoption methodology, see Harding and Hsiaw (2011)

[94] However, in programs outside the scope of this report such as critical peak pricing or critical peak rebates, a pre-post method may be less biased. This method should only be considered when the experimental factor can be presented repeatedly so that the difference between the behavior when it is present and when it is not present is observable. It is not really appropriate for circumstances where the effect of the experimental factor is expected to persist for a long period of time after exposure or is continuously presented throughout the experiment (e.g., time of use or information feedback).

This protocol strongly recommends that evaluators estimate robust standard errors clustered on subjects (the randomized unit in field trials) to account for within-subject correlation. Most statistical software programs, including STATA, SAS, and R, have regression packages that output clustered standard errors.

Clustered standard errors account for the fact that in a panel with N subjects and T observations per subject there is less information about energy consumption than in a data set with N*T independent observations. Because clustered standard errors account for these within-subject energy-use correlations, they are typically larger than OLS standard errors. When there is within subject correlation, OLS standard errors are biased downward and overstate the statistical significance of the estimated regression coefficients.[95]

2.7.4.4.12 Opt-Out Subjects and Account Closures

Many BB programs allow subjects to opt out and stop receiving the treatment. This section addresses how evaluators should treat opt-out customers in the analysis, as well as utility customers whose billing accounts close during the analysis period.

As a general rule, evaluators should include all subjects initially assigned to the treatment and control groups in the savings analysis.[96] Specifically, evaluators should keep opt-out subjects in the analysis sample. Opt-out subjects may have different energy consumption characteristics than subjects who remain in the program and dropping them from the analysis would result in nonequivalent treatment and control groups. To ensure the internal validity of the savings estimates, opt-out subjects should be kept in the analysis sample.

Sometimes treatment or control group subjects close their billing accounts after the program starts. Account closures are usually unrelated to the BB program or savings; most are a result of households changing residences. Subjects in the treatment group should experience account closures for the same reasons and at the same rates as subjects in the control group; evaluators can thus safely drop treatment and control group subjects whose accounts close from the analysis sample.

When dropping customers who close their accounts during the treatment from the regression estimation, evaluators should still count the savings from these subjects for periods during treatment when their accounts were active. To illustrate, when estimating savings for a 1-year BB program, evaluators can estimate the savings from subjects who closed their accounts and from those who did not as the weighted sum of the conditional average program treatment effects in each month:

$$Savings = \sum_{m=1^{12}} - \beta_m \times Days_m + N_m$$

Where:

m = Indexes the months of the year

$-\beta_m$ = The conditional average daily savings in month m (obtained from a regression equation that estimates the program treatment effect on energy consumption in each month)

---

[95] Bertrand et al. (2004) show when DiD studies ignore serially correlated errors, the probability of finding significant effects when there are none (Type I error) increases significantly.

[96] This protocol urges evaluators not to arbitrarily drop outlier energy consumption observations from the analysis unless energy consumption was measured incorrectly, the customer was not a residential customer, or the sample size is small enough that the outlier strongly influences the estimated savings. If an outlier is dropped from the analysis, the reasons for dropping the outlier and the effects of dropping it from the analysis on the savings estimates should be clearly documented. Evaluators should test the sensitivity of the results to dropping observations.

$Days_m$= The number of days in month m

N_m = The number of treatment group subjects with active accounts in month m.

2.7.4.5 Program Uplift and Double Counting of Savings

Many BB programs cause participants to increase their participation in other utility energy efficiency programs, a phenomenon often referred to as efficiency program uplift. For example, most home energy report programs encourage recipients to participate in other utility energy efficiency programs that provide cash rebates in exchange for adopting efficiency measures, such as efficient furnaces, air conditioners, wall insulation, windows, and light-emitting diodes (LEDs). The savings from this efficiency program participation caused by HERs are often referred to as joint savings or uplift savings. Quantifying the effects of BB programs on efficiency program participation is important for two reasons:

- Uplift can be an important effect of BB programs and a potential additional source of energy savings.
- Savings from efficiency program uplift may be double counted. When a utility customer participates in an efficiency program because of a BB program intervention, the utility may count the program savings twice: once in estimating BB program savings and again in estimating the rebate program savings. To avoid double counting, evaluators must estimate savings from program uplift and subtract these savings from the behavior program savings or the uplifted program savings or from both programs.[97]

2.7.4.5.1 Estimating Uplift Energy Savings

For BB programs implemented as randomized experiments, estimating savings from uplift is conceptually straightforward. To illustrate, suppose that a utility markets an energy efficiency measure to treatment and control group subjects identically through a separate rebate program. Customers in the behavioral treatment group also receive messaging encouraging them to adopt the measure. Because customers were randomly assigned to the treatment and control groups, the groups are expected to be equivalent except for the treated customers having received the BB program encouragement. Therefore, in comparing BB program treatment and control group customers, evaluators can attribute any difference in the uptake of the measure between the groups to the behavioral treatment. To improve the accuracy of the uplift estimate, evaluators can estimate the impact as a DiD, by comparing the change in uptake of the measure between the pre and post-treatment periods for treatment and control group customers. The DiD estimate will account for any preexisting differences between treatment and control groups in the tendency to adopt the measure. If data are not available on the installation of the measure in the pretreatment period (for example, if it was not rebated at that time), evaluators should estimate uplift savings based only on post-treatment differences.

Figure 2-13 illustrates the logic for calculating behavior program savings from the efficiency program as a DiD. The figure shows energy savings from utility rebate program participation for treatment and control group customers during the pretreatment and treatment periods. Although customers had been randomly assigned to

[97] This protocol does not take a position on which program gets credit for the uplift. When a BB intervention causes participation in an energy efficiency program, we know that the program participation would not have occurred without the intervention. However, the amount of uplift caused by the BB intervention may depend on the dollar incentives provided by the efficiency program. For example, the BB program may produce greater lift in participation for a program incentive of $200 than $100. To determine the relationship between uplift and the incentive amount, it would be necessary to randomize the incentive amount and to study participation as a function of incentives and who receives the BB intervention. It is possible to subtract the uplift savings from either the behavior program or the uplifted program. However, it is common practice for evaluators to attribute all joint or uplift savings to other energy efficiency programs by subtracting them from the BB program savings. This is a simple and convenient approach for avoiding double counting of savings.

receive treatment, treatment group customers had a slightly higher tendency to participate and greater savings (=5) during the pretreatment period than the control group (=4). In this case, estimating program uplift by taking the simple difference in post-treatment savings between the treatment and control groups (8-4) would ignore the higher savings for the treatment group that would have occurred in the absence of the BB treatment and yield a slightly biased uplift savings estimate of 4. The true uplift savings equal 3, and an accurate estimate can be obtained as a DiD: (8-5) – (4-4).

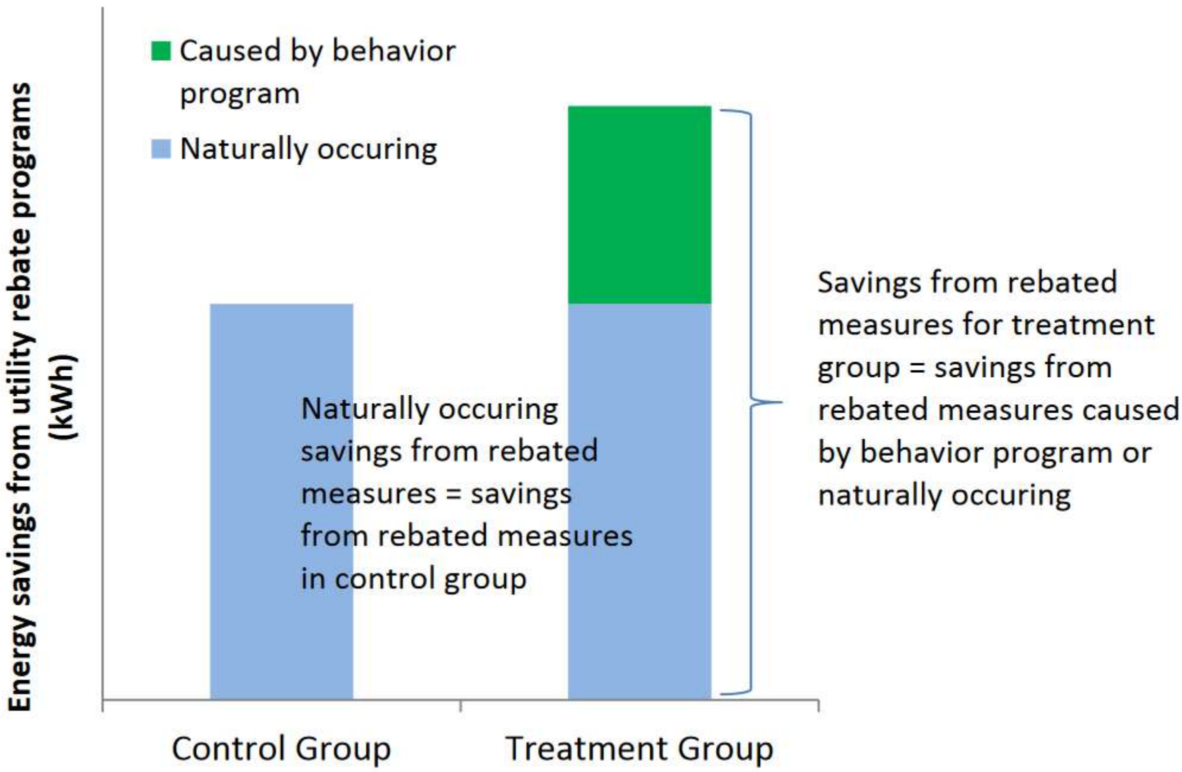

Figure 2-13 Calculation of Double-Counted Savings[98]

To estimate BB program savings from efficiency program uplift, evaluators should take the following steps:

- Collect energy efficiency program tracking data for treatment and control group customers for the year before treatment and all years of treatment.[99] Match the BB program treatment and control group subjects to the utility energy efficiency program tracking data.
- Calculate the average uplift savings per treatment group customer as the DiD between treatment and control groups in average efficiency program savings per customer, where the savings are obtained from the utility tracking database of installed measures.[100] The averages should be calculated over all treatment group customers and all control group customers, not just those who participated in efficiency programs.

[98] *Ibid.*

[99] These data should include a customer account number and premise number for linking the records to individual customers and homes, a measure description and category, the installation date, the quantity installed, and a unit annual savings value

[100] A simple difference can be used if evaluators verify that pretreatment energy efficiency program participation and savings are equal for treatment and control group customers or if pretreatment energy efficiency program data are not available. Pretreatment data will be unavailable for new programs.

Evaluators can calculate the average uplift savings per treatment group customer as a difference in unconditional means between treatment group and control group customers or in a regression. As described in the next few paragraphs, it may be necessary to adjust the deemed savings values in the utility tracking data for measures installed for less than 1 year.

- Multiply the uplift savings per treatment group customer by the number of customers who were in the treatment group to obtain the total uplift savings.

Evaluators can estimate BB program uplift savings for efficiency measures that the utility tracks at the customer level. Most measures for which utilities offer rebates—such as high-efficiency furnaces, windows, insulation, and air conditioners—fit this description. Also, evaluators can perform the uplift analysis for individual efficiency measures or programs or in aggregate across all programs and measures. Performing the analysis for individual measures or programs may provide useful insights about interactions between the BB program and other efficiency programs that an aggregate analysis cannot provide.

Evaluators should be mindful of specific reporting conventions for efficiency program measures in utility tracking databases. For example, many jurisdictions require utilities to report weather-normalized and annualized measure savings, which do not reflect when measures were installed during the year or the actual weather conditions that affect savings. In contrast, regression-based estimates of energy savings, such as from Eq. 4, will reflect installation dates of measures and actual weather. Evaluators should therefore adjust the annual deemed savings in the program reporting database to account for when measures were installed during the year and weather.

In addition, for BB programs treating customers for longer than 1 year, evaluators should account for the savings from uplift in previous years if uplift savings are subtracted from the behavior program. Measures with a multiyear life installed in previous program years will continue to save energy for the remaining life of the measure. Depending on the utility's conventions for reporting savings, it may be necessary to account for savings from program lift in previous program years from the BB program savings estimate.[101]

2.7.4.5.2 Estimating Uplift for Upstream Programs

Upstream measures are those that the utility does not track at the customer level. The most important of such measures are high-efficiency lights such as LEDs that are rebated through utility upstream programs. Most utilities provide incentives directly to retailers for purchasing these measures, and the retailers then pass on these price discounts to utility customers at the point of sale. Estimating behavior program savings for upstream measures is conceptually similar to that for downstream measures but requires a different data collection approach. Data on the purchases of rebated measures by treatment and control group subjects can be collected through customer surveys, store intercept surveys, or home site visits.[102]

Evaluators wanting to estimate the lift in LED adoption from upstream programs should be aware that it may be necessary to collect data for large numbers of customers to detect small BB program treatment effects. If evaluators perform surveys, they should size their survey samples with the objective of being able to detect small but economically significant effects. However, if the treatment effect is small, the uplift savings from LEDs will also be small, and it may not be worth conducting surveys to measure it. Also, evaluators should adjust the

[101] For an example of a HER program evaluation that makes these adjustments, see Cadmus (2018) and DNV-GL (2018).

[102] See PG&E (2013) for an example of a study employing home visits.

lighting purchases impact estimates for in-service rates and the percentage of high-efficiency lamps sold in the utility service area that received rebates.[103] Evaluators should also be aware that some energy savings from purchasing LEDs may be offset by reductions in the hours of use of those bulbs by treated customers. LEDs may save less because treated customers light their homes less than before.

2.7.4.6 Savings Persistence and Measure Life

Most behavior-based program administrators and utility regulators assume a 1-year measure life for HERs and other residential BB measures. Administrators and regulators have been conservative in their assumptions about measure life for several reasons. First, doubts exist about the persistence of behavioral savings after treatment ends for utility customers who had been changing thermostat settings, turning lights off in unoccupied rooms, or modifying other energy consumption behaviors. Second, until recently, there has been a lack of evidence demonstrating that BB savings persist. Finally, HERs and other BB measures are fundamentally different than home improvements, such as LEDs or air source heat pumps. This difference is because BB measures attempt to influence behaviors, which often requires repeated treatments to be effective. Further, their effects can decay. For all these considerations, the default assumption for most BB program administrators has been that behavioral savings do not persist and that measure life is 1 year.

However, in the last 7 years, researchers have conducted highly credible RCT studies demonstrating that HER customers continue to save energy after treatment ends and that savings may persist for several years. In addition, researchers have developed frameworks for estimating BB savings persistence and implementing a multiyear measure life (Khawaja and Stewart 2014; Jenkins et al. 2017). These frameworks account for repeated, multiyear program treatments and the gradual decay of behavior-based measure savings.

Because of this research, some BB program administrators and regulators have begun to reconsider the assumption of a 1-year measure life and allowed for savings persistence. For example, the Illinois (IL) TRM was recently updated to require adjustments to HERs savings for persistence. Other states previously or recently adopted a multiyear measure life for HERs or other BB measures or are proposing to adopt one.[104]

This part of the protocol provides evaluators with guidance about HER savings accounting and designing experiments to estimate BB savings persistence and measure life and about estimating BB savings when savings from previous treatments persist. The protocol does not recommend specific savings decay or measure life values.

2.7.4.6.1 BB Savings Persistence and Measure Life Concepts

BB savings persistence and measure life concepts presented in this section are meant to be illustrative of how program administrators can perform BB savings accounting with a multiyear measure life. BB savings accounting methods are still evolving, and there is not yet consensus about the details. Sections (i) *IL TRM* and (ii) *PA TRM* describe approaches that two states have implemented for performing HER savings accounting.

---

[103] Upstream lighting savings captured in the BB program savings calculation equals the product of the BB treatment effect on upstream lighting purchases (in bulbs, estimated from the comparison of treatment group and control group purchases), the in-service rate, and the unit savings. The portion of these upstream lighting savings claimed by the upstream lighting program equals the product of the upstream lighting savings, the ratio of upstream sales to total market sales, and the upstream program net-to-gross ratio.

[104] As of July 2019, Illinois, Connecticut, New Hampshire, and Minnesota have adopted a multiyear measure life for home energy reports. Pennsylvania is considering an HER multiyear measure life.

Figure 2-14 illustrates the measure life, savings persistence, and savings decay concepts for a multiyear BB program. The figure shows the average annual savings per customer for the first five program years. Suppose in the first year that the BB treatment generates 100 kWh of savings. Assume that savings from this treatment and all subsequent treatments partially persist, decaying at a 20% annual rate. In the second year, the BB treatment generates 150 kWh, but not all these savings are attributable to the second year 2 treatment. Eighty kWh of savings are from the year 1 treatment, and the remaining 70 kWh of savings are new savings attributable to the year 2 treatment. In year 3 (and years 4 and 5), the same logic applies. Only a portion of the annual savings are attributable to that year's treatment. In year 3, 64 kWh of savings are from year 1 treatment, 56 kWh of savings are from the year 2 treatment, and 68 kWh are new savings, attributable to treatment in year 3.

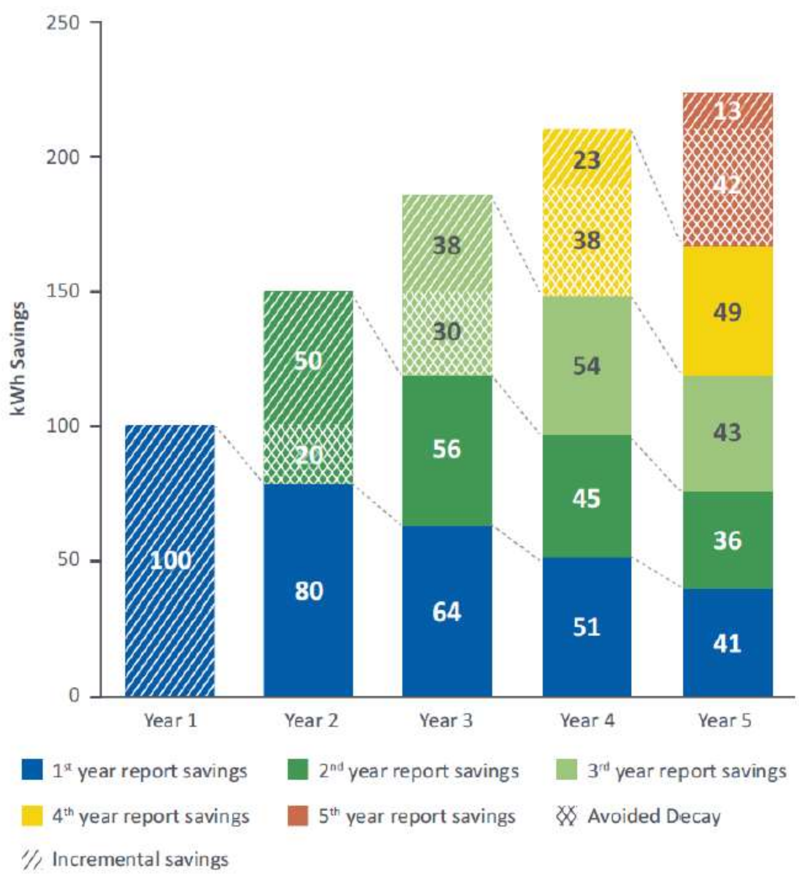

Figure 2-14 Illustration of Savings Persistence[105]

Figure 2-14 shows: (1) for each year of BB treatment new savings are generated and that a fraction of the new savings persist in future years, and (2) for BB programs in which the same customers receive treatments in multiple years, some annual savings may be attributable to treatments provided in previous years. This implies that after year 1 only a portion of annual savings will be attributable to treatment in that year.

In this example, with a constant annual savings decay rate, the lifetime savings from the year t treatment is the sum of year t new savings (sn,t) and savings in future years from persistence of year t savings:

[105] *Ibid.*

$$Lifetime\ Savings = s_{n,t} + s_{n,t}(1-\delta) + s_{n,t}(1-\delta)^2 + \cdots = \frac{S_{n,t}}{\delta}$$

Where δ, 0≤ δ < 1, is the savings decay rate. For instance, if δ=0.2, lifetime savings would equal $5s_{n.t}$.

Also, with measurements of the annual savings in year t, new savings from previous years of treatment, and the savings decay rate, it is possible to deduce new savings in the program's t[th] year.

$$New\ savings\ in\ year\ t = s_t - \sum_{k=1}^{t}(1-\delta)^k s_{n,t-k}$$

Where s_t is an estimate of the annual savings.

For example, the new savings in program year 3 equals:

$$s_{n,3} = s_3 - (1-\delta)^2(1-\delta)^2 s_{n,1} - (1-\delta)(1-\delta)\ s_{n,2}$$

In the example, to estimate the year 3 new savings, the evaluator would estimate the year 3 annual savings (188 kWh) by regression analysis and then subtract the decay-adjusted year 1 and year 2 new savings, which equal 64 kWh and 56 kWh, respectively. The year 3 new savings equal 68 kWh.

The previously mentioned formulas for lifetime and new savings neglect that most behavioral energy-efficiency programs experience attrition in the number of program participants because of customers moving residences and closing their accounts. A more accurate estimate of these savings would account for this attrition. If the annual rate of customer attrition equals α, 0≤ α < 1:

$$Lifetime\ Savings = s_{n,t} + s_{n,t}(1-\delta)(1-\delta) + s_{n,t}(1-\delta)^2(1-\delta)^2 + \cdots = \frac{S_{n,t}}{\delta + \alpha - \delta\alpha}$$

$$s_{n,t} = s_{n,t} - \sum_{k=1}^{t}(1-\delta)^k(1-\delta)^k s_{n,t-k}$$

For example, with customer attrition, new savings in program year 3 would equal:

$$s_{n,3} = s_3 - (1-\delta)^2(1-\delta)^2 s_{n,1} - (1-\delta)(1-\delta)\ s_{n,2}$$

Evaluators may also need an estimate of BB measure life. The measure life of a behavior-based treatment (e.g., reports sent in the second year of a program) can be defined as the lifetime savings expressed in terms of first-year savings equivalents.[106]

$$Measure\ life_{n,t} = \frac{lifetime\ savings_{n,t}}{s_{n,t}} = \frac{1}{\delta + \alpha - \delta\alpha}$$

[106] TPA typically define measure life using the concept of effective useful life: "the median length of time (in years) that an energy efficiency measure is functional." (Hoffman et al. 2015) Because it is not possible to directly observe functionality of BB measures in contrast to a efficiency product, it is necessary to estimate BB measure life in terms of first-year savings.

For example, for δ = 0.2 and α = 0.1, measure life for a behavior-based treatment in year t would equal 3.57 years. The lifetime savings from the year t treatment equal approximately 3.5 times the new savings in year t.

This illustration of the savings persistence and measure life concepts has assumed that savings decay indefinitely at a constant annual rate and the customer attrition rate is constant, but these assumptions, while simplifying the savings accounting, may not hold and need not be used. For example, the savings persistence rate may change over time, savings may persist for a finite number of periods, or customer attrition rates may vary. Evaluators can relax the assumptions and adapt the previously mentioned framework to their own situations. However, even with alternative assumptions, the concepts of new savings, lifetime savings, and measure life described earlier are still valid, and with modifications, the formulas for these concepts can be applied.

The following section describes how Illinois and Pennsylvania have conducted HER savings accounting with a multiyear measure life.[107]

(i) IL TRM

The IL TRM incorporates the HER savings accounting framework with several modifications.[108] The TRM assumes that HER electricity and gas savings only persist for 5 years and that the electric savings decay at 20% in the first year after treatment and then at a higher rate for the second, third, fourth, and fifth years after treatment. After the fifth year, the savings completely decay. Gas savings decay at a faster rate.

Table 2-26 presents the TRM persistence factors[109] for new savings as a function of years since the savings were first realized. The savings persistence factors equal one minus the cumulative savings decay rate.

Table 2-26 Illinois TRM HER Savings Persistence Factors

Fuel	Persistence factor for year t new savings[110] in year t	Persistence factor for year t new savings in year t+1	Persistence factor for year t new savings in year t+2	Persistence factor for year t new savings in year t+3	Persistence factor for year t new savings in year t+4
Electricity	100%	80%	54%	31%	15%
Gas	100%	45%	45%	9%	4%

For example, with an annual customer attrition rate of α, new savings in program year 3 in Illinois would equal:

$$s_{n,3} = s_3 - 54\% * (1-\alpha)^2 s_{n,1} - 80\% * (1-\alpha) s_{n,2}$$

In this calculation, before accounting for attrition in the number of treated customers, 80% of year 2 new savings are assumed to persist to year 3 and 54% of year 1 new savings (all savings are new) are assumed to persist to year 3.

107 Also, see NMR (2017) for application of this protocol's framework to a HER program in Connecticut and ADM (2018) for application to a Utah HER program.

108 See Jenkins et al. (2017) for a description of the Illinois TRM framework development.

109 See: https://www.ilsag.info/wp-content/uploads/IL-TRM_Effective_010121_v9.0_Vol_4_X-Cutting_Measures_and_Attach_09252020_Final.pdf

110 New savings are the sum of avoided decay savings and incremental savings in year t.

The IL TRM determined the HER savings persistence factors based on empirically estimated HER savings persistence factors for electricity and gas utilities inside and outside of Illinois. The TRM persistence factors will be updated as findings from new studies about HER savings persistence become available.

(ii) PA TRM

The Pennsylvania (PA) TRM also assumes a multiyear HER measure life and incorporates, with modifications, the previously described savings accounting framework. The TRM assumes that HER savings decay continuously at a linear rate of 31.3% per year for program populations treated for 2 or more years. The savings decay factor was based on analysis of HER savings decay for Pennsylvania electric utility HER programs that paused delivery of energy reports. The savings persistence rate is assumed to be 0% for the first year of treatment.

Table 2-27 Pennsylvania TRM HER Electricity Savings Persistence Factors[111]

Fuel	Persistence factor for year t new savings[112] in year t	Persistence factor for year t new savings in year t+1	Persistence factor for year t new savings in year t+2	Persistence factor for year t new savings in year t+3	Persistence factor for year t new savings in year t+4
Electricity	100%	0%	0%	0%	0%
Gas	100%	84%	53%	22%	0%

The continuous linear decay rate means implies that following the second treatment year, 15.65% of second-year savings, all which are assumed to be “new,” decay in the next year or, equivalently, that 84.4% (=1-0.313*(1-0.5)) persist.[113] Similarly, 53.1% (=1-0.313*(2-0.5)) of second-year savings persist after 2 years.

As an example, with an annual customer attrition rate of α, new savings in program year 4 would equal:

$$s_{n,4} = s_4 - 53.1\% * (1-\alpha)^2 s_{n,2} - 84.4\% * (1-\alpha) s_{n,3}$$

Because it is assumed none of the annual savings from the first program year persist, the first-year savings do not enter the calculation of savings for program year 4.

2.7.4.7 Estimating BB Savings Persistence

This section describes how evaluators can design studies to obtain estimates of savings persistence and savings decay for BB measures.

2.7.4.7.1 Study Design

This protocol recommends that evaluators employ RCTs to estimate the persistence of BB savings after participants stop receiving treatment. The implementation of an RCT to estimate savings persistence should proceed similarly to the implementation of RCTs previously discussed in this protocol.

---

111 The savings persistence factors were calculated using the default annual decay assumption of 31.3% and the persistence formulas in the PA TRM. New savings are the sum of avoided decay savings and incremental savings in year t.

112 New savings are the sum of avoided decay savings and incremental savings in year t.

113 If 31.3% of HER savings decay after 1 year, the average rate of savings decay over the year is 31.3%*0.5.

Figure 2-14 illustrates an RCT savings persistence experiment. The program administrator is assumed, as in Figure 2-9, to have implemented the BB program as an RCT with an opt-out design: customers from the study population were randomly assigned to receive the treatment or to a control group and treated customers can opt out of the program. To economize on space, Figure 2-14 does not show the utility's option at the beginning of the program to screen customers or that after treatment begins customers can opt out of the program.

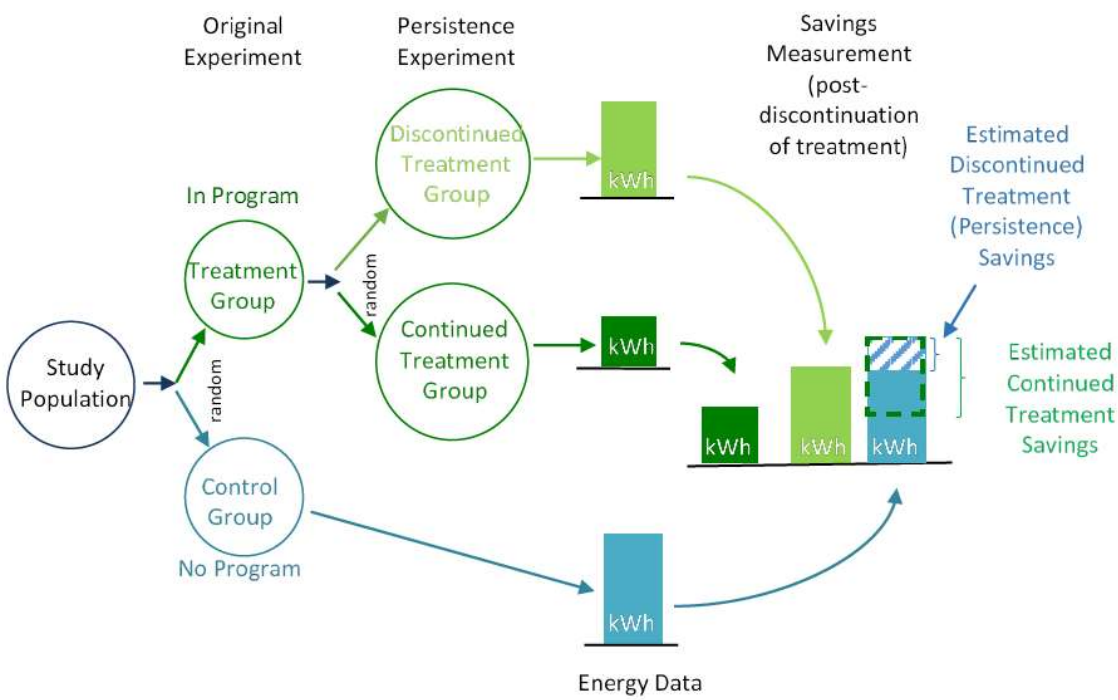

Figure 2-15 Illustration of Savings Persistence Study Design[114]

The persistence study starts after treatment group customers have received treatment for some duration (e.g., 1, 2, or 3 years). Though not illustrated, the utility may choose to screen the treatment group (and the control group) and study persistence for a specific subpopulation (e.g., by an energy use, socio-demographic, or housing characteristic). Also, the persistence study population must include treatment group customers who opted out, because evaluators will need to make energy use comparisons between the persistence study population and the original control group, which includes customers who would have opted out if they had been treated.

The next step is to randomly assign customers in the persistence study population to one of two groups. Customers in the "discontinued treatment" group will stop receiving the treatment; customers in the "continued treatment" group will continue receiving it. Evaluators should size, that is, assign enough customers to, the continued and discontinued customer treatment groups to detect the expected savings. The utility then administers the experiment and after enough time has passed collects energy consumption data for the report discontinuation period for control customers, discontinued treatment group customers, and continued treatment group customers to estimate savings persistence.

[114] *Ibid.*

To estimate savings for discontinued customers ("persistence savings"), the evaluator should compare the energy consumption of customers in the discontinued treatment group with the energy consumption of customers in the original control group during the discontinuation period. Under Savings Measurement in Figure 2-15, this difference is shown as the dashed, light-green box and represents the post-treatment savings for customers who no longer received the treatment.

The savings persistence rate can be estimated in two ways. This protocol recommends that evaluators compare the savings of the continued and discontinued treatment groups after treatment is discontinued. The continued treatment group savings represent the savings that the discontinued treatment group would have achieved if treatment had continued. Therefore, the ratio of the savings shows the percentage of the continued treatment group savings that persist after customers stop receiving treatment.

Evaluators can also estimate savings persistence by comparing the savings of the discontinued treatment group after treatment was discontinued with the group's savings before treatment was suspended. For evaluators wanting to measure savings persistence after a program administrator stops treating all customers in the behavior program, this approach is the only option. A limitation of this approach is, however, that savings may depend on weather, program implementation changes, or other time-varying factors, which, if not accounted for when comparing savings over time, can bias estimates of the savings persistence.

Both ways of calculating savings persistence only measure savings persistence rates for customers whose treatment was discontinued after a certain length of treatment (e.g., 2 years). Evaluators would need to conduct a series of discontinuation experiment to measure savings persistence for customers receiving treatment for fewer or greater number of years.

2.7.4.8 Estimating Savings Persistence

Suppose a utility started the treatment in period t = 1 and administered it for t*>0 periods. Beginning in period t = t*+1, the utility stopped administering the intervention for a random sample of treated customers. Evaluators can estimate the average savings per customer for a customer who continues to receive the treatment (continuing treatment group) and for those who stopped receiving the treatment after period t* (discontinued treatment group).

Assuming pretreatment energy consumption data are available, the following fixed effects DiD regression equation can be used to estimate savings during treatment and savings after treatment stops. This specification is estimated with consumption data for treatment and control group customers.[115]

$$kWh_{it} = \alpha_i + \tau_t + \beta_1 P_{1,t} \times T_{ci} + \beta_2 P_{1,t} \times T_{di} + \beta_3 P_{2,t} \times T_{ci} + \beta_4 P_{2,t} \times T_{di} + \varepsilon_{it}$$

Where:

kWh_{it} = electricity consumption by customer i in period t

α_i = A customer fixed effect (an unobservable that affects energy use for customer i); these effects can be estimated by including a separate intercept for each customer

[115] Evaluators can also implement a variant of the lagged dependent variable model (Eq. 7) to estimate savings persistence

τ_t = The time-period fixed effect (an unobservable that affects the consumption of all subjects during time period t); the time period effect can be estimated by including a separate dummy variable for each time period t, where t = -T, - T+1, …, -1, 0, 1, 2, ..., T

β_1= The average energy savings per continued customer caused by the treatment during periods t = 1 to t = t*

$P_{1,t}$ = An indicator variable for periods when customers in the continued and discontinued treatment groups received the treatment; it equals 1 if period t occurs between periods t = 1 and t = t* and equals 0 otherwise

T_{ci}= An indicator for whether customer i is in the continued treatment group; the variable equals 1 for customers in the continued treatment group and equals 0 for customers not in the continued treatment group β2 = The average energy savings per discontinued customer caused by the treatment during periods t = 1 to t = t*

T_{di} = An indicator for whether customer i is in the discontinued treatment group; the variable equals 1 for customers in the discontinued treatment group and equals 0 for customers not in the discontinued treatment group

β_3 = The average energy savings from the treatment for customers in the continued treatment group when t>t*

$P_{2,t}$= An indicator variable for periods when continued treatment group customers received the treatment and discontinued treatment group customers did not receive the treatment; it equals 1 if period t occurs after t = t* and equals 0 otherwise

β_4 = The average energy savings for customers in the discontinued treatment group when t>t*

If the persistence study is implemented as an RCT, OLS estimation of Eq. 9 is expected to yield unbiased estimates of savings for customers in the continued treatment group (β3) and discontinued treatment group (β4) after the discontinued group stops receiving treatment.[116] To estimate savings persistence after treatment stops, evaluators can take the difference between savings during treatment (β2) and post-treatment savings (β4) for subjects in the discontinued treatment group or the difference between post-treatment savings for the discontinued treatment group (β4) and the same period savings for the continued treatment group (β3).

2.7.4.9 Practical Evaluation Considerations

Evaluators conducting experiments to measure BB savings persistence should be mindful of several issues. First, stopping delivery of HERs or other BB treatments to estimate savings persistence may involve loss of some energy savings from discontinued customers, especially if the measure life is 1 year or program administrators are prevented from claiming persistence savings from discontinued customers. Also, the suspension of treatment may not result in a commensurate reduction in program administration and implementation costs, so

[116] 60 Evaluators can test the identifying assumption that assignment of treatment group customers to the discontinued treatment group was random by comparing the consumption of continuing and discontinuing treatment group subjects prior to the first treatment. If assignment was done at random, there should not be statistically significant differences in consumption between the two groups during this period.

that the program's cost-effectiveness may be adversely affected. It is also possible that suspending reports or treatment may dissatisfy some utility customers grown accustomed to receiving treatment.

Program administrators not wanting to conduct their own experiments can use findings about savings persistence from other studies but should borrow from studies that are valid for their own programs. As the rate of BB savings persistence may depend on climate; presence of other efficiency programs; BB program implementation strategies, including the frequency of prior treatment (e.g., quarterly vs. monthly); duration of prior treatment (number of years of treatment); and the form of the treatment (e.g., electronic or paper HERs); savings persistence estimates for one group of utility customers may not apply to other groups.

Finally, the equation in section Estimating Savings Persistence2.7.4.8 *Estimating Savings Persistence* above estimates savings for continued and discontinued treatment group customers for the actual weather during the analysis period, but evaluators may want to normalize the persistence savings estimates for year-to-year variation in weather. To obtain weather-normalized savings, evaluators can estimate savings as a function of cooling and heating degrees by adding stand-alone heating and cooling degree variables and three-way interaction variables of degrees with each of $P_{1,t}*T_{ci}$, $P_{1,t}*T_{di}$, $P_{2,t}*T_{ci,}$ and $P_{2,t}*T_{di}$ to the right side of the equation in section Estimating Savings Persistence2.7.4.8 *Estimating Savings Persistence*. All independent variables in the equation in section Estimating Savings Persistence2.7.4.8 *Estimating Savings Persistence* would also remain in this enhanced specification. For example, in a savings model estimated with monthly billing data, the coefficients on the interaction terms would indicate how savings before and after discontinuation of treatment depended on HDDs and CDDs. The coefficients on the two-way interaction variables $P_{1,t}*T_{ci}$, $P_{1,t}*T_{di}$, P2,t*T_{ci}, and $P_{2,t}*T_{di}$ would indicate the average savings unrelated to weather. Estimating this specification requires within-time period (e.g., a day or month) variation between customers in heating and cooling degrees. Without such variation, it is not possible to isolate the consumption's impact of weather from the impacts of other time-specific factors, which the time-period fixed effects account for.

2.7.5 REPORTING

BB program evaluators should carefully document the research design; data collection and processing steps; and analysis methods; and plan for calculating savings estimates. Specifically, evaluators should describe:

- The program implementation and the hypothesized effects of the behavioral intervention
- The experimental design, including the procedures for randomly assigning subjects to the treatment or control group. This should also include a careful description of the impacts measured by the experiment.
- The sample design and sampling process
- The processes for data collection and preparation for analysis, including all data cleaning steps
- Analysis methods, including the application of statistical or econometric models and key assumptions used to identify savings, including tests of those key identification assumptions
- Results of the savings estimation, including point estimates of savings and standard errors and full results of regressions used to estimate savings. Evaluators should clearly state the time periods to which the savings estimates pertain.
- Assumptions about measure life and savings persistence. If a behavior-based measure has a multiyear measure life, evaluators should describe the calculation of persistence savings and new savings.

A good rule of thumb is that evaluators should report enough detail such that a different evaluator could replicate the study with the same data. Every detail does not have to be provided in the body of the report; many of the data collection and savings estimation details can be provided in a technical appendix.

2.8 Protocols for Evaluating Demand Response Programs and Projects

2.8.1 INITIALISMS AND DEFINITIONS

ARC Aggregator of Retail Customers

Businesses that combine one or more retail customers and represent those customers' combined capabilities for demand response in the wholesale markets

BLDR Batch Load Demand Response

A special category of DRR-Type I resource that can reduce its load, or maintain its already reduced load beyond the normal BLDR duty cycle, to provide Demand reduction for economic, reserve, or Emergency services

BPM Business Practices Manual

A set of manuals designed to provide Market Participants with detailed information regarding how to conduct business in the various markets administered by MISO

BTMG Behind the Meter Generation

(1) General: Electrical generation that due to its location and metering is not "seen" by MISO through telemetry. (2) Specific: A defined term in the Tariff that refers to Behind the Meter Generation participating as a Load Modifying Resource in the MISO markets.

CPNode Commercial Pricing Node

A nodal level created for commercial purposes that aggregates certain EPNodes; all Market Settlement activity is performed at a CPNode, and it is the level where LMPs and MCPs are publicly available

DR Demand Response

Interruptible Load or Direct Control Load Management and other resources that can reduce Demand during an Emergency

Demand Response Event - A period of time defined by the System Operator, including notifications, deadlines, and transitions, during which Demand Resources provide Demand Response. All notifications, deadlines, and transitions may not be applicable to all Demand Response products or services.

DRR Demand Response Resource

Retail customer facilities or operations that are capable of voluntarily reducing their demand on the system

DSRI Demand Side Resource Interface On-line User Guide

The guide to manage Load Modifying Resources, including obtaining access state availability of assets, receive and respond appropriately to scheduling instructions, deploy resources, view event history, and participate in drills.

EDR Emergency Demand Response (Initiative or Resource)

A MISO-classification that provides for load reductions under Emergency conditions

EPNode Elemental Pricing Node

The lowest level of nodal relationship in the MISO market; EPNode are modeled as part of the Physical Network Model to represent points on the Transmission System where energy is injected or withdrawn

FERC Federal Energy Regulatory Commission

ISO Independent System Operator

LBA Local Balancing Authority

An operational entity that is responsible for compliance to NERC for certain Reliability Standards

LMP Locational Marginal Price

A nodal price for energy that combines the price of energy, transmission losses, and congestion

LMR Load Modifying Resource

A Tariff term that refers to resources that have qualified as planning resources, that is, resources that contribute towards the system's ability to meet the resource adequacy requirement. LMRs consist of two distinct resource types: Demand Resources and Behind the Meter Generation.

LSE Load Serving Entity

The business that provides power to retail customers

MCP Market Clearing Price

An equilibrium price paid for various ancillary reserves

MECT Module E Capacity Tracking tool

The Web-based computer program and interface that allows Market Participants to enter various data related to their loads and Module E requirements.

MISO Midcontinent Independent System Operator, Inc.

The operator / administrator of the transmission grid

MP Market Participant

A legal entity that is qualified, pursuant to procedures established by MISO to: Submit Bilateral Transaction Schedules; Submit Bids to purchase, and /or Offers to supply electricity in the Day-Ahead and/or Real-Time Energy Markets; Hold Financial Transmission Rights (FTRs) and submit bids to purchase, and /or offers to sell such rights; and Settle all payments and charges with MISO

NAESB North American Energy Standards Board

The NAESB Business Practice Standards developed terms for product/service categories demand response resources may provide, evaluation of performance, and other aspects of M&V to establish common terminology and criteria that could be used for wholesale and retail demand response programs.

NERC	North American Electric Reliability Corporation
PRA	Planning Resource Auction
	An annual auction held to allow Load Serving Entities an opportunity to meet their obligations for obtaining required capacity for a given Planning Year
PRMR	Planning Reserve Margin Requirement
	The total capacity requirement, measured in MW, for an LSE, based on its customers' load coincident with MISO's peak during the planning year
RSG	Revenue Sufficiency Guarantee
	RSG is the financial mechanism through which MISO obtains and transfers funds to offset direct costs incurred by suppliers that are not compensated through normal market prices.
RTO	Regional Transmission Organization
SCED	Security Constrained Economic Dispatch
	A model that selects units for dispatch from among those previously committed on the basis of their marginal economic costs
SCUC	Security Constrained Unit Commitment
	A model that selects units for commitment on a co-optimized basis, based upon their economic offers, operational parameters, and congestion
Tariff	Open Access Transmission, Energy and Operating Reserve Markets Tariff
	The FERC-approved set of rules under which MISO operates
TDRL	Targeted Demand Reduction Level

2.8.2 BACKGROUND

2.8.2.1 Purpose of this Protocol

This document provides guidance on methods for M&V of demand response (DR) in wholesale and retail markets. The document is intended for use by designers and operators of DR programs and market mechanisms, by regulators, and by participants or potential participants in wholesale and retail DR program offerings.

Measurement and verification for DR means the determination of the demand reduction quantities. This document addresses M&V for DR in 2 broad contexts:

1. Settlement, meaning determination of the demand reductions achieved by individual program or market participants, and of the corresponding financial payments or penalties owed to or from each participant.
2. Impact estimation, meaning determination of program-level demand reduction that has been achieved or is projected to be achieved, used for ongoing program valuation and planning.

Some parties are accustomed to thinking of M&V primarily in the context of settlement, and some primarily in the context of impact estimation. In this document, we recognize the importance of measured reductions in both contexts for effective DR design and operation and draw linkages between the two.

This work is based on a product of the National Forum for the National Action Plan on Demand Response (NAPDR) which was developed with a goal of helping states to advance the development and deployment of demand response resources. This work contributes to that goal by helping to establish credible measurement of demand reductions provided by DR resources. This document describes M&V methods that work best in various market and program contexts, as well as identifying the types of inaccuracies to which different methods are subject. Also addressed are the relationships among different aspects of DR program design (e.g., payment/penalty levels and structure, characteristics of demand response resources (e.g., weather sensitivity and variability of load, and M&V method specification)).

The intent of this document is to provide common language and guidance on best DR M&V practices in various market and program contexts including wholesale capacity or Measurement and Verification for Demand Response energy markets, and DR programs in retail markets, all with varying operating rules. The document generally follows the terminology and framework of the NAESB Business Practices Standards document on Measurement and Verification for DR, and provides additional guidance, and the Midcontinent Independent System Operator, Inc. (MISO) Business Practices Manual (BPM) for Demand Response in the second half of this protocol provides guidance for demand response when a utility is connected to a system aggregator.

2.8.2.1.2 Areas Addressed

This work includes:

- A framing discussion of demand response as a resource, with an overview of the role of M&V, also referred to as performance evaluation
- A review of the NAESB Business Practice Standards for DR M&V. These Business Practice Standards are directed to the determination of achieved DR demand reduction quantities and provide some basic terminology for describing M&V methods
- Guidance on M&V methods for settlement, including design considerations and continuing challenges.
- Guidance on impact estimation methods
- Guidance on conducting DR programs while participating in Independent System Operator (ISO) and Regional Transmission Organizations (RTO)

2.8.2.2 Organization

A brief description of each section is listed below.

- Section 2.8.3 *The Role of M&V for a Demand Response as a Resource* discusses demand response as a resource, an overview of measuring demand response and applications for M&V.
- Section 2.8.4 *NAESB Business Practice Standards* provides a review of the NAESB Business Practice Standards for DR M&V. These Business Practice Standards are directed to the determination of achieved DR demand reduction quantities.
- Section 2.8.5 *M&V Methods for Settlement* provides detailed information on developing an M&V methodology, from fundamentals through design considerations and continuing challenges.
- Section 2.8.6 *Impact Estimation* discusses the purpose of impact estimation, impact estimation methods for DR, and suggested applications of impact estimation methods.
- Section 2.8.7 *DR When Connected to System Aggregator* Discusses rules, regulations, and procedures for demand response when Entergy New Orleans is connected to a system aggregator.

Supporting appendices, contained in Volume 3 of this TRM:

- Appendix B: Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments
- Appendix C: Prior work in DR M&V Methods

2.8.3 THE ROLE OF M&V FOR A DEMAND RESPONSE AS A RESOURCE

2.8.3.1 Demand Response as a Resource

With proper program and M&V design, demand response can be a reliable, measurable, and verifiable resource in retail and wholesale markets. The challenge program designers and administrators face are that treating load as a supply resource creates a fundamental evaluation problem: how to accurately measure that which cannot be directly observed (i.e., the “but-for” load). There is no unambiguous, incontrovertible way to measure what the load otherwise would have been. The goal of M&V design is to develop a performance evaluation methodology that can provide the best estimate of what the load would have otherwise been, appropriate for the product or service being provided.

Some wholesale or retail electric systems rely upon reduced demand (as an alternative to increased supply) and pay participants based on the amount reduced. A measurement of the quantity of demand reduced relative to a customer-specific baseline is used for the operation and settlement of these systems. Historical performance can be evaluated to estimate expected response of an individual resource, or to adjust the amount of capability that a resource is able to offer into a market in a future period. Historical performance can also be used to estimate the amount of demand response for planning and forecasting. Transparency and fairness of baselines, retrospective assessments, and the accuracy of short-term forecasts all contribute to resource reliability and market confidence. Providing guidance on developing a performance evaluation methodology is a major focus of this document and is addressed in detail in Section 2.8.5 *M&V Methods for Settlement*.

The quantity of demand reduced for a program or market mechanism as a whole and by component is determined via impact evaluation. This aggregate measurement is needed for a range of purposes, from retrospective regulatory oversight to long-term planning studies and day- or hour-ahead operator forecasts. Section 2.8.6 *Impact Estimation* describes uses of and methods for DR impact evaluation.

2.8.3.2 Measuring Demand Response[117]

Measurement of any demand response resource typically involves comparing observed load during the time of the curtailment to the estimated load that would otherwise have occurred without the curtailment. The difference is the load reduction. The load reduction is positive if the observed load is less than the estimated load absent a curtailment, negative if the observed load is greater.

For demand response, the market product defines how the load reduction is valued and measured. Many demand response programs use a baseline methodology to estimate the load level without a curtailment for each participating resource. Other performance evaluation methodologies may also be used, depending on the product or service provided (see Section 2.8.4 *NAESB Business Practice Standards*). Actual metered load data, or

[117] Although the term “measurement” is widely used in the industry, DR reduction quantities cannot be measured in the same sense that load and generation quantities can be measured through precise metering. Rather, DR “measurement” is in most cases an estimation process, as described further in this document.

an alternative value, is compared to the "no-curtailment" estimate to determine the reduction amount for performance and settlement.

Any estimate of what the load would have otherwise been subject to some error. This error should neither be ignored nor exaggerated. Rather, the estimation error can and should be understood and managed.[118]

This document provides general guidance to help understand how various features of program design, performance evaluation method design, and participants affect estimation error in different contexts. The document also offers methods for assessing the estimation errors in a specific context and suggests strategies for managing and mitigating these errors through design choices and revisions.

As background for the discussion of alternative M&V approaches, general concepts for understanding DR estimation error are discussed in Section 2.8.3.4 *Understanding and Managing Estimation Error For DR*. First, we review the different uses of M&V for DR.

2.8.3.3 Applications For M&V

M&V for DR is used for:

- Establishing the eligibility or capability of resources;
- Retail settlement;
- Wholesale settlement;
- Projecting the future performance of an individual resource based on its past performance relative to its capability;
- Impact estimation of a program or product as a whole; and
- Forecasting and planning.

Different methods may be used for each of these purposes. Across these applications, the M&V methodology and its accuracy affect incentives and payments to participants, costs borne by the market as a whole, program operations, forecasts, and re-design. The purposes are described further below.

2.8.3.3.1 Establishing Resource Capability

For most products and services that demand response can provide, the capability of the resource needs to be established before the resource can participate in the demand response program. The methodology for capability measurement may be applied for an individual end user participating as a resource, or for an aggregated resource as a whole. The capability assessment may be as simple as the deemed capability of the appliance that is being controlled through direct load control. The assessment may be something more complex like determining the maximum demand over a fixed period of time so that a resource can offer its capacity into a wholesale market. Alternatively, either a retail or wholesale program might require an actual demonstration of capability before the resource is permitted to offer the demand reduction into the program.

[118] Throughout this document, the term "error" is defined as difference between the estimated value and the actual value of interest. Although the actual value may not be observable, there are means of assessing the magnitude of the estimation error, as described in section **Error! Reference source not found.**.

2.8.3.3.2 Settlement

DR settlement is the determination of demand response quantities achieved, and the financial transaction between the program or product operator and the participant, based on those quantities.[119] The wholesale market operator settles the market and determines the financial flows to and from the wholesale market DR participants for their performance. Retail DR program operators determine performance-based settlement with their program participants.

For demand response programs that pay an incentive for load reductions provided, the estimated load without curtailment determines the calculated reduction quantity that is the basis for settlement with each demand response resource. In the wholesale market, the DR resource may be an individual end-use customer, but more commonly is an aggregate of end-use customers operated by a DR aggregator, or the total of a DR program operated by a retail Load Serving Entity (LSE). Wholesale settlement is between the market and the market-participating DR resource. Retail settlement is between the DR aggregator or retail program operator and the end-use customer participating in the aggregation or the retail program.

In retail demand response programs, payment to end-use customers may not depend on each customer's estimated load reduction but may be based only on participation. For example, a direct load control program may pay a single seasonal incentive for the right to control load or may pay a fixed amount for each control event. However, if the retail program is offered into the wholesale market as an aggregated DR resource, the program operator will typically be settled according to an estimate of the load reduction quantity for each wholesale DR event. In wholesale markets, settlement often includes not only payments for load reductions achieved, but also penalties if the reduction achieved is below a committed amount. More generally, different M&V may be used to settle between a retail program operator and its customers than is used to settle that program as an aggregated resource in the wholesale market.

An LSE operating a retail DR program does not necessarily offer that program as a wholesale market resource. Rather, the retail operator may use DR to manage its own supply costs and settle in the wholesale market only for the actual load of its customers (i.e., the final aggregated load of its customers after DR reductions). In this case, the measurement needed for load settlement in the wholesale market is the LSE's aggregated load by interval (by market zone or node). The aggregated interval load comes either directly from summing interval meters, or from a load profile estimate. However, even if measured reductions are not required for settlement either with retail participants or with the wholesale market, DR M&V via impact estimation is valuable for assessing program effectiveness and for ongoing planning.

Table 2-28 below indicates some common retail DR structures, and the corresponding M&V needed for retail and wholesale settlement. The M&V needs for these different contexts are discussed further below. Also indicated in the table is the M&V need for impact estimation. Impact estimation itself has multiple uses and methods, as discussed in Section 2.8.6 *Impact Estimation*.

As the table indicates, there are a variety of arrangements a retail operator may have with its DR customers; many of these program structures do not require measurement of demand reduction as the basis for settlement

[119] More generally, for example, an ISO "administers and oversees the commodity market for buying and selling electricity within [a]. . . region. The ISO settlement process is used to determine the charges to be paid to or by a market participant to satisfy its financial obligations. The process measures the amount of energy purchased and sold through the energy market and arrives at each market participant's payment." http://www.iso-ne.com/nwsiss/grid_mkts/how_mkts_wrk/multi_settle/index.html

with the retail customer or DR aggregator. However, when the program- or segment-level reduction is offered as a wholesale resource, the measured demand reduction amount for the program or segment is typically needed for wholesale settlement. For all program types, if impact estimation is conducted, its primary purpose is to determine the quantities of demand reduction achieved by the DR program. The focus of this document is on measuring the quantity of demand reduction for settlement and for broader impact estimation contexts. Particular emphasis is placed on wholesale and retail settlement using baseline methods (see highlighted cells in Table 2-28).[120]

Table 2-28 M&V Needs for Common DR Contexts

Retail Program or Service Structure	Common Applications	M&V Needed for Participant Settlement with Retail Program Operator	M&V Needed for Participant Settlement with Wholesale Market	M&V Needed for Program-Level Impact Estimation
Customer or retail DR aggregator is paid per demand reduction amount	Demand Bidding/Buyback, Peak-Time Rebate	Measured demand reduction for the individual customer or DR aggregator	Measured demand reduction for aggregate	Measured demand reduction for aggregate
Customer is paid based on participation metrics	Mass market Direct Load Control	Verification of event participation	Measured demand reduction for aggregate	Measured demand reduction for aggregate
Customer pays for usage by time interval	Dynamic of fixed time-varying rates	Metered usage by time interval	Measured demand reduction for aggregate	Measured demand reduction for aggregate
Customer pays a penalty/surcharge above a pre-set load level	Contract for differences, firm load demand response, curtailable rates	Metered usage by time interval	Measured demand reduction for aggregate	Measured demand reduction for aggregate
None-- end-use customer participates directly in the wholesale market	Large customer as direct wholesale market participant	N/A	Individual demand reduction	Individual demand reduction
End-use customer participates in the wholesale market via a DR Aggregator	End-use customer enrolled in wholesale DR aggregator and rewarded through agreed sharing of DR payments	Measured demand reduction for the individual customer	Measured demand reduction for aggregate	Measured demand reduction for aggregate

2.8.3.3.3 Impact estimation

Impact estimation is the determination of the response that occurred to a given event, curtailment instruction, dispatch or set of events. At its most granular level, impact estimation estimates the demand reduction of a single demand response resource for a given interval. However, the purpose of impact estimation is ordinarily to

[120] Goldberg, Miriam L, and G. Kennedy Agnew. Measurement and Verification for Demand Response (2013)

provide estimates for a program or product as a whole, or for market segments, across a program season or year.

Impact estimation can support reporting of response on an event, daily or longer period, for a program or product overall. This information is used by stakeholders, system planners, reliability organizations, and regulators. Impact estimation is used not only as a “scorecard” on past performance, but also to develop or revise policies about the eligibility, treatment, and levels of demand response.

Ex post or retrospective estimation is the determination of savings achieved by a product or program over a particular span of time. This result is used to confirm or revise the *ex ante* or prospective assessment of program effectiveness or cost-effectiveness. *Ex post* estimation may also provide the basis for adjusting projections for future program operations.

Ex ante models can also be developed from impact evaluation results, to estimate demand reduction quantities as a function of event conditions including participation and weather. As described in section B.8.6, the resulting program-level *ex ante* estimates can be used to settle a retail program in a wholesale market.

In many instances, impact evaluation estimates of demand reduction are distinct from the estimates of demand reduction for settlement. Estimates of demand reduction for settlement need to occur within a short time of each curtailment event and must use calculation methods explicitly specified as part of the program rules. These requirements limit the range of feasible methods for securing the estimates. Impact evaluation demand reduction estimates can represent a more accurate estimate of load reduction given more data, a longer time frame, and sufficient time to apply more rigorous methods than are feasible for short term settlement.

Impact estimation is discussed further in Section 2.8.6 *Impact Estimation*.

2.8.3.3.4 Projecting Individual Resource Performance

For an individual DR resource, the estimated demand reduction quantities for individual events can be used not only for settlement, but also to assess the resource’s performance over a period of time. For each resource, a performance factor can be calculated reflecting the load reduction achieved compared to the resource’s committed reduction. For example, the NYISO calculates a performance factor for each individual resource as the maximum observed load reduction amount over a season, as a fraction of the commitment. Such “performance factors” can be used by aggregators and program administrators to assess the dependability of the individual resource to provide the level of reduction that it has committed to the demand response program.

To calculate performance factors, the “observed” load reduction may be the quantities used for settlement, as in the case of the NYISO, or could be determined by a more comprehensive impact evaluation. The design of this performance evaluation method needs to ensure consistency with the objective of the program, provide an accurate estimate of the “but-for” load, and align with treatment of other suppliers of the same products.

2.8.3.3.5 Forecasting and Planning

Load forecasting is estimation of load on an hourly and daily basis in advance of the operating day. Load forecasting is conducted on a long-term basis of one or more years ahead as part of resource planning, as well as on a day- and hour-ahead basis for operations.

In this context, DR M&V is used primarily to develop *ex ante* estimates of future load reduction capability for long-term forecasts, and to estimate reductions that will be achieved if an event is called in short-term operations.

DR M&V is also needed to construct the "reconstituted" total load that would have occurred in each control area, zone, or node if past DR the events had not been called. This reconstituted load is the basis for projecting the total future load to be served by the combination of supply- and demand-side resources.

Errors in estimates of past load reductions will also affect load forecasts developed from the reconstituted load determined from those estimates. The resulting load forecast errors may either overstate or understate the load, and in the short term may result in under-scheduling or over-scheduling of supply to meet the forecasted load.

System planners may also include demand response as a supply resource in resource adequacy planning. The M&V designed for measuring response of the individual or aggregated resource then affects long-term planning functions.

2.8.3.4 Understanding and Managing Estimation Error For DR

2.8.3.4.1 Measuring What Can't Be Observed

When creating mechanisms for load to participate in wholesale markets as a resource, a general principle is that load should be subject to the same requirements as generation, to the extent practical. It therefore may seem natural to require that load reductions be measured with the same accuracy as is required for metering of generation.

However, as noted above, there is a fundamental difference between load reduction and generation as resources: *It is not possible to meter or otherwise directly observe load reductions*. Rather, measurement of the performance of any demand-side resource necessarily means comparing observed load to an estimate of the theoretical load that would have occurred absent the resource's being dispatched—that is, compared to a calculated baseline.

This baseline is an estimate of load at a condition we can't observe and is necessarily subject to some estimation error. Even though the theoretical load can't be observed, it's nonetheless possible to measure and manage the estimation errors. In the discussion that follows, we review the relationships among the key quantities produced by DR M&V, and the relationships among their estimation errors. We then describe broad strategies for understanding and mitigating the effects of estimation errors. These strategies are revisited in more detail in later sections of this paper.

2.8.3.4.2 Key Quantities Produced by DR M&V

Key quantities produced by DR M&V include:

- The calculated baseline load. This is the estimate of the theoretical load that would otherwise have occurred, or the "but-for" or "no-event load."
- The calculated reduction, or difference between the calculated baseline load and the observed load. This is the estimated reduction from the theoretical no-event load
- The financial settlement amounts, that is the payments and penalties based on the calculated reduction.

All of these quantities are subject to estimation error, and these estimation errors are directly related to one another. The discrepancy between the calculated baseline and the theoretical no-event load produces a discrepancy in the calculated load reduction of the same MW magnitude: If the load estimate is high or low by 20 MW, the load reduction calculation will be off by the same 20 MW in the same direction. The discrepancy in the calculated reduction in turn results in a discrepancy between the financial settlement amounts compared to the settlements that would be made if the theoretical no-event load were observed.

In this document, when we refer to M&V accuracy, we mean how close the calculated baseline, load reduction, or financial settlement is to the value that would be obtained if the theoretical no-event load were observable. We discuss how to assess and manage DR M&V accuracy below.

How load reduction discrepancies translate into financial settlement discrepancies depends on the program rules and market conditions. Over- and under-payments mean that the price signals given to participants are distorted or blurred. The result is a weakening of the price response, a possible reduction in cost-effectiveness of the program, and/or a shifting in benefits and costs among stakeholders. How severe these effects are depending on the size of the financial discrepancy. M&V, and M&V accuracy, are important for getting the financial transactions as close to “right” as possible.

2.8.3.4.3 Bias and Random Error

Measurement or estimation error consists of systematic and “random” components.

- Systematic error or bias is a tendency for the estimate to be higher on average or to be lower on average than the actual value. A measure of bias is the average difference between the estimate and the actual value.
- Random or nonsystematic errors are deviations up and down that on average are zero. A measure of the magnitude of random error, the typical level of variability.
- Up and down, is the standard deviation of differences between estimates and actual values.

The level and direction of systematic error and the level of variability for a particular estimation method usually depends on the characteristics of the participating resource, and on the operating conditions including time of day, calendar, and weather. For example, some methods will tend to overstate baselines on very hot days and understate on mild days, and the degree of this bias will vary across resources of different types. Resources with more regular load patterns will tend to have baselines with smaller random errors than those with more variable operations.

If the baseline estimate is systematically overstated or biased upward, the load reduction estimate will be systematically overstated by the same MW amount. Incentive payments to the participant will be biased upward as well. Conversely, if the baseline estimate is systematically understated or biased downward, the load reduction estimate will be systematically understated, and the incentive payments will be biased downward. Likewise, variability in the baseline translates into variability in calculated load reduction and in the corresponding incentives.

For both systematic and random error, a given magnitude error in the baseline becomes a proportionately much larger error in the estimated load reduction. For example, for a load of 200 kW with a 40kW reduction, a 20 kW error in the baseline is a 10 percent error in estimating load but a 50% error in estimating the load reduction.

The up and down random errors in baseline and in corresponding load reduction estimates will tend to balance out over events and customers. However, the effects on incentives may not balance out. For payments tied to market prices, an error in one direction may be settled at a high market price while an equal error in the opposite direction may be settled at a low market price. In addition, program payment and penalty schemes may involve threshold requirements that result in higher consequences for errors in one direction or the other.

Managing DR M&V Estimation Errors

The means by which the effects of M&V error can be managed and mitigated include the following four practices:

1. Assessing the magnitude of the systematic and random estimation error.
   Impact evaluation reports provide confidence bands7 for *ex post* and *ex ante* estimates and compare evaluated savings with the nominal DR quantities based on program settlement rules. This information can be used to adjust settlement procedures or quantities, or to modify the baseline estimation method used for settlement on a going forward basis.
   Baseline method assessment studies can provide estimates of systematic and random errors for different types of resources, in terms of demand level, reduction quantity, or payments for demand reduction. Methods for conducting such assessments are described in section B.8.5.5, Means to Assess Settlement M&V Accuracy.
2. Operational adjustments based on assessment of estimation errors.
   Dealing with systematic estimation errors for demand reduction can take multiple forms. One is to de-rate individual resources for observed and projected under- or over-achievement. Another is to incorporate adjustment factors into operational forecasts. Still another is to modify the program or demand reduction calculation methods to reduce these systematic errors.
   Systematic errors can be addressed by applying adjustment factors once the degree of bias is determined. Residual uncertainty can be mitigated in part by aggregating over many different resources. However, even in aggregate, the amount of DR that has been provided will typically have more measurement/estimation error than a corresponding supply-side resource. Nonetheless, even with some uncertainty in the measurement of the actual reduction delivered, the magnitude of the DR resource may still be sizable, and the DR can provide a valuable and reliable resource as long as the associated measurement error magnitude is known.
3. Program adjustments to mitigate effects of M&V errors.
   Programs can reduce the effects of M&V errors by a number of means. One is to change the baseline specifications to reduce some of the sources of error identified. Another is to change program rules to eliminate some of the factors that contributed to baseline errors. Another, when allowed, is to try to direct potential participants into the type of DR program best suited to them. Program design features that can improve M&V accuracy are discussed in Section 2.8.5.5 *Means To Assess Settlement M&V Accuracy.*
4. Program design as an iterative process.
   Program design, including M&V methods for settlement, must be subject to ongoing re-assessment and refinement. Programs are designed and prospectively assessed based on an expected participant profile. As programs are modified to address the issues experienced by current program participants, the participant mix may change as a result of the modifications. The next round of program design in turn addresses the issues and behavior of the new set of participants, and the cycle continues.

2.8.4 NAESB BUSINESS PRACTICE STANDARDS

2.8.4.1 Overview

The electricity industry has been moving towards development and adoption of a common set of terminology, definitions, analysis methods and protocols for DR products and services in recent years. The North American Energy Standards Board (NAESB) has developed Business Practice Standards for DR Measurement and Verification for wholesale and retail markets. The wholesale and retail standards were developed to be nearly the same, with some additional elements specific to retail business practices. A primary focus of the NAESB business practice standards is on M&V methods used for market operations and settlement, but the terminology applies also to other M&V applications.

The FERC, which regulates wholesale markets only, has adopted the Phase 1 version of the NAESB Business Practice Standards for DR M&V in wholesale markets, and has issued a Notice of Proposed Rulemaking (NOPR) to adopt the Phase 2 version. The Phase 2 standards, ratified by NAESB membership, expand, and clarify criteria described in the Phase 1 Business Practice Standards. This document uses the framework and terminology of the NAESB standards and offers additional discussion and guidance. Recommendations in this document are not proposed as standards.

2.8.4.1.1 Goals of the NAESB Business Practice Standards

Goals of the M&V standards are defined by NAESB as providing a common framework to ensure:

- Transparency: Facilitate market transparency by developing accessible and understandable M&V requirements for Demand Response products.
- Accountability: Promote accurate performance measurement of DR resources by system operator(s), in dispatch, operations management and market settlements.
- Consistency: Develop uniform and consistent methods and procedures applicable across all wholesale markets.

2.8.4.1.2 Scope of the NAESB DR M&V Standards

The NAESB DR M&V Business Practice Standards cover the following aspects of M&V:

- Provide standard terminology for defining program requirements, measurement methods, and data requirements;
- Identify elements that System Operators or Governing Documents must specify for each broad type of program and performance evaluation methods;
- Identify which elements and requirements are applicable to which broad types of methods (unless otherwise specified by the System Operator);
- Specify particular requirements for metering accuracy and granularity; and
- Identify five broad types of performance evaluation methodologies and related criteria.

The standards were not developed to provide specific requirements or guidance on how to specify particular elements of the performance evaluation methodologies. As a result, the NAESB Business Practice Standards do not:

- Provide guidance on best specifications for particular market/program rules and resource characteristics;
- Address the relationship between retail and wholesale DR M&V; or

- Address the relationship between M&V for settlement and program evaluation.

This document builds on the NAESB framework, adopting the terminology where applicable, to provide discussion and guidance on issues that were considered out of scope for the NAESB Business Practice Standards developed to date.

2.8.4.2 Key Terminology

The NAESB Business Practice Standards developed terms for product/service categories demand response resources may provide, evaluation of performance, and other aspects of M&V to establish common terminology and criteria that could be used for wholesale and retail demand response programs. Terminology from the NAESB Business Practice Standards has been incorporated into many demand response programs since the NAESB Business Practice Standards were ratified by NAESB members and incorporated into regulation by the FERC. The focus for this section will be on the terms relevant to performance evaluation methodologies.

- NAESB defines demand response as, “a temporary change in electricity usage by a Demand Resource in response to market or reliability conditions. For purposes of these standards, Demand Response does not include energy efficiency or permanent Load reduction.”
- NAESB defines a demand response event as, “a period of time defined by the System Operator, including notifications, deadlines, and transitions, during which Demand Resources provide Demand Response. All notifications, deadlines, and transitions may not be applicable to all Demand Response products or services.”
- An important distinction is required between demand response and demand reduction value which is defined as, “the measurement of reduced electricity usage by a Demand Resource during a Demand Response Event or Energy Efficiency performance hours expressed in MW.”

Demand response is the more general term, while demand reduction specifically refers to load reduction during a demand response event. Throughout this document, we attempt to be consistent regarding this usage.

Figure 2-16 adapted from the NAESB *Business Practice Standards for Measurement and Verification of Wholesale Demand Response*[121], illustrates the general framing of a Demand Response Event, and associated terminology. This chart is intended to illustrate event-based demand response, not the dispatch of demand response that is scheduled and dispatched in real-time as a supply resource. Not every demand response event will include every component shown in the chart.

---

[121] https://www.naesb.org//pdf4/dsmee_group2_040909w5.pdf

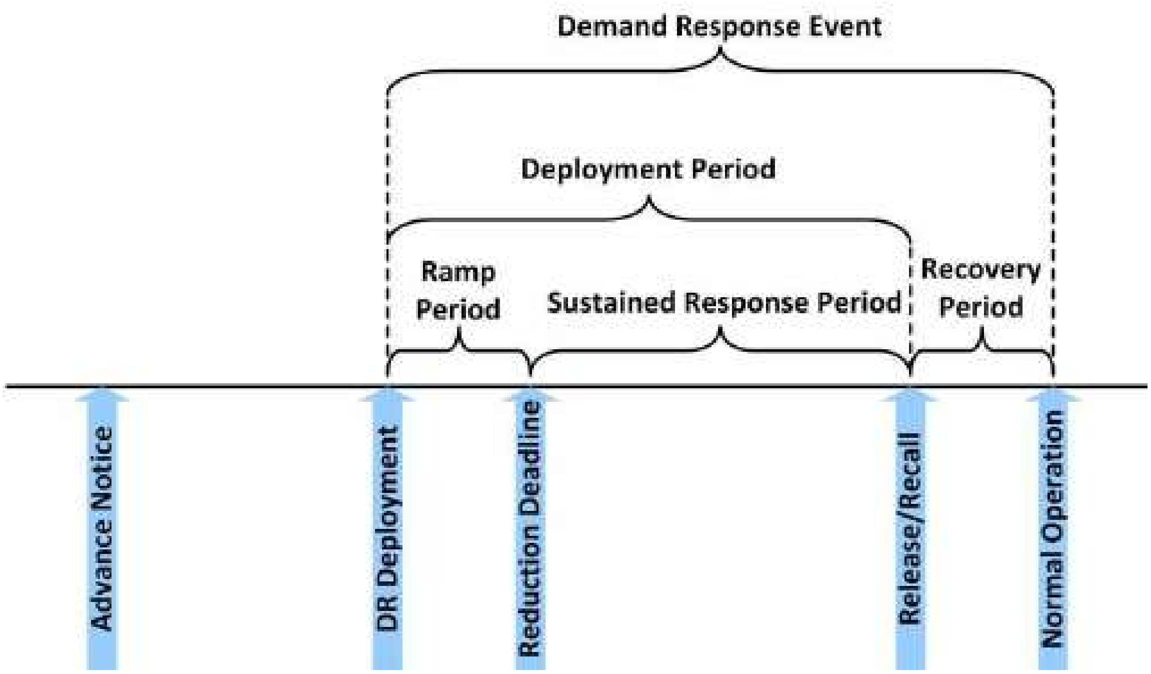

Figure 2-16 NAESB Demand Response Event Terms

2.8.4.2.1 Performance Evaluation Methodologies

Performance evaluation methodology refers to the approach taken to estimate the demand reduction value of the product/service provided by a demand response resource. Five performance evaluation methodologies have been defined in the NAESB Business Practice Standards:

- Maximum Base Load: A performance evaluation methodology based solely on a Demand Resource's ability to maintain its electricity usage at or below a specified level during a Demand Response Event.
- Meter Before / Meter After: A performance evaluation methodology where electricity Demand over a prescribed period of time prior to Deployment is compared to similar readings during the Sustained Response Period.
- Baseline Type-I: A Baseline performance evaluation methodology based on a Demand Resource's historical interval meter data which may also include other variables such as weather and calendar data.
- Baseline Type-II: A Baseline performance evaluation methodology that uses statistical sampling to estimate the electricity usage of an Aggregated Demand Resource where interval metering is not available on the entire population.
- Metering Generator Output: A performance evaluation methodology in which the Demand Reduction Value is based on the output of a generator located behind the Demand Resource's revenue meter.

These five performance evaluation methodologies are shown with the four service types defined for demand response in Table 2-29. The check marks indicate whether a performance evaluation methodology is applicable to specific product type.

Table 2-29 NAESB Service Types and Applicable Performance Evaluation Methodologies[122]

Performance Evaluation Methodology	Valid for Service Type			
	Energy	Capacity	Reserves	Regulation
Maximum Base Load	✓	✓	✓	
Meter Before/Meter After	✓	✓	✓	✓
Baseline Type-I Interval Metering	✓	✓	✓	
Baseline Type-II Non-Interval Metering	✓	✓	✓	
Metering Generator Output	✓	✓	✓	✓

2.8.4.2.2 Criteria for Performance Evaluation Methodologies

For each performance evaluation methodology, the *NAESB Business Practice Standards* provide applicable criteria to define; not all criteria are applicable to every performance evaluation methodology. The criteria are grouped together in three main categories: Baseline Information, Event Information, and Special Processing (see Table 2-30).

Table 2-30 NAESB Criteria for Performance Evaluation Methodologies[123]

Baseline Information	Baseline Window
	Calculation Type
	Sampling Precision and Accuracy
	Exclusion Rules
	Baseline Adjustments
	Adjustment Window
Event Information	Use of Real-Time Telemetry
	Use of After-the-Fact Metering
	Performance Window
	Measurement Type
Special Processing	Highly-Variable Load Logic
	On-Site Generation Requirements

Baseline Information. The criteria in this category cover the components used development of the estimated (“but-for”) load.

- Baseline Window: The range of data used for estimating the “but-for” load.
- Calculation Type: The arithmetic method used to compute the “but-for” load.
- Sampling Precision and Accuracy: Any sampling and accuracy requirements, if applicable, as for Baseline Type-II where interval meter data is not used.
- Exclusion Rules: Allowances for excluding any historic load data from the Baseline Window.

[122] Goldberg, Miriam L, and G. Kennedy Agnew. Measurement and Verification for Demand Response (2013)
[123] *Ibid.*

- Baseline Adjustments: Any calculations, based on a variety of conditions (such as temperature, humidity, event day operating conditions) for making adjustments to the baseline on the day of the event.
- Adjustment Window: The time period from which the adjustment data can be evaluated.

Event Information. This set of criteria covers the metering, data and measurement used for evaluating response.

- Use of Real-Time Telemetry: Specifies whether or not, real-time two-way communication with the program administrator is required for performance evaluation.
- Use of After-the-Fact Metering: Specifies whether or not after-the-fact metering can be used for performance evaluation.
- Performance Window: The period of time during the event that is used to evaluate the performance of the demand response resource.
- Measurement Type: The arithmetic method used to compute the demand reduction.

Special Processing. These additional considerations may need to be specified for demand response resources with highly variable load or behind-the-meter generation.

- Highly-Variable Load Logic: Any additional data requirements or calculations for treatment of highly variable loads providing demand reduction, either during an event or for determining the capability of the demand response resource.
- On-Site Generation Requirements: Any additional requirements for reporting the performance on on-site generation during an event.

2.8.4.3 Applications of NAESB performance Evaluation Methodologies

2.8.4.3.1 Energy Performance Evaluation Methodologies

The NAESB performance evaluation methodologies serve as a way to characterize the type of measurement used to estimate the reduction of a demand response resource. This report focuses on Baseline Type I and Type II to estimate energy response because they are the most common performance evaluation methodologies in use; these methods are typically used to estimate the amount of energy provided by a demand response resource during an event or schedule. Some demand response programs also use the Baseline Type I or Type II methodology to calculate the capacity provided during a demand response event, as described later in this section in *Capacity Performance Evaluation Methodologies*. Baseline Types I and II are frequently referred to as the Customer Baseline Load, or CBL.

The other three performance evaluation methodologies that are in use may be combined with a Baseline Type I or Type II. Metering Generator Output may be used in combination with a Baseline method for a generator that is used outside of DR events as well as to respond to these events. Products and services that require historical data beyond the data used in a Baseline Type I or Type II may incorporate a Maximum Base Load calculation Service types that require information closer to the real-time conditions of the demand response resource may use Meter Before/Meter After). As Table 2-29 indicates, most of the performance evaluation methodologies are applicable to all products and services. The design of the demand response program and the environment in which that program operates often provide the context for the performance evaluation methodology that will best align with the objectives of the program.

For Baseline Type I and Type II, the baseline calculation method can take many forms. The calculation method is specified by a combination of the baseline window, the exclusion rules, the calculation type, and the baseline adjustments and adjustment window. The combination of the baseline window and exclusion rule is intended to select days and hours that are similar to what the event day or period would have been absent the event. In many cases, the adjustments can make the baseline calculation less sensitive to the selection rules. Examples of criteria for Baseline Type I are provided below.

Baseline Window: A period of time preceding and optionally following a Demand Response Event over which electricity usage data is collected for the purpose of establishing a Baseline.

Examples of baseline windows include:

- the last 10 non-holiday weekdays;
- the 10 most recent program-eligible non-event days;
- the 10 most recent program-eligible days beginning 2 days before the event;
- the last 45 calendar days; or
- the previous year.

Exclusion Rules: Rules for excluding data from the Baseline Window. Common exclusion rules include:

- Excluding days with DR events.
- Excluding days with outages, or force majeure events.
- Excluding days with extreme weather.
- Excluding days with the highest or lowest loads.

Calculation Type: The method of developing the Baseline value using the data from the baseline window.

Examples of calculation types include:

- Average value: for each hour of the day, calculate the average of the load at that hour over the included days.
- Regression: calculate load by regressing the load from the included days on weather and other variables, usually with separate regression coefficients by hour of the day.
- Maximum value: take the maximum of the loads in the included period.
- Rolling average: the updated unadjusted baseline for an operating day is equal to 0.9 times the prior unadjusted baseline plus 0.1 times the most recent included day.

Baseline Adjustments: An additional calculation applied after the basic Calculation Type, to align the baseline with observed conditions of the event day. Factors used for adjustment rules may be based on but are not limited to; Temperature; Humidity; Calendar data; Sunrise/Sunset time and/or; Event day operating conditions.

Examples of baseline adjustments include:

- Additive: add a fixed amount to the provisional baseline load in each hour, such that the adjusted baseline will equal the observed load at a time shortly before the start of the event period.
- Scalar: multiply the provisional baseline load at each hour by a fixed amount or scalar, such that the adjusted baseline will equal the observed load on average during a window of time shortly before the start of the event period.

Adjustment Window: The period of time for which the adjusted baseline matches the observed load. The NAESB guidance is that the adjustment window shall begin no more than four hours prior to deployment. Examples of adjustment windows include:

- The hour before the event (hour -1)
- The 2 hours before the event (hours -1 to -2)
- The two hours that end two hours before the event (hours -3 to -4)

Sampling Precision and Accuracy: If the aggregate baseline is calculated from a sample of interval metering data (as for baseline Type II) the M&V method specification should include the statistical precision required. A common sampling precision requirement is that the load should be estimated so as to have a confidence interval that is +/- 10 percent of the estimate at a 90 percent confidence level. However, this precision standard, which derives from PURPA load research requirements, may or may not be appropriate for the operation of a particular program or market. Moreover, sampling accuracy is only one component of baseline accuracy. In general, better precision requires larger samples with higher associated metering costs.

The specific confidence and error levels of 90/10 precision are artifacts from PURPA and the world of load research. They may or may not serve the needs of DR M&V and, as a result, should be given due consideration.

Examples of baseline calculation methods, specifying data windows and exclusion rules, as well as the calculation method and adjustments are given in Volume 3, Appendix B: Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments. In addition, the ISO/RTO Council has a detailed table that lists the NAESB M&V parameters for the wholesale demand response programs across North America (link available in Appendix B: Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments.

2.8.4.3.2 Capacity Performance Evaluation Methodologies

This report does not address in detail the application of performance evaluation methodologies for estimating capacity response other than Baseline Type I or II approaches used to estimate the energy reduction provided by a demand response resource that has a capacity obligation. This is, in part, because the uses of performance evaluation methodologies for estimating capacity vary greatly.

Wholesale market demand response programs use a variety of methods to estimate the capacity of the resource from a comparable period, usually from the prior year. The program administrator may use the coincident peak load of the demand response resource, the average of multiple coincident peak loads, or something more complex that utilizes criteria of a Baseline Type I to estimate the maximum capacity of the resource.

For demand response resources that offer capacity, this maximum capacity often provides the upper bound that is used in conjunction with a Maximum Base Load performance evaluation methodology. The difference between the maximum capacity value and the Maximum Base Load that the resource can achieve during an event is the amount of capacity that the resource can enroll. For example, the Maximum Capacity Value may be the resource's historic peak load, while the Maximum Base Load is a demand level the resource commits not to exceed during an event. This relationship is illustrated in Figure 2-17.

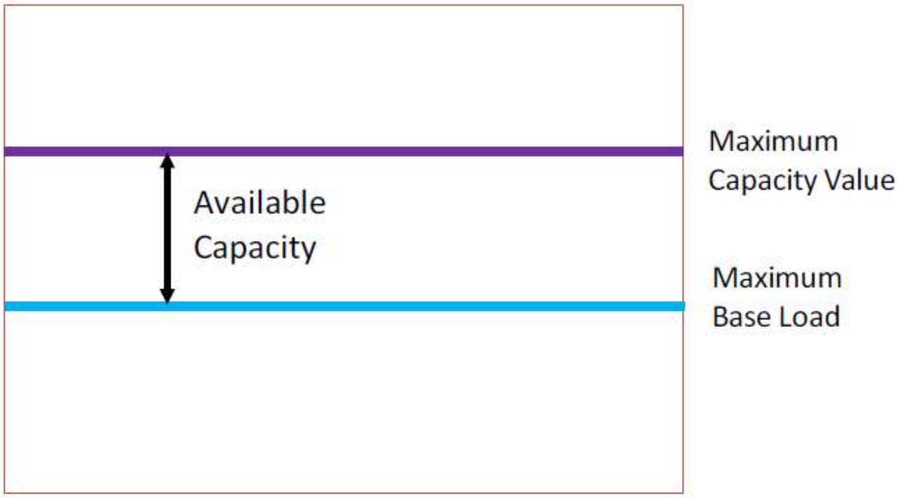

Figure 2-17 Illustration of a Maximum Base Load Performance Evaluation Methodology[124]

To estimate response after an event, the program administrator may use an energy baseline calculation, such as Baseline Type I or II. Alternatively, the program may calculate the demand reduction as the difference between the Maximum Capacity Value and the maximum interval metered load during the event; this measured reduction is then compared to the amount of capacity committed. For example, if a resource has a Maximum Capacity Value of 400 kW and a Maximum Base Load of 300 kW, the Available Capacity is the difference, 100 kW; if that resource has metered load of 320 kW during an event, the calculated demand reduction is 80 kW, or 80% of the committed amount. The Maximum Capacity Value, used to estimate the amount of available capacity in the illustration, may also be based on one of the types of performance evaluation methodologies, such as a Baseline Type 1 that uses a simple average of metered loads during certain peak hours,

Some capacity programs allow the resource to nominate the amount of capacity they can provide; these programs typically use the Baseline Type I energy performance evaluation methodology to estimate response.

2.8.4.4 Performance Evaluation Methodologies for Operating Reserves and *Regulation Service*

Demand response has demonstrated its potential in the ancillary services market by providing non-spinning reserves and regulation services in many markets.[125] For demand response resources that provide ancillary services, the performance evaluation methodologies may be similar to Baseline Type I, where the amount of energy reduction is measured from an estimated "but-for" load or may use any of the other applicable methods. The real-time nature of demand response providing these two services may lend itself to the use of the Meter Before/Meter After performance evaluation methodology, where change from a previous interval is measured, similar to a traditional supply resource. At the time of this report, the penetration of demand response providing ancillary services and details on common performance evaluation methods for these services are limited.

[124] Goldberg, Miriam L, and G. Kennedy Agnew. Measurement and Verification for Demand Response (2013)

[125] For example, PJM -- http://www.pjm.com/markets-and-operations/demand-response/dr-synchro-reserve-mkt.aspx, and ERCOT -- http://www.ercot.com/services/programs/load/laar/index Available CapacityMaximum Capacity ValueMaximum Base Load

2.8.4.5 Applying The NAESB M&V Terminology to Common Demand Response Program Concepts

Administrators of demand response programs may initially find it challenging to categorize their performance evaluation methodologies using the NAESB terminology. Table 2-31 lists some of the more common types of demand response programs and how those programs or program mechanisms align with the NAESB terminology and whether further discussion of the demand response program or program mechanism is included in this document. This summary indicates common examples and is not meant to be exhaustive of possible M&V applications to program mechanisms.

Table 2-31 Summary of Common DR Mechanisms and NAESB DR M&V Methods[126]

Program Mechanism	Market/Service Type	Resource/Customer Type	Applicable DR M&V Method	Further Guidance in this Document
Firm load: Reduce to pre-specified load on notification	Retail or Wholesale/Energy Capacity, Reserves	Any	Maximum Base Load Evaluation	Impact Estimation Approaches
Reduction from baseline	Retail or Wholesale/Energy Capacity, Reserves	Individual or aggregate loads, individually interval metered	Baseline Type 1 (interval meter)	Baseline methods by customer and program characteristics
		Individual or aggregate loads, NOT individually interval metered	Baseline Type 2 (not interval meter)	Baseline methods by customer and program characteristics
Reduction from baseline, short events	Retail or Wholesale/Energy Capacity, Reserves	Individual or aggregate loads, individually interval metered	Meter before/Meter after	None
		Aggregate loads, NOT individually metered	Baseline Type 2 (not interval meter)	Application of Meter before/Meter After for sample
Behind-the-Meter Generation	Retail or Wholesale/Energy Capacity, Reserves	Customer-sited generation	Metering Generator Output	Baseline methods applied to generation
Direct Load Control	Retail	Individual end users	N/A	Impact Estimation Approaches
Direct Load Control	Retail or Wholesale	Aggregate of retail participants	Baseline Type 1 or Type 2	Impact Estimation Approaches

In this table, a “Retail” market or service refers to a program or service operated by a load serving entity or DR aggregator to serve end use customers; A “Wholesale” market or service refers to a program or service operated by a wholesale market operator. In each case, the applicable DR M&V methods are the methods the operator would use to measure performance of the DR provider. A retail program may be offered as an aggregate DR

[126] Goldberg, Miriam L, and G. Kennedy Agnew. Measurement and Verification for Demand Response (2013).

resource in the wholesale market. Different M&V methods may be used for retail settlement than for wholesale settlement, or for determination of demand reduction quantities for individuals than for aggregates. Direct Load Control (DLC) is not ordinarily offered by wholesale markets. Wholesale Direct Load Control in the table refers to aggregated DLC participating as a DR resource in a wholesale market. While NAESB Baseline Type 1 could in principle be applied to individual DLC end users, this practice is neither common nor recommended for retail settlement.

As indicated in the table, guidance in this document focuses primarily on specification of baseline methods, and on program-level impact estimation, We turn first to methods for settlement, which are primarily baseline methods.

2.8.4.5.1 Firm Load

Demand response programs that require participants to reduce load to a pre-specified, individually negotiated “firm” level during the event window, upon notification from the program administrator are effectively using the Maximum Base Load performance evaluation methodology. For many of these programs, M&V for settlement with the participating load is a straightforward observation of how much the load exceeded the firm level. Typically, this determination is based on the maximum metered load during the event window.

2.8.4.5.2 Reduction From Baseline

Many DR programs require participants to reduce load relative to a baseline during a performance window after notification by the program administrator. These DR programs reward participants according to the amount of their demand reductions during that window. These programs include many wholesale demand response programs, and retail programs, including Peak Time Rebate programs.

- For a participant that is an individual end user with interval metered load data, the baseline is calculated from the participant’s individual interval load data and settlement is usually based on the magnitude of the reduction. This is an application of the NAESB Baseline Type I method.
- For a demand response program that permits the aggregation of individually metered end users, an aggregate baseline may be calculated from the aggregate of the individual end users’ interval load data and compared with the aggregate observed load to determine the demand reduction. Alternatively, the aggregate demand reduction may be calculated as the sum of individual end user reductions, each calculated from its own baseline and own actual load. These are also applications of the NAESB Baseline Type I method.
- For a participant that is an aggregate of individual end users who are not all on interval meters, interval metering may be required for a statistical sample of the end users. The baseline is calculated from the interval load data for the sample. This is an application of the NAESB Baseline Type II method.
- For short term demand reductions, such as ancillary services, NAESB Meter Before/Meter After method may be used, and may be used in conjunction with another performance evaluation methodology to ensure the best estimate of the response and to mitigate gaming opportunities. The method can be used directly when the end user(s) all have individual interval metered load. Although not in widespread use at this time, it is possible that for an aggregation of end users who do not have interval metered load, Meter Before/Meter After can be applied to the aggregate load estimates from a statistical sample of end users. The use of data from the sample makes this approach an application of the NAESB Baseline Type II method in combination with Meter Before/Meter After.

2.8.4.5.3 Behind-the-Meter Generation

If the use of behind-the-meter generation is permitted in the demand response program, specific performance evaluation methodologies may apply to the output of the behind-the-meter generation during a demand response event or schedule. The applicable NAESB DR M&V method is Metering Generator Output. However, depending on how the participant uses the generator absent an event, a baseline calculation may still be needed. The same performance evaluation methodologies that are used for load participating as a resource may be applied to behind-the-meter generation. The value contributed to the program is measured as the difference between the metered generator output and the baseline generation for the event window. For wholesale demand response, measuring only the metered generation does not capture the impact of the total demand response resource’s load on the wholesale power grid. As a result, Metering Generator Output may be used in combination with another performance evaluation methodology when the demand response resource reduces load in addition to its behind-the-meter generation. Or metering at the retail delivery point may be used in lieu of separate metering of the behind-the-meter generator.

2.8.4.5.4 Direct Load Control (DLC)

Direct load control (DLC) programs allow the program operator to control customers’ equipment directly via communicating technology that signals equipment to turn off and then releases the control at the end of the event window. Initially, control devices were radio-signaled switches that turned equipment off entirely or limited how much the equipment could run in each hour. Most commonly controlled equipment types were residential central air conditioners, water heaters, pool pumps, or heat pumps. More advanced control equipment includes re-setting thermostats rather than restricting equipment duty cycle, and two-way communication to allow customers to over-ride control and programs to record customer control status.

Most DLC programs do not pay individual participants for their individual amounts of load reduction. Rather, as noted above, payment is typically some type of fixed participation credit per season, event, or event hour. As a result, DLC programs may not require measurement of reduction amounts as a basis for settlement between the retail program and the end-use participant. However, to determine the amount of credit to provide or to determine the benefit of the program, an estimate of the aggregate load reduction is needed, and this can be determined using a baseline.

If the total DLC program reduction is offered into a wholesale market as a demand response resource, a method for determining the reduction quantity during each event is necessary for settlement of the program with the wholesale market. Currently, DLC performance in wholesale energy markets is measured using a variety of methods, discussed in Section 2.8.5 *M&V Methods for Settlement*. Some of these methods can broadly be interpreted as applications of Baseline 1 (for customers who all have interval metering data) or Baseline 2 (when a sample of customers is metered).

2.8.5 M&V METHODS FOR SETTLEMENT

2.8.5.1 Fundamental Method Design Concepts

Designing a performance evaluation methodology for demand response program settlement starts with basic criteria:

- Accuracy – the method should provide an accurate estimate of the load so that demand response resources are credited only for load reductions associated with the event and baseline manipulation is minimized.

- Flexibility – the method should provide an accurate estimate of the load for all types of demand response resources that are expected and take into consideration extraordinary circumstances such as excessively high load on event days and exclusions that may reduce the accuracy of the estimate.
- Simplicity/Comprehensibility – the method should be able to be conveyed in straightforward language so that the requirements and calculations are readily understood
- Reproducibility – the performance evaluation calculation should be reproducible by the demand response resource, aggregator and program impact evaluator

The criteria outlined in the *NAESB Business Practice Standards* were developed to provide the structure for designing performance evaluation methodologies that support these fundamental criteria. The performance evaluation methodology used for settlement of the demand response program is vital to the success of any demand response program; being able to estimate the available reduction capability and making payment for the amount of reduction at the time of the event are key aspects of demand response programs.

As illustrated in Figure 2-18, DR M&V methods and results affect and are affected by many aspects of program planning, design, and operations. The M&V method specification for settlement, program structure and rules, and cost-effectiveness analysis all need to be considered jointly as part of program design.

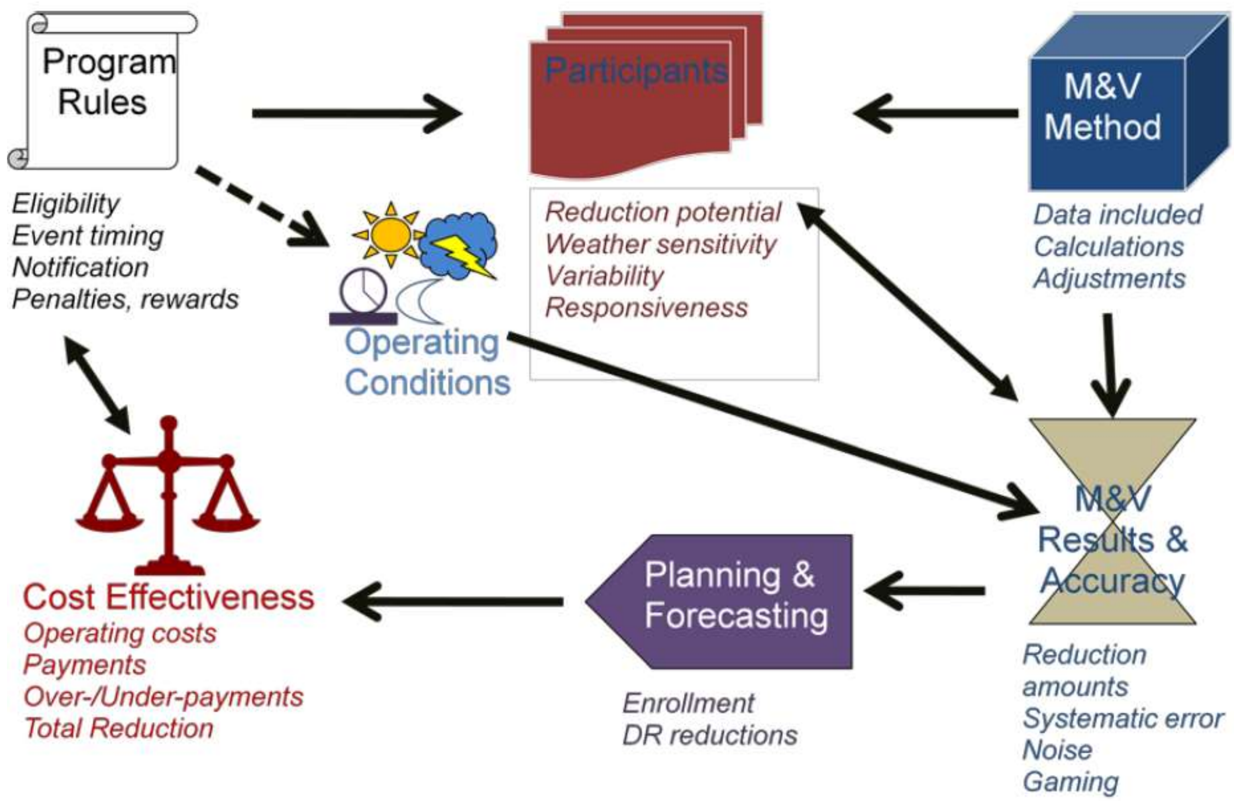

Figure 2-18 Methods and Results Affect and are Affected by Program Planning, Design, and Operations[127]

[127] *Ibid.*

Program rules, including measurement methods, payments, and penalties based on those measurements, affect the types of participants that will be interested in joining and staying in the program. Program rules also specify the conditions under which events are called, which can affect the results of M&V. M&V results and the accuracy of those results depend on the operating conditions as well as on the participant characteristics and M&V methods themselves. The M&V results may be incorporated into planning and forecasting, as well as the assessment of cost-effectiveness. Cost-effectiveness is the assessment of whether or not the benefits of the program outweigh its costs. Inaccurate M&V can result in over- or under-paying program participants and affect the level of program costs, program participation (i.e., over-paying will likely attract participation, and under-paying may reduce participation), and benefits computation. Over-estimated savings may result in over-stated benefits of avoided generation costs, which also reduces the benefit/cost ratio.

M&V method specification is an iterative process, as is all program design. After the initial design and implementation, modifications are suggested based on experience. Participant enrollment levels and behavior change in response to those program changes. The program rules and measurement methods must be re-evaluated and potentially revised based on customer response to changes in program design. The remainder of this section addresses baseline method specification for settlement. This specification is a primary challenge for designing DR programs that settle based on measured reductions. We first review the elements of baseline estimation error, and general means of managing those errors. We then discuss how the characteristics of participating resources and program rules can affect DR M&V accuracy. For each set of issues discussed, we provide recommendations.

2.8.5.2 Load Characteristics That Affect DR M&V Choices and Accuracy

As described in Section 2.8.4 *NAESB Business Practice Standards*, baseline calculation methods are specified by the combination of the data selection rules (baseline window and exclusion rules), the calculation type, and the adjustments (adjustment window and baseline adjustment method).

Simple baseline calculations support transparency. A variety of simple baselines are in use, using as the calculation method a simple or rolling average of load in each hour over days in the baseline window, subject to exclusion rules. Often an additive or scalar adjustment to recent pre-event hours is also included. Examples of such methods are included in Volume 3, Appendix B: *Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments.*

Empirical studies of baseline accuracy for commercial and industrial customers have shown that many simple baseline methods of this type for individual loads can have acceptable accuracy for program operations under a wide variety of loads and conditions. These studies have also found that, as long as a symmetric day-of adjustment is included, regression-based methods are no more accurate than these simpler averages. Additive adjustments are generally preferred to scalar adjustments because the resulting baseline can become volatile under a scalar adjustment.

For residential customers, however, simple baselines based on averages of recent eligible days have been found to have substantial biases for individual customers and, to a lesser extent, for program-level aggregates.[128] These biases are somewhat mitigated but are still substantial when day-of adjustments are used. While there

[128] See Oklahoma Corporation Commission Staff Report, Assessment of a Peak Time Rebate Pilot by Oklahoma Gas & Electric Company. Prepared by Dr. Stephen S. George, November 2, 2012.

are potentially ways to improve on these baselines, effective alternatives with much lower errors include the use of unit estimates based on prior evaluation work that incorporates more complete weather regression modeling, and the use of experimental design. Use of experimental design is discussed later in Section 2.8.4 *NAESB Business Practice Standards* and further in Section 2.8.5 *M&V Methods for Settlement*.

The types of loads participating in the DR program affect the types of baselines that can be effective, and the issues that need to be addressed in designing the program rules and baseline methods. Issues and methods associated with different load characteristics are discussed in what follows.

2.8.5.2.1 Business or Customer type

Business or customer type affects baseline accuracy primarily through its operational characteristics. Thus, if baseline methods are to be assigned based on customer type, this assignment is most effective if it is based on observable load characteristics, rather than a reported business category. For example, as noted, an industrial customer might have very consistent, non-weather-sensitive load patterns, weather-sensitive but otherwise consistent patterns, or highly variable patterns. Different methods will be most effective for these different customer types.

There are, however, broad differences between customer classes that relate to baseline method accuracy. Air conditioning tends to be a larger fraction of summer load for residential customers than for commercial customers, and many industrial customers have minimal weather sensitivity. Residential customers also use air conditioning more variably. Both these factors can make baseline accuracy more of a challenge in the residential sector compared to larger customers, for programs directed to summer peak use.

Recommendation: Business or Customer Type – If baseline methods are to be assigned based on customer type, this assignment is most effective if it is based on observable load characteristics and broad revenue class, rather than on a reported business category or customer segment. Key qualities that can be determined from customer load data include:

- Weather sensitivity
- Seasonality unrelated to weather
- Variability unrelated to season or weather.

2.8.5.2.2 Weather sensitivity

Residential and small commercial customers tend to have more weather sensitivity than large industrial loads. However, some large industrial facilities do include substantial weather sensitivity.

For weather-sensitive loads, it is particularly important to have days in the baseline calculation from the same season and with similar weather. In particular, as discussed above if events are called or bids clear on all hot (or cold) days, the accuracy of almost any baseline method is likely to be poor for weather-sensitive loads. Baselines for moderately weather sensitive loads work best when they include symmetric adjustments that reflect the weather of the event day. Without a day-of-event adjustment, reductions on very hot (or very cold) days can be substantially understated. This understatement occurs even if recent days are used and only higher-load days are included in the baseline computation.

Day-of-event adjustments will tend to over-state reductions for customers who pre-cool/heat in response to notification or in anticipation of a likely event. Customer-specific symmetric adjustments tend to understate reductions for customers who cancel work shifts before an event in response to notification. For this reason, it is recommended that adjustments rely on observed load in a time interval prior to the time of notification, or else use system or weather characteristics rather than the participants' pre-event load.

A common type of baseline is a simple average for each hour, taking the highest-load subset of X days in the baseline window of Y days. This "High X of Y" approach selects for days that are more like a peak day when events may be more likely. For weather-sensitive loads, however, this type of baseline still tends to understate baselines and corresponding load reductions on extreme hot days. On the other hand, "High X of Y" baselines will tend to be overstated on event days that are mild compared to recent days.

The inclusion of a day-of-event additive adjustment will substantially correct the understatement on peak days and the overstatement on mild days, though the load at the peak hours will still tend to be somewhat under- and over-stated in these respective cases.

Day-of-event adjustments do have some limitations (discussed later in this section, in Shift cancellation and other operational response to event notification or anticipation). Weather-based adjustments reflecting the load's historical relation to weather have been implemented successfully and provide an alternative for these scenarios (PJM weather sensitive adjustment method is discussed later in this section in Notification Rules and day-of-event adjustments, and in Volume 3 Appendix B: *Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments).* For residential customers with substantial weather sensitivity, baselines based on averages of recent days have been found to perform poorly, even with day-of-event adjustments. To calculate program-level reductions for programs with large numbers of homogenous customers, effective alternatives with higher accuracy are experimental design, or use of unit savings calculations determined from prior studies using regression analysis.

Recommendation: Weather-Sensitive Load – To reduce biases for moderately weather-sensitive C&I loads, include a symmetric day-of-event adjustment. Where anticipatory load changes are considered to be likely for many participants, a weather-based adjustment not affected by the customer's event-day load in pre-event hours should be considered. For program-level reductions for programs with large numbers of homogenous customers, use either unit savings calculations determined from prior studies using regression analysis, or experimental design.

2.8.5.2.3 Seasonality

Some loads have seasonal variations in operating patterns unrelated to weather. For such loads, baseline calculations that depend explicitly on weather variables, such as degree-day regressions or the PJM THI adjustment method, could create distortions. However, it is important to ensure that the data used in the baseline calculation are from the season of the event day.

Recommendation: Seasonal Non-Weather-Sensitive Load – To reduce biases for seasonal, non-weather-sensitive loads, include a symmetric day-of-event adjustment that is not explicitly related to weather terms.

2.8.5.2.4 Operational Variability—Highly Variable Loads

Some loads are very consistent for a given day, hour, and season, or can be well predicted using weather variables. Other loads are highly variable in ways that are not readily described by calendar and weather factors.

Loads that are highly variable apart from systematic weather response are a challenge for any performance evaluation methodology. For such assets, general customer baseline methods tend to produce demand reduction estimates with limited relationship to actual DR actions. The resulting disconnect between actions taken and payments to the participant can result in participant dissatisfaction, as well as detracting from market efficiency. If there are no penalties to the participant for under-performance, the highly variable asset is likely to stay in the program and receive erratic payments, without necessarily providing value to the market.

If a DR program is open to customers with highly variable loads, one strategy is to include a non-performance penalty to discourage customers who are unlikely to have a meaningful baseline from participating. Other strategies have been the subject of informal discussions by practitioners, but do not necessarily have any experience as of yet.

One potential strategy is to allow a procedure for customized baselines, to shift more of the prediction burden to the participant. For example, a customer may know what factors affect its load variations and may be able to provide operational data that allow a more meaningful baseline to be constructed. The customer would then be required to submit its planned levels of these operating conditions prior to bid submittal or the event notification. A simple example is that a plant with frequent, irregular shutdown periods might be required to provide advance notice of a pending shutdown and would be penalized for shutting down without prior notice if there is no DR event called.

Alternatively, the customer would be required to offer its own load prediction. If the participant is providing predictions of operations or load that will be the basis for calculating a baseline for settlement, the participant must also face a penalty if actual operations or load depart substantially from the prediction if a load reduction is not called. This approach is not currently in use, and details remain to be developed.

Another strategy is to establish formal criteria for measuring the predictability of a participant's load. Assets whose load does not meet the predictability criteria either would not be allowed to participate or would have their calculated reductions de-rated. A variant of this approach would be to count load reductions only if they are beyond an uncertainty band for the baseline.

Highly variable loads are inherently problematic for baselines intended to represent the load absent the DR event. In terms of program operations and settlement with the participant, such loads may be better engaged in other DR strategies, such as critical peak pricing or a firm load requirement program. Even if baselines are not needed to operate those other types of DR, impact estimation of DR performance from highly variable loads remains a challenge for all program types.

Many program operators must accept any eligible customer, and do not actively target, encourage, or discourage particular participants. For those operators, the only means of restricting or directing customers is through meaningful and defensible program rules.

Recommendation: Highly Variable Loads – For resources with highly variable loads, to ensure that incentives payments are meaningfully aligned with demand reduction actions taken, the following strategies may be considered:

- Establish a "predictability" requirement for program eligibility.
- Allow a customized baseline that uses additional operational information supplied by the participant.
- Require the participant to provide its own baseline prior to notification and penalize large departures from the participant's "scheduled" load on non-event days.

- If allowed, encourage the customer to participate in other types of DR programs that do not require calculation of demand reduction for program settlement.

2.8.5.2.5 Presence of Facilitating Technology

It is generally recognized that facilitating technology that allows customers to respond automatically to an event signal increases the responsiveness of participating customers. Automating technology also makes participation more attractive to customers. To a certain extent, facilitating technology can also improve the quality of M&V. A customer with effective control systems in place will tend to have more consistent operations during non-event periods, and more consistent response to events.

The control systems also may offer the opportunity to record additional operating parameters that can be useful in a more comprehensive impact estimation, or for other aspects of settlement not associated with baseline calculations. At a minimum, the program operator will typically have data on when control signals were sent. If the control signal technology is two-way, the operator may also have data on signal receipt and over-rides, if that is an option. Payments to customers can then be adjusted for failed signal receipt or over-rides. For example, some direct load control programs using two-way communicating thermostats allow customers to over-ride the thermostat re-set signal, and the customer pays a penalty or gives up an incentive payment for doing so. As described in Section 2.8.5, this system information on signal receipt and over-ride can be used for impact estimation, and for settlement based on *ex ante* unit savings and the number of units.

Recommendation: Facilitating Technology – For load control programs settled in the wholesale market based on the number of units controlled, information from the control system on control over-ride, success, or magnitude should be used as an input to the settlement calculation.

2.8.5.2.6 Shift Cancellation and Other Operational Response to Event Notification or Anticipation

As discussed above related to notification and adjustment timing, different types of customers have different inclinations to modify their load in preparation for or anticipation of a DR event. Participants who have to deal with shift scheduling will have different pre-event behavior from those who can turn major loads on and off on short notice. For customers with substantial heating or cooling of the premise or energy storage capability, pre-heating or pre-cooling is a consideration for baseline accuracy.

Some plants want to be able to respond to a DR notice by canceling a shift that is scheduled to start well before the event window. If the adjustment window would include part of the cancelled shift, the plant's baseline will be reduced by the shift cancellation. For this reason, it is recommended that participant-specific adjustments are based on pre-notification periods. For demand response resources that participate through offers to the market, consider allowing participants to specify a notification/start up time as part of their offer.

A plant with stable operating patterns and no weather sensitivity is likely to be better represented by a baseline with no day-of-event adjustment. Using the unadjusted baseline would allow the plant to cancel shifts before the event window without a negative effect on its calculated reduction.

Long-term shutdowns may affect the baselines of DR resources in programs where historical data from a prior period, such as the same season in the prior year, is used in a baseline calculation. Establishing procedures for reporting such planned shutdowns in advance can reduce opportunities for a baseline to be overstated.

2.8.5.2.7 DR Resources Providing Load Reduction Every Day

In principle, any DR resource with a capacity obligation must be available to provide demand reduction during all times covered by its obligation. Otherwise, demand response used as a capacity resource may not be able to displace the need for generation capacity (i.e., additional generation may need to be acquired to cover the hours that demand response resources were unavailable). Likewise, entities offering demand resources typically want to minimize restrictions on their opportunity to sell this service.

Some demand response resources are indeed in a position to provide demand reductions consistently every day. For example, a customer with behind-the-meter generation potentially could use its own generation, within the constraint of environmental permitting rules, to reduce load taken from the market on as many days as required by DR calls, but otherwise use its own generation only in emergencies. Even without onsite generation, a facility might have the ability to shift loads such that it could go to a lower level of operation during any period called, on any number of successive days, but would stay at a higher operational level if not called.

Meaningful measurement of load reduction requires observation of “non-dispatched” operating conditions. A resource that is in reduction mode on a continual or daily basis no longer has a “no-dispatch” state of operation against which the reduction can be measured. However, setting explicit rules to limit how frequently a resource may offer reductions is at odds with the principle of resources being available across all times covered by the DR program.

To address this issue, ISO NE has established rules that limit the number of successive days on which an entity can participate as a demand resource before its baseline must be refreshed. Baseline refreshment means inclusion in the baseline calculation of meter data from a present operating day, even if the operating day included a dispatched load reduction—in this case, meaning that the resource was instructed to reduce load as a result of its demand reduction bid clearing in the energy market. The extent to which this rule is sufficient or excessive and its applicability to other systems and services are open empirical questions.

Further exploration is needed of mechanisms for ensuring that adequate “non-dispatch” days are available for baselines, and to assess how many days are “adequate.” Such studies can lead to guidance on the types of mechanisms to use and how to specify them in detail based on program experience.

2.8.5.3 Program Design Features Affecting M&V Choice and Accuracy

As described in Section 2.8.4 *NAESB Business Practice Standards*, performance evaluation methods using Baselines are specified by the combination of the data selection rules (baseline window and exclusion rules), the calculation type, and the adjustments (adjustment window and baseline adjustment method). All of these specifications are part of the program design. Other program rules affect how frequently and under what conditions events can occur, or the frequency that a demand reduction bid from a particular asset can clear in a market that incorporates DR in its energy market. The combination of these program rules and baseline specification, along with the characteristics of the participating loads discussed above, affect the baseline accuracy. Program design elements are discussed below in terms of their interaction with baseline rules and accuracy.

2.8.5.3.1 Rules to Ensure “Comparable” Days in Baseline Calculations

The baseline window is specified to select days that are in some sense similar to the event day, such as recent business days. Exclusions are sometimes applied to eliminate anomalously high or low load days. Typically, event days are also excluded from baseline calculations, since the baseline is intended to represent a participant's consumption absent the event. Depending on the program rules and operating practices, these selection approaches can lead to a shortage of similar days in the baseline calculation, as described further below.

(i) Challenges if DR is Dispatched on Every Extreme Day

A common challenge is that DR events are often called on system peak days, which tend to be particularly hot summer days or cold winter days. The weather on recent non-event days will typically not be as extreme as on event days. If dispatchable events are called, or a particular bidding asset clears, on all of the most extreme weather days, it is difficult for any baseline methodology to provide accurate baselines for weather-sensitive loads for those days. This situation is a problem for impact estimation as well as for settlement baselines.

Baltimore Gas & Electric (BGE) addresses this issue by including weekends in the baseline calculation for a residential Peak Time Rebate (PTR) rate that has events only on weekdays, to ensure inclusion of hot days for each customer. An alternative approach, if program operators have discretion on when to call an event, is to operate the program in a way that ensures some event days and some non-event days for extreme weather conditions, as well as for mild conditions. For homogeneous customer groups, experimental design methods discussed in Section 2.8.6 *Impact Estimation* can provide this structure.

As described earlier in Section 2.8.5.2.2 *Weather sensitivity*, baseline methods based on averages of recent days, even with day-of-event adjustments, will tend to understate baselines on extreme weather days, and overstate on mild days, for highly weather sensitive loads. For weather-sensitive loads where this type of baseline is used, program rules that result in event days on a mix of extreme and mild weather days tend to produce a mix of over- and under-stated load reduction estimates. This mixing does not improve the accuracy of load or financial settlement for any single day but can improve the overall accuracy over a season. Of course, how over-, and under-stated reductions translate into net financial errors depends on the prices that apply to the different days.

If extreme weather days occur in sequential clusters, leaving one or more of the days in the cluster as a non-event day can partially improve the baseline accuracy for the event days that are called.

Recommendation: Program Operation to Reduce Baseline Error for Weather-Sensitive Loads – To improve the overall accuracy of settlement for weather-sensitive loads, if the baseline method is an average of recent days with possible exclusions and day-of-event adjustments, program dispatch rules that allow the following can be considered.

- Ensure that events are likely to be called on a mix of extreme and mild weather days.
- If extreme weather days are projected over several days in a row, leave one or more of these days as a non-event day.
- Even if there are no strings of sequential extreme days, ensure that some extreme days are not called as event days, for eventual impact evaluation.
- For residential programs, include weekend days in the baseline calculation even if they are not program-eligible days.
- For all but the last of these, a trade-off that must be recognized is that these practices to improve baseline accuracy would come at the cost of restricting the use of the DR resource.

(ii) Challenges from too few Recent Non-Event Days -- Static Baselines

For loads that vary seasonally, whether or not they are strongly weather sensitive, a related problem is frequent DR events. In demand response programs based on bids submitted by the demand response provider, some program rules may make it possible to bid in such a way so that events are called on every program-eligible day for several months. When event days are excluded from baseline calculations, as is commonly done, the result is a baseline frozen at the point before the string of DR event days began. In this case, there may be too few recent non-event days to provide the basis for an accurate baseline.

This problem will be partly ameliorated by use of a symmetric day-of-event adjustment, which roughly aligns the load level to conditions of the event day prior to the event. Day-of-event adjustments do not, however, address the changes in shape of the baseline over time. As a result, even with an adjustment, bias can increase as the source of baseline data become more distant from the event.

The frozen baseline phenomenon arises with the combination of:

- DR assets clearing every day in a bidding program
- Event days excluded from the baseline calculation
- Weather sensitive DR assets

In an example provided by ISO NE, several DR assets showed a pattern of bidding into the market every day at a price point that virtually assured they would be cleared, starting in the first cool period in the fall and continuing through the winter. Because these assets cleared every day, and prior event days were excluded from baseline calculations, baselines were fixed at their summer load levels. Thus, the assets received payments for the difference between summer and fall/winter load levels, even if they made no reduction in response to their bids clearing.

At the time, ISO-NE had an “asymmetric” day-of adjustment, meaning the adjustment was applied if it would increase the baseline, but not if it would decrease it. This adjustment method exacerbated the issue. Analysis of simulated load reductions and baseline calculations performed with program data explored the potential for frozen baselines. This analysis determined that applying a symmetric rather than asymmetric adjustment decreased the extent of the bias substantially but did not remove bias completely. The weather sensitive load shape underlying the static summer baselines remained quite different from the fall and winter load shapes and continued to show reduction according to the baseline calculation, where no true reduction had been made. The simulation data indicated that changing the baseline method to require a minimum number of program-eligible baseline days prior to the events would more effectively address this bias. Other alternative design criteria, such as changing the exclusion rules may provide a solution to reduce the likelihood of a static baseline when demand response is deployed frequently.

Thus, program rules can limit opportunities for static baselines by avoiding or limiting any of the bulleted conditions above. For example, ISO NE proposed incorporating cleared days (i.e., prior event days) to address baseline bias resulting from clearing every day. In this case, the main objective was to address the baseline bias.

Recommendations: Limiting Static Baseline Opportunities – To limit opportunities for “static baselines,” the following approaches can be considered.

- In programs where other program rules and requirements allow, and where event days will be excluded from baseline calculations, limit how frequently a given asset is allowed to clear or to have events.

- Incorporate event days or recent non-eligible days in the baseline calculation for assets that have too few recent non-event days in their baseline window. This should only be used in extreme situations, as doing so may increase the bias of the baseline calculation, reducing its accuracy and further understating the estimate of the load.
- For programs that have the flexibility to target particular types of customers, target loads with minimal weather sensitivity or other seasonality. This approach is not practical for all programs, but for large, non-seasonal industrial facilities, the static baseline phenomenon is unlikely to be a problem.

To determine if a static baseline may be an issue for program participants, model the proposed baseline calculation under extreme scheduling conditions to test its resilience to frequent scheduling. If a persistent bias develops under these conditions, one of the solutions listed above may be necessary to avoid paying for non-existent load reduction.

2.8.5.3.2 Notification Rules and day-of-event adjustments

Day-of-event adjustments are often included in baseline calculations to align the baseline calculated from recent non-event days with the conditions of the event day to improve the estimate of the “but-for” load level. The typical adjustment shifts or scales the baseline by a fixed amount so that it matches the actual load during a period before the event start (the adjustment window). This adjustment can help correct for load changes due to weather, as well as for variable operations.

In simulation studies of loads that are not participating in a DR program, symmetric day-of adjustments have been shown to improve the accuracy of a wide range of baseline calculations, including those that use explicit weather models, for a wide range of load types. However, for an asset that is in a DR program, there is the possibility that the load during the adjustment window will itself be affected by the event or the expectation of an event. The extent and nature of these effects is difficult to measure, but conceptually depends on the timing of the notification along with the specification of the adjustment window and method.

Event effects during the adjustment window can occur in a number of ways including the following:

- Preparatory increase in response to notification: A building is pre-cooled to a cooler than usual level from the time of event notification up to just before the event. This is a legitimate, reasonable response that makes program participation more viable for the building. However, if the adjustment window includes hours between notification and the event, the baseline will be inflated.
- Preparatory decrease in response to notification: A plant cancels a shift upon notification of an event. Facility load drops prior to the event start. If the adjustment window includes hours between notification and the event, the baseline will be substantially understated.
- Anticipatory increase prior to notification: A building is pre-cooled to a cooler than usual level beginning in the early morning whenever a very hot day is forecasted, which makes a DR event likely. As long as some hot days do not have DR events, the pre-cooling can be expected to occur in at least some of the non-event days used to calculate the baseline. The more routine the pre-cooling is, and the more the baseline window and exclusion rules select similarly hot days, the less bias there will be in the adjusted baseline.
- Anticipatory decrease prior to notification: A plant cancels a shift based on forecast conditions that suggest a likely event. Facility load drops prior to the event start. If the adjustment window includes hours between notification and the event and symmetric adjustment, the baseline will be substantially understated.

- Manipulative increase: A DR asset deliberately ramps up load during the adjustment window after event notification or based on its determination that an event is likely. The baseline is artificially inflated. This behavior may be difficult to distinguish from appropriate preparatory or anticipatory increases.

Setting the adjustment window to end prior to notification can limit opportunities for deliberate manipulation. On the other hand, the earlier the adjustment window, the less effective it may be in adjusting the baseline to estimate day-of load conditions.

Day-ahead notification is more attractive to participants who want more time to respond to events and is common in bidding programs. With day-ahead notification, any day-of-event adjustment is subject to preparatory effects, both legitimate and manipulative.

PJM's alternative weather sensitive adjustment16reflects the conditions of the event day without allowing pre-event responses to distort the baseline. This method uses a simple regression of load on whether to compare event-day weather conditions during the event window to the conditions during the baseline window at the same hours. The ratio of the regression-based load estimates for the two periods provides the adjustment. The approach has the advantage of adjusting to the event day weather conditions without requiring pre-event load to be informative. The disadvantage is that it adjusts only for weather and does not adjust for an asset's natural, non-distorting operations on the event day.

Some programs have used asymmetric adjustments, which apply the adjustment if it will increase the baseline but not if it would decrease the baseline. This practice avoids penalizing early shut-downs, but in general creates upward-biased baselines and can contribute to static baselines, discussed above.

Recommendations: Baseline adjustment methodologies by notification and load characteristics – To improve accuracy and reduce bias for almost any baseline method, use an additive, symmetric day-of-event adjustment. Table 2-32 summarizes recommended adjustment window and basis, based on the notification timing, and the likely accuracy problems remaining for different types of assets.

Table 2-32 Recommended Baseline Adjustment by Notification Timing and Load Characteristics[129]

If Notification Is-	For Load Characteristics		A Useful Adjustment Basis is-	Likely Accuracy Problems After Adjustment are-
	Non-Weather Variability	Weather-Sensitivity		
Same day	Low	Low	None or own load, 1-2hrs pre-notification	Minimal
	Low	High	Own load, 1-2 hrs. pre-notification or weather	Anticipatory pre-cooling can inflate baseline
	High	Low	Own load, 1-2 hrs. pre-notification	Underlying variable load

[129] *Ibid.*

	High	High	Own load, 1-2 hrs. pre-notification or weather	Anticipatory load shifting can inflate baseline, underlying variable load
Day Ahead	Low	Low	None	Minimal
	Low	High	System or weather, 1-2 hrs. pre-notification	Pre-cooling in response to notification/clearing inflates baseline; added variability compared to same-day notification, own-load adjustment
	High	Low	System or weather, 1-2 hrs. pre-notification	Underlaying variable load; added variability compared to same-day notification, own-load adjustment
	High	High	System or weather, 1-2 hrs. pre-notification	Pre-cooling in response to notification/clearing inflates baseline; added variability compared to same-day notification, own-load adjustment

(i) Concerns Related to Gaming Opportunities

A concern for any baseline method is that participants may manipulate their baselines to reap greater incentive payments. No baseline calculation method can eliminate the possibility of manipulation. However, such manipulation or "gaming" does not happen unless it is worth the trouble to the manipulator. The added energy costs and the operational inconvenience of changing load patterns simply to inflate a baseline have to be less than the expected excess payment. A DR aggregator attempting to adjust load for purposes of manipulating baselines needs the cooperation of its customers. While some end users, especially larger organizations, may find it worthwhile to follow a baseline manipulation strategy, this practice does not appear to be widespread in existing programs.

Bidding program participants typically want to know what baseline their reductions will be measured against prior to submitting a bid. This practice assures that even if the methods have biases, the participant has visibility to the results and can make an informed decision whether to offer a load reduction relative to that baseline. However, to reduce the incentive for selective bidding based on methodologically overstated baselines, the participant should not be able to submit a bid that is guaranteed to clear.

Recommendations: Limiting Gaming Opportunities – Elements that can reduce opportunities for baseline manipulation by participants include the following:

- Use a baseline calculation method that's fair on average on likely event days, absent any gaming.
- Ensure that baseline calculation data include recent "similar" days and are limited in how far back the "look-back" period can be so that data from another season cannot be used to overstate the baseline.
- Use rules that have the effect of limiting participants' ability to control or predict what days they will be called on to reduce.
- Investigate load and bidding patterns that seem perverse based on customer characteristics.
- Require advance notice of scheduled shut-downs.

2.8.5.4 Settlement Issues and Approaches For Particular Program Types

The settlement issues discussed above play out in different ways for particular program types. The following is a brief discussion of M&V issues for key types of DR programs. For each, we present a general discussion of the program type and outstanding issues to be addressed. We also identify some additional general issues requiring consideration.

2.8.5.4.1 Direct Load Control

As noted in Section 2.8.4.5 *Applying The NAESB M&V Terminology to Common Demand Response Program Concepts*, *Applying the NAESB M&V Terminology to Common Demand Response Program Concepts – DLC,* DLC programs typically pay incentives to participating customer based on participation only, and not based on a measurement of each customer's load reduction. However, DLC programs offered as DR resources in wholesale markets require a basis for measuring the reduction achieved by the program for a particular event. A variety of methods are currently in use for this purpose.

(i) *Ex ante* Unit Estimates and Current Participation

With this method of measuring DLC program load reduction, an *ex ante* estimate of savings per participant is multiplied by the number of successfully controlled participants. The unit savings estimate may come from engineering estimates at the start of a program, or from *ex post* program evaluation after some experience with the program. The average reduction per unit can be based on end-use metering, whole-premise metering, or other methods.

The *ex ante* estimates provide the average reduction per unit, typically by time of day or for the peak hour, and possibly also by temperature condition, by customer climate zone, or by equipment capacity. The number of successfully participating units begins with the enrollment level. This participant count should be adjusted by the rate of over-ride, if allowed by the program, and by signal success rates. These adjustment factors may be estimated from prior impact evaluation, or by event-specific information collected by the DLC program's control system, depending on the system capabilities.

Ex ante unit savings by geography, time of day, and weather condition based on analysis of multiple prior impact evaluations is the basis for PJM's "DLC method" for wholesale settlement. This method is used to settle DLC with PJM for participants who don't have interval metering in place as of the start of the season.

(ii) Firm Service Level

For retail customers who have interval meters, PJM uses another method, based on Firm Service level. The retail program operator determines the total Peak Load Contribution (PLC) of its DLC participants. This PLC serves as a Maximum Capacity Level. The operator commits to reduce the total load of the participants to a Firm Service Level during events, effectively the same as a Maximum Base Load. Performance relative to this committed reduction is calculated from the sum of the metered loads of the participants during the event.

(iii) General NAESB Baseline I or Baseline II

In principle, a Baseline method could be used that calculates a simple average of recent days, with adjustment to the event day, similar to many of the methods listed in Volume 3 Appendix B: *Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments*. This approach could be applied to individual customers with interval metering as a NAESB Baseline I method, or to a sample of customers who don't have interval metering, as a NAESB Baseline II method. However, application of these baseline methods to DLC programs for wholesale settlement does not appear to be in use currently and is not recommended. DLC programs that control air conditioning or heating involve loads and load impacts that are highly weather dependent. Simple baseline methods generally do not represent such loads as accurately as can the weather models used for the *ex ante* estimates.

(iv) Experimental Design

Experimental design, or the random assignment of eligible participants to treatment and control groups, has been used in recent years as an impact evaluation method. Operating a DR program using experimental design means that during each DR event, a randomly selected subset of participants is not dispatched, thereby serving as a control group. This approach can be useful for programs with large numbers of relatively homogeneous customers, primarily residential and small commercial.

For instance, some California direct load control programs have held back a random subset of participant households from each event activation. The event- period load for these non-activated but program participant households provide a statistically unbiased baseline for those households that were activated. This approach is not directly addressed in the *NAESB DR M&V Business Practice Standards*, though it could broadly be interpreted as an application of Baseline II method. Experimental design applications are discussed in (iv) *Experimental Design*.

2.8.5.4.2 Peak Time Rebate

Peak Time Rebate (PTR) is a retail rate or program that provides rebates to participants who reduce their use during an event window after notification that an event will be in effect has been issued. Retail settlement with participants requires a customer-specific baseline. The general baseline methods and issues described above apply in this context.

PTR often is available to smaller customers than have historically participated in DR programs (other than DLC). For these customers, reducing air conditioning use by raising summer thermostat settings can be a key part of their response strategy.

Common baseline methods used for PTR settlement are based on averages of metered consumption data from recent non-event days, with a baseline adjustment, or data exclusion rules to select hotter days. As discussed in Section 2.8.5 *M&V Methods for Settlement*, most of these methods tend to understate baselines on extreme hot days, resulting in penalties or lack of reward for customers who reduced energy consumption (and consequently made themselves uncomfortable) on very hot days. Understating the baseline and associated reduction in energy usage could be expected to lead to appreciable program dissatisfaction, though this response has not been seen in recent pilots.

Smaller load reductions that get lost in the noise can also result in underpayment. Further, customers with significant day-to-day variations in energy use could receive payments for naturally lower loads on days with event windows. In general, if the scale of reductions available to the customer is small compared to the customer’s overall variation in energy usage, establishing meaningful baselines for PTR will be challenging. This problem of small responses relative to the customer’s natural variability in energy usage is exacerbated if the PTR program is established as a default rate, with many non-engaged customers.

This issue was demonstrated in analysis of a proposed default residential PTR rate[130], with a baseline defined as the average of the highest 3 out of the most recent 10 eligible days, beginning 3 days before the event day, with no adjustment. The analysis of customer load on twelve key summer days showed that:

- 60% of customers would have received incentive payments based on the calculated baseline despite not reducing load at all during an event window. This would lead to incentive payments totaling $41 million each year to customers with no load reduction.
- Some customers who reduced their use (compared to a peak day with no event called) would receive no payment.

With this level of mismatch between actions and payments, this particular PTR program appears to provide little incentive to move this class of customers toward more efficient consumption behavior, in line with supply costs. Payments to customers who have not performed are costly to all ratepayers. Lack of payment to customers who have made reductions could dissuade customers from responding to future events.

The mismatch might be less severe with a different baseline method. However, even with a better baseline, there will still be payments to customers who took no action and non-payments to customers who did act for almost any PTR program.[131]

One reason PTR pilots have found high participant satisfaction despite baseline inaccuracies likely has to do with customer expectations.[132] Customers are not necessarily guaranteed a payment if they take certain actions but are paid if they beat their baselines. Moreover, baseline errors are not necessarily all in the same direction for a particular customer. In terms of the monthly bill, customers who tend to take actions during PTR events tend to see savings. Customers who respond minimally, if at all, to PTR events may or may not receive payments, and are not penalized.

Whether the baseline errors are too large for a particular program ultimately comes down to the question of whether the program is cost-effective with these baselines and the associated customer responses.

(i) Outstanding Issues for Peak Time Rebate

More study is needed to assess the accuracy of common baseline methods for the residential sector across a range of climate conditions. Future studies should include the implications for the monetary transfers and overall cost-effectiveness, under appropriate pricing assumptions.

More study is also needed on customer load and operating characteristics that make the customer a good PTR candidate. These characteristics include not only the ability and willingness to respond to events with observable demand reductions, but also predictable usage patterns outside of event days that will tend to result in stable and meaningful baselines. Understanding these characteristics can guide policies on whether and for what customer segments PTR should become a default rate.

[130] https://www.pge.com/regulation/RateDesignWindow2010/Testimony/PGE/2012/RateDesignWindow2010_Test_PGE_20120403_234258.pdf

[131] For a more detailed assessment of alternative baseline methods, see Oklahoma Corporation Commission Staff Report, Assessment of a Peak Time Rebate Pilot by Oklahoma Gas & Electric Company. Prepared by Dr. Stephen S. George, November 2, 2012.

[132] "BGE's Smart Energy Pricing Pilot," Cheryl Hindes, PLMA Panel, November 8, 2012.

Cost-effectiveness assessments are needed for PTR programs, based on impact estimations of load reductions actually achieved, as well as on observed customer acceptance rates from programs that have run for one or more seasons.

2.8.5.4.3 Ancillary Services

Ancillary Services is a relatively new product space for demand response, thus information on common performance evaluation methods for these new DR services is limited.

The Meter Before/Meter After performance evaluation methodology may prove to be a viable method for accurately estimating the response of DR resources under real-time dispatch conditions. Clearly Meter Before/Meter After requires demand resources with relatively flat load profiles during the time period of the dispatch. If a resource has periods of ramping up or down or general variability, the meter Before/Meter After approach can over- or under-estimate the actual level of load reduction even for the shorten period.

2.8.5.4.4 Programs Using New Control/Communication Technologies

New control and communication technologies that are being incorporated into demand response include:

- Remote control of equipment by customers;
- Automatic dispatch of demand reduction signals to customer equipment based on a price or command signal to the customer's meter, following a customer-specified response strategy;
- Communication that a control signal has been received or that specific equipment usage has been curtailed; and/or
- Real-time, two-way continuous communication with a system operator for dispatch of energy and/or ancillary service products.

The same general M&V methods can be applied for settlement (as well as for impact estimation) when these technologies are used as when they are not. However, these control and communication technologies also offer additional opportunities in the settlement context for verifying demand response and in the broader contexts of impact estimation for understanding demand response patterns.

The most useful information for M&V provided by this technology is the communication back to the program operator through new DR communication standards like OpenADR (Open Automated Demand Response). This information can be used for immediate verification of curtailment and identification of failed or over-ridden signals. As described in section B.8.6, this information can be used to determine DLC program accomplishment for wholesale settlement.

The operator may also receive more detailed information, such as the degrees of thermostat re-set, or particular pieces of equipment put into standby mode. This type of information is not currently being used for settlement but could be.

In the impact estimation and forecasting context, relating the equipment response information to empirical observations on load reductions over time allows more fine-grained forecasts of reductions for specific customers and for future customers. Comparing the equipment changed with the measured load reduction can also provide another level of verification of the load reduction measurement.

2.8.5.5 Means To Assess Settlement M&V Accuracy

As noted, there is no direct measurement of M&V accuracy. Only consumption can be metered directly, not reduction` in consumption. However, by using a form of load simulation it is possible to assess in general how well a particular baseline method represents what would have happened absent a DR event. The simulation calculates baselines according to the prescribed method for a set of customers and days when no DR event occurred. Comparisons to actual load during the DR event can then be made. Following are general steps for conducting such an assessment.

1. Obtain interval load data for a set of customers similar to those expected to be in the program. For an existing program, these customers might be actual participants on non-event days. For a prospective program, the customers who will be targeted, or a similar group of customers may be used. The more similar the customers used in this analysis are to the actual (likely or targeted) program participants, the more informative the analysis will be.
2. For days similar to days when DR events are likely to be called by the program, but when no DR event is affecting the study customers, use the designated baseline method to calculate the baseline for each customer and day. If events are likely to be called under a broad range of conditions, it is important to examine baseline performance for different conditions, including frequent successive deployments. If events are likely to be targeted to extreme weather days or system peak load days, it is important to examine baseline performance under these conditions.
3. For each customer in the study data set and each study day, calculate the following for one or more event hours:
    a. Calculated baseline using the baseline methodology;
    b. A simulated actual load reduction quantity assuming (for example) a 20% reduction from the actual load (actual load is known in the simulation exercise);
    c. The simulated actual event load with that simulated load reduction quantity;
    d. The simulated load reduction calculation using the baseline methodology: the difference between the calculated baseline and the simulated actual event load;
    e. The participant payment or penalty corresponding to the simulated actual load reduction quantity, applying the program payment/penalty rules to the actual reduction; and
    f. The participant payment or penalty corresponding to the simulated calculated actual load reduction quantity, applying the program payment/penalty rules to the calculated reduction using the baseline method.
4. Calculate the following accuracy metrics from the quantities in Step 3:
    a. Difference between (3a) the calculated baseline and actual load;
    b. Difference between (3d) the load reduction calculated from the baseline and the (3b) actual reduction. This metric translates (4a) the error in estimating load into (4b) the error in estimating the load reduction; and
    c. Difference between (3e) customer payments or penalties based on the reduction from the calculated baseline and (3f) what those payments or penalties would be if based on the actual reduction amount. This metric translates (4b) the error in estimating load reduction into (4c) the error in estimating the financial impacts.
5. Examine the distribution across customers and days for each of these accuracy metrics in terms of parameters such as the following:

a. Systematic errors or bias: average difference between the calculated value using the baseline method and the actual value.
b. Variability: what is the typical level of error for load, load reduction, and payment quantities?
c. What fraction of customers or what types of customers showed no positive load reduction using the calculated baseline?
d. What fraction of customers would produce a baseline load estimate that would require no actual reduction to achieve a positive payment?

Examples of such studies are discussed in Volume 3 Appendix C: *Prior work in DR M&V Methods*. An important point that emerges from studies of this type is that a modest error in estimating the load itself can become a much larger error in the calculated reduction. For example, for a 20% actual load reduction, a 10% error in the estimated load level is a 50% error in the calculated reduction. These errors in measuring reductions translate into misalignments between payments and actual load reductions. Even with these imperfect calculations of reductions, the DR program may still provide benefits to the program administrator and to the market.

Several simulation studies of baseline accuracy are described in Volume 3 Appendix B: *Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments*. Each of these studies examines both systematic errors and the level of random error or variability. However, there are a variety of ways to summarize the "typical" errors across multiple customers, days, and event conditions. Different studies have used different metrics in line with the general guidance above. Development of a standardized analysis and reporting approach would improve comparisons across such studies.

2.8.6 IMPACT ESTIMATION

Impact estimation at the program level is another instance of measurement and verification and plays an important role in ongoing program assessment and improvement. As indicated in Figure 2-16 above, M&V methods for settlement should be considered in the context of program planning, design, and operations. In this context, program-level impact evaluation is a key element in the ongoing cycle of program development.

Impact estimation broadly speaking means determination of program effects. For DR programs, these effects can include load reductions (or load increases) related to a particular event or set of events, energy savings (positive or negative), monetary effects, and other impacts. The effects may be determined at the program level or at any level of granularity. For purposes of this document, we consider impact estimation primarily for calculation of load reductions (positive or negative) for a program as a whole or for specific customer segments (e.g., geographic regions, low income customers, etc.).

The discussion here focuses on event-based programs. To a large extent, similar issues and methods apply to impact evaluation of alternative rate designs that are not event-based. However, issues specific to the evaluation of alternative rate designs are not examined in this report.

Impact evaluation in general measures load reduction achievement, not load reduction capability. The discussion below does not address capacity markets, though results of an impact evaluation could be used to assess capacity performance.

2.8.6.1 Impact Estimation Purposes and Contexts

Impact estimation is used in a variety of contexts and for a variety of purposes. The estimation can be described in terms of the following dimensions:

- Purpose: how will the reduction determination be used, and by whom?
- Perspective: retrospective (*ex post*) or prospective (*ex ante*).
- Level of customer aggregation: individual retail customer, entire program, aggregations of customers by the DR provider, or customer segments.
- Level of event aggregation: individual event, summary of events in various forms (overall averages, averages as a function of temperature, customer segment, location (etc.) in a projection table or formula).
- Timing of impact determination (e.g., day after event, end of season, etc.).

These dimensions are discussed below.

2.8.6.1.1 *Ex Post* Impact Estimation and *Ex ante* Impact Estimation

Ex post impact estimation determines demand reductions retrospectively. *Ex post* estimation for a program season or year is commonly used as part of regulatory or stakeholder due diligence to determine if a program performed as planned and may be the basis for payments to program operators.

Ex post estimation not only provides the retrospective scorecard of what did happen, but also is typically the foundation for developing *ex ante* impact estimates and for understanding how to make a program perform better going forward. Explicit projections of impacts under future conditions are *ex ante* impact estimates.

Ex ante impact estimation provides projected demand reduction estimates for future program periods and/or for specific event conditions (e.g., normal weather, extreme weather, etc.). These projections may be functions of enrollment levels, participant characteristics, or event conditions.

Ex ante estimates also are important for assessing the cost-effectiveness of programs. DR resources have option value – that is, they are designed to be used under extreme conditions (e.g., system emergencies, high priced periods, etc.). In any given year, such conditions may not occur frequently or be as extreme as the conditions for which the program was designed. As such, for any particular year, the average impacts per unit may understate the true value of the program. Cost-effectiveness analysis using the *ex post* impacts specific to any particular year thus has limited use.

For programs with relatively homogenous participants such as residential programs, *ex ante* methods typically consist of projected savings per participant, together with projected enrollment numbers. The projected savings per participant and enrollment is likely to vary by geography and potentially other characteristics. Savings per participant also typically varies by time of day and weather conditions.

Ex ante impact estimation can be used as the basis for retrospective settlement. In this case, application of an *ex ante* projection table or formula to observed conditions and actual enrollment provides an *ex post* impact determination. For programs that allow dispatch to be over-ridden, enrollment is adjusted by the fraction responding or projected to be responding.

For example, PJM uses the “DLC method” to settle with utilities operating Direct Load Control programs. Prior *ex post* impact evaluations from the PJM region were mined to determine *ex ante* savings per participating unit for each utility as a function of a temperature-humidity index. Under the PJM DLC method, *ex post* savings for settlement are calculated by multiplying this unit savings by the number of participants and adjusting for over-ride rates where applicable.

2.8.6.1.2 Individual and Aggregate Impacts

Impact estimation is typically not concerned with accuracy for individual customers so much as accuracy of aggregate estimates at the program or participant subgroup level. Even when individual customer baselines for settlement have noise and recognized biases, impact estimation for the program as a whole can demonstrate DR as a reliable, measurable resource.

Often impacts are determined not only for the program as a whole but also by participant segments defined by program options, geography, and other customer characteristics. The segment-level analysis can provide insight into conditions where greater reductions are achieved. In addition, segmentation provides a basis for more meaningful *ex ante* estimates as the mix of participating customers' changes.

2.8.6.1.3 Timing of Impact Determination

Comprehensive aggregate *ex post* and *ex ante* impacts may be determined after the end of each program year or season or less frequently. Seasonal impacts may be summarized in terms of the maximum, average, or total reduction over all events in the season. Future impacts, as noted, may be expressed as functions of customer characteristics and event conditions.

Many programs determine *ex post* impacts within a few days of each event. Some programs need immediate impact calculations for settlement with participants. Methods commonly used for settlement with program participants are the focus of Section 2.8.5 *M&V Methods for Settlement*. For both program and participant operations, day-ahead *ex ante* estimates are important. Program operators need to know how much of each resource is likely to be delivered in response to an event call. Program participants, both DR aggregators and individual customers, need to know what their own resources are likely to deliver to make bid decisions and other operational choices.

2.8.6.1.4 Summary of Impact Estimation Applications

Table 2-33 summarizes the ways that impact estimation is used, and the associated perspectives, aggregation, and timing. The *ex ante* perspective refers to *ex ante* estimates developed from *ex post* impact estimations.

Table 2-33 DR Impact Estimation Methods By End-Use Participant Type and Perspective[133]

Purpose	Perspective	User	Level of Customer Aggregation	Event Aggregation	Timing
Annual or Seasonal due diligence program measurement	Ex Post	Program operator, Regulator	Program or specified aggregated load	Summary over events	End of season
Settlement with individual end users	Ex Post	Program operator	Individual account	Individual event	Day(s) after event or monthly
Settlement with DR aggregator	Ex Post	Program operator	Aggregated load	Individual event	Day(s) after event or monthly

[133] *Ibid.*

Day-ahead or shorter operational planning	Ex Post	Program operator	All DR resources or targeted subset	Individual (possible) event	Day or hour(s) ahead
Daily Bidding and operators	Ex Post	Program participant (individual or aggregator)	Own resource	Individual (possible) event	Day or hour(s) ahead
Annual planning	Ex Post	Program operator	All DR resources	Rangers of potential events under various scenarios	Season ahead
Annual planning	Ex Post	Program participant (individual or aggregator)	Own resource(s)	Rangers of potential events under various scenarios	Season ahead up to long term planning horizon

2.8.6.2 Impact Estimation Methods

For DR programs settled based on calculated reductions, the *ex post* impact can be calculated as the simple sum of the demand reductions determined for each participant using the program's settlement methods. This method is used, for example, by the NYISO for its Emergency Demand Response Program. With this approach, there is no difference between the total settled amount and the program-level impact.

Some programs, however, conduct a program-level impact estimation that does not rely on the settlement method or settled quantities. *Ex post* program-level impact estimation is not subject to many of the constraints of participant settlement. These constraints include the need for simplicity, rapid results, reduction amounts for each participant and event, and timely feedback to customers for an effective behavioral change program.

More accurate program-level results can typically be obtained by using impact estimation methods that are not practical for settlement applications. These methods include:

- Individual or pooled regression analysis involving more complex models and data from a broader span of time than typically used in settlement calculations that may provide *ex ante* and *ex post* results from the same model;
- Day matching to identify one or more non-event days that are similar to each event day, usually from a full season of data;
- Incorporation of supplemental information about customers, such as survey data, end-use metering data, or program tracking data; and
- Experimental Design, treatment/control group analysis.

These methods are discussed below. This guidance document does not attempt to specify analytic forms in detail or to identify the preferred analytic approach. Rather, the advantages and disadvantages of general methods in different contexts are described.

2.8.6.2.1 Individual Regression Analysis

Individual regression analysis fits a regression model to an individual customer's load data for a season or year. A basic model describes load at each hour of the day (or perhaps the average for an event window) as a function of weather terms such as cooling degree-days. More elaborate models can allow the cooling degree-day base to be determined by the regression best fit, and might include calendar and day of week effects, lag terms reflecting temperature over multiple hours, and humidity. An example of a basic individual hourly load regression model is shown in the equation below:

$$L_{jdh} = \alpha_{jh} + \beta_{jh} C_d + \varepsilon_{jdh}$$

Where:

L_{jdh} = is the load of customer *j* at hour h of day d

C_d = is the cooling degree-days for the day

α_{jh} = is the base coefficient for each hour of the day, specific to customer *j*

β_{jh} = is the cooling coefficient for each hour of the day, specific to customer *j*

ε_{jdh} = is the residual error

Typically, the individual regression models are fit to loads on non-event days. The model is then applied with the conditions of each event day to provide an estimate of the customer's load that would have occurred on that day absent an event. The impact is calculated as the difference between the modeled and observed load for each hour of the event period. Post-event rebound (increased load to make up for foregone load during the event period) can also be calculated.

When load data are available for a sample of participating customers, the program-level results are estimated by sample expansion from the individual customer impacts. When load data are available for all participating customers, program-level results are the sum of the individual customer impacts.

The individual regression model can also include event-day terms and be fit across both event days and non-event days. In this case the event effect is the difference between the model applied to the event-day conditions with and without the event-day terms in effect. The second equation (below) provides a simple example. However, unless there are multiple event days spanning a wide range of the other terms in the model, including event-day terms in individual regressions will provide no more information than the average over event days of the modeled versus observed approach from the previous equation:

$$L_{jdh} = \alpha_{jh} + \beta_{jh} C_d + \delta_{jh} E_d + \varepsilon_{jdh}$$

Where:

L_{jdh} = the load of customer j at hour h of day d

C_d = the cooling degree-days for the day

α_{jh} = the base coefficient for each hour of the day, specific to customer j

β_{jh} = the cooling coefficient for each hour of the day, specific to customer j

E_{dh} = a 0,1 dummy variable indicating that an event occurred on day d

δ_{jh} = the event effect for hour h

ε_{jdh} = is the residual error

Advantages of the individual regression method are:

- Results are determined for each customer, which provides a basis for richer analysis, including looking at distributions of results rather than averages only. Individual customer results can also be related to other customer information.
- Meaningful results can more easily be developed for groups of customers whose load patterns are dissimilar, since each is modeled separately.
- Results can be aggregated into any segments that are subsequently determined to be of interest after the initial analysis is completed.
- Customers for which the basic regression structure is not a good description can be identified by model diagnostics and treated separately.
- Weather response terms such as the best degree-day base can be determined separately for each customer, leading to better and more meaningful overall fits.
- *Ex ante* results can be derived by fitting individual regressions to design or extreme temperature data and then aggregating the resulting estimates.
- Results can be analyzed to understand relative customer engagement in programs that promote behavioral changes.

On the other hand, model fits for an individual customer are subject to a higher level of estimation error than are the fits from a pooled model. Examination of distributions across customers' needs to consider that the spread of observed results reflects both the spread of individual responses and the estimation "noise" or random errors.

Moreover, if event-day effects are estimated for an individual customer, these individually estimated effects can often be lost in the noise—that is, not be statistically significant—even if across all customers there is an effect. The opposite can also occur, where statistically significant effects are found for large numbers of control group customers who had no event to respond to. That pattern indicates a systematic modeling error, which would affect a pooled model just as much as it would affect the average of individual models.

In general, if the same model structure is applied with individual fits and with a pooled fit, the coefficients of the pooled fit will be approximately the average coefficients of the individual fits. This equality will be strictly true if the individual and pooled fits all use the same degree-day base and other variables, the individual fits all have the observations in the same hours, and all observations have equal weights. In particular, any bias in the individual fits will be present for the pooled fit as well.

A disadvantage of the individual regression approach is that it does not take advantage of the power of a pooled regression approach.

2.8.6.2.2 Pooled Regression Analysis

Pooled regression analysis uses a similar model structure to the individual regression analysis but fits a single model across a large group of participants and hours. In this case, a single set of coefficients is used to describe all customers' average load pattern. With a pooled analysis, it is more common to include event-day terms in the regression model. With the larger pooled sample, terms that might not be well determined for an individual customer can be estimated. A simple example is illustrated below:

$$L_{jdh} = \mu_j + \tau_{dh} + \alpha_h \beta_h C_d + \delta_h E_d + \varepsilon_{jd}$$

Where:

L_{jdh} = the load of customer *j* at hour h of day d

τ_{dh} = an incremental fixed level for customer *j*

μ_j = fixed effect terms for affecting all customers for a particular day and hour (reducing the residual correlation for repeated observations at the same day and hour)

α_h = the base coefficient for each hour of the day, specific to customer *j*

β_h = the cooling coefficient for each hour of the day, specific to customer *j*

E_{dh} = a 0,1 dummy variable indicating that an event occurred on day d

δ_h = the event effect for hour h

ε_{jdh} = is the residual error

α_h, β_h and δ_h are not customer-specific.

Advantages of the pooled regression method are:

- The coefficients utilize information across all customers, so that effects that might be poorly estimated by each individual regression can be well determined.
- Segment level effects can be obtained by including segment indicators in the model, or by fitting the model separately by segment.
- Overall results are provided even if there are some customers for which the basic regression structure is not a good description.
- *Ex ante* estimates can be obtained directly from the event-day terms in the model.

Disadvantages of the pooled regression method include:

- Segments of interest need to be identified in the model development stage and cannot be easily estimated after the fact from the basic results.
- Weather response terms are estimated only in aggregate, which can reduce the model accuracy.
- The method works best when pooling is across a group of fairly similar customers, such as residential or small commercial.
- A pooled model approach has an added degree of complexity relative to the individual approach. Even with the inclusion of customer-specific intercepts (μj) and time-period terms (τdh) there will still tend to be serial correlations and patterns in the regression residuals (εjdh). If these correlations are not appropriately accounted for, the regression estimates can appear to be much more precise than they really are, especially if many thousands of customers are included in the regressions. That is, the calculated standard errors for the regression terms and associated savings estimates may be understated.

2.8.6.2.3 Match Days

Match day methods identify one or more non-event days that are similar to each event day, based on various criteria. Common bases for identifying match days for a given event day include:

- Similar temperature or temperature-humidity index;
- Similar system load; or
- Similar customer load at non-event hours for the individual customer.

For each participating customer, that customer's load on the match day (or average of the match days if there are multiple) serves as the baseline or reference load. Demand reductions are calculated as the difference between the (average) match day and event day load at each hour.

A key advantage of match day methods is their simplicity and transparency. In addition, for variable loads that are not well described by hourly or weather models, match day methods may be more accurate than regression models if the matching criteria include characteristics of the individual customer's load.

Disadvantages of match day methods include:

- For loads that can be reasonably well described in terms of hourly loads and weather patterns, regression methods will tend to be more accurate. Match days are limited to actual observed days, and averages of those days. Regression models, if properly specified, effectively interpolate between particular observed conditions, and extrapolate from them. (It's easy to construct examples of weather models that consistently understate load in extreme weather conditions. A matched day could provide a better estimate at those conditions than such a model. However, a better model that does not systematically understate load at the conditions of interest, possibly by using only data from more extreme conditions, in most cases will be more reliable than a single best-fit day. Any basis for selecting match days should, in principle, be possible to capture more systematically and comprehensively in a regression framework.)
- Match day methods do not provide a direct basis for producing *ex ante* estimates. If a regression will be used to extrapolate from the match-day results, it may make more sense to use a regression for the *ex post* results to begin with.
- Assessing the accuracy of a match-day estimate is more problematic than assessing the precision of a regression model. Testing for lack of fit or systematic bias is not as straightforward with a matching procedure as with an explicit model, and is not commonly included in match-day analysis. Measuring the precision or level of random variability of a match-day estimate is also not as clear-cut. It's possible to calculate a standard deviation across match-day estimates from multiple event days, but it's not clear to what extent this variability reflects differences in event-day conditions versus random variations on the particular event days versus particular conditions or random variation on the non-event days used for matching. If the analysis is done for a sample of customers rather than for the full population, variability across different match days does not reflect the sampling errors (that is, the differences that would be expected with the same methods if different random samples were selected). As a result, determining the true uncertainty of both *ex post* estimates and projections based on those estimates is challenging.

2.8.6.2.4 Experimental Design

For DLC as well as other mass market programs, comprehensive interval metering offers the opportunity to use experimental design for M&V. This approach can be used to determine program-level reductions for individual events. It has begun to be used for *ex post* impact estimation and offers substantial promise. As noted in Section 2.8.4 *NAESB Business Practice Standards,* direct use of experimental design has not yet been seen as a basis for market settlement, though *ex ante* estimates based on experimental design may be.

Experimental design is random assignment of customers into two groups, one of which is "treated" and the other remains as a "control" group. In the case of DLC, customers enrolled in the program are randomly assigned to subgroups, and during any dispatch event one or more of the randomly assigned groups is not dispatched while the remainder are. That capability depends in part on the program's control technology, and in part on the operational capacity of the program. Thus, an essential feature of this impact estimation method is that it must be built into the program operation.

The average demand reduction per participant is calculated as the difference between the averages for the groups that are dispatched and those which were not. An alternative calculation with this design is a difference of differences method. A baseline calculation or load model constructed for each participant, in both the dispatched and non-dispatched groups (treated and control groups, respectively). The impact is then calculated as the difference between the dispatched group's modeled and observed load, minus the corresponding difference for the control group. With this approach, the departure of the control group from its modeled load essentially provides an estimate of how the treatment group's actual load would have been higher or lower than its model, absent a DR event.

With customers who all have interval metering via Advanced Metering Infrastructure (AMI), this type of design and analysis has been used to determine impacts of large-scale residential and/or commercial direct load control programs at PG&E, SDG&E and across multiple utilities in Ontario Canada for the Ontario Power Authority's (OPA) PeakSaver Program. The approach has been used also with a sample of interval metered customers prior to the implementation of AMI, for SDG&E.

In many contexts, randomly assigning customers to different rates or different dispatch regimes is not possible. In these cases, comparison groups of customers identified as similar to the participants after the fact are sometimes used for impact estimation. However, without true random assignment there are always unknown underlying differences between participants and nonparticipants, and these differences can bias any estimate based on comparing the groups. The remainder of this discussion focuses on the use of randomized treatment-control experimental design. In such a design, customers originally in a common pool are randomly assigned to either the treated or comparison (control) group, with minimal subsequent opportunity for customers to opt in or out of their assigned group.

The randomized control experimental design is conceptually the gold standard of evaluation approaches but has been limited in its practical applications until recently. The practical limitations result from the fact that most full-scale program applications and regulatory contexts don't allow for random assignment of customers to participate in a program or not. A recent exception in the energy efficiency context is behavior-based programs offering information to large numbers of randomly selected residential customers. The experimental design of the program offering establishes the basis for measuring the effect of the information program.

Where feasible, experimental design has the potential to produce the most accurate results possible for estimating load reduction. The method is valuable because it virtually eliminates any systematic difference between treatment and control, providing an unbiased estimate, and with sufficiently large samples can provide very high precision.

Experimental design is effective for impact estimation of relatively homogeneous groups of customers, such as residential or small commercial, where several hundred or several thousand customers participate in a program. The method is less effective for evaluating smaller numbers of customers or large commercial or industrial customers, because the treatment-control differences will have too much random error to be reliable.

When most participants have interval metered data available, experimental design offers many advantages including the following:

- First, because the M&V is conducted separately for each event day, participants do not have to be assigned to treatment or control permanently. In fact, it is more appropriate to have the control group be a different, randomly selected set of participants for each event. This approach best assures that the treatment and control group are the same in all ways other than being dispatched on a particular day, including that they have otherwise equivalent program experience.
- Second, for a large scale program, large control samples can be used to provide highly accurate results without substantially reducing the total dispatched resource. When load control programs had to be evaluated using metering samples installed specifically for that purpose, samples on the order of a few hundred (depending on the level of granularity desired) were sufficient to provide adequate accuracy for the estimated reductions. A program with 50,000 customers enrolled could easily have a control sample of 1,000 customers for each event day to produce accurate estimates of program load reductions.
- Third, for *ex post* estimation or for settlement directly based on the metering sample, determining savings based on a randomly assigned treatment-control difference provides a highly accurate estimate of the reduction without requiring explicit weather modeling. If weather modeling is used, the difference of differences method ensures that any systematic bias in the modeling can be corrected by subtracting the difference between the modeled and actual load of the control group from the difference between the modeled and actual load of the control group of the dispatched group.
- Fourth, for *ex ante* estimation, observing large numbers of both dispatched and non-dispatched customers during each event provides a much more accurate basis for modeling event effects as functions of weather or other conditions. This type of modeling can be very challenging in particular if all participants are dispatched on the few hot days.
- Fifth, as an extension of the last point, with a random control group as the basis for settlement and evaluation, calling events on every hot day does not create a problem for M&V.
- Finally, the experimental design approach can allow good load reduction estimates to be developed for a wide range of conditions, while exposing any individual customer to a limited number of control events. This feature can allow the method to be used to define *ex ante* estimates for a range of operating parameters and weather conditions. Implementing this aspect of the approach requires close coordination with the program operation.

The best ways to produce *ex ante* estimates based on experimental design are still to be explored. The per-unit results from different event days can be averaged, or a simple temperature regression can be fit to the results.

A more complete approach could be to fit a pooled model across all customers and days. Having treated and control customers on each event day as well as having both event and non-event days for each customer strengthens this analysis. The pooled model could provide *ex ante* estimates per unit as a function of weather conditions.

This type of analysis is relatively straightforward to conduct with a sample of a few hundred or even several thousand participating customers but may be computationally challenging for a large residential program with universal hourly load data available. Possible ways of addressing that challenge include:

- Conduct the analysis using data from a large sample of participants, not all of them.
- Aggregate the load for groups of customers who had the same DR dispatch schedule. Conduct a pooled analysis on the groups.

(i) Use of Experimental Design

Experimental design utilizes established statistical methods to produce unbiased, highly accurate *ex post* impact estimates. Key outstanding issues for increased use of this approach include:

- Explore with program operators the challenges of and potential for dispatching the program following an experimental design protocol.
- Work with wholesale markets to establish protocols that will allow use of experimental design as a basis for settlement.
- Establish recommended strategies for developing *ex ante* estimates when *ex post* or settlement is based on experimental design.

2.8.6.2.5 Applications of End-Use Metering for DR Impact Estimation

Until the last few years, interval load data has not been available for most small customers. Impact estimation for residential DR programs such as DLC has typically relied on metering samples installed for this purpose. In areas without AMI, that will still be the case in the future.

Since DLC programs control a particular end use, impact estimation can be conducted by metering only the affected end use(s). Many DLC evaluations have taken this approach. Advantages of end-use metering include the following:

- A single end-use can typically be modeled more accurately than whole-premise data, resulting in better precision for the overall estimates for a given sample size.
- Equipment operating characteristics such as duty cycle and connected load can be identified, providing additional insight into event response patterns.
- Load curtailment can be observed directly if end-use metering data are collected at 1-minute intervals.

On the other hand, whole-premise metering captures other effects in the home that are not reflected in the end-use metering. For example, control of the air conditioner compressor could result in increased use of fans or even room air conditioners.

When interval load data are broadly available via AMI, investment in end-use metering for impact estimation becomes more difficult to justify. Moreover, the large numbers of metered customers available with AMI makes up for the reduced resolution for individual customers in an impact evaluation. However, even on a small sample basis, supplemental end-use metering can provide finer grained understanding of load response patterns and mechanisms. In particular, modeling duty cycle and connected load as functions of temperature provides a strong basis for projecting the effects of alternative air conditioner control strategies, as described below.

End-use metering data can be analyzed using the same types of modeling approaches as whole-premise data, including use of a randomized treatment/control methodology. This approach has been used for example in the evaluation of the SDG&E Smart Thermostat program.

For air conditioner DLC, end-use metering analysis can take more complete advantage of the physical relationships that drive air conditioning. One such approach fits 2 types of models to 15-minute or finer air conditioning metering data for each unit in a metering sample:

1. A model that estimates the connected load of the air conditioner, the kW draw when the unit is running, as a function of current outside temperature. This connected load is not constant but increases by 1 to 2 percent per degree Fahrenheit.
2. A model of duty cycle, or the fraction of each hour the unit runs, as a function of daily weather conditions. The duty cycle model uses a structural form that recognizes that the duty cycle must be between 0 and 100%.

Advantages of this analysis approach include:

- The analysis reveals detailed patterns of customer equipment use at different conditions.
- These patterns can be related to other customer characteristics.
- Projected reductions can be estimated by time of day and weather condition, at any level and strategy of duty cycle control, not just those observed in the evaluation. That is, this approach more accurately models the technical limits of AC units thus more effectively accounting for units reaching full cooling capacity at extreme temperatures.

2.8.6.2.6 Custom Engineering and Field Studies

For individual large loads, special studies can be conducted to assess load impacts. These studies would typically include a site visit to identify what loads are controlled, together with end-use metering or extraction of existing operating log data to document load at event and non-event conditions. Analysis to estimate the load that would have occurred absent an event is specific to the operations of the facility. While this approach is not common, it may be the only practical method for large loads with irregular operating patterns.

2.8.6.2.7 Composite studies

An approach that has been used for *ex ante* impact estimation in the PJM market is to consolidate the results of multiple end-use metering studies conducted for *ex post* impact evaluations. The consolidated metering analysis was used to develop *ex ante* estimates for DLC programs, for several utilities operating in that market. This approach can provide a more robust result than any single study.

2.8.6.3 Guidance Summary

Table 2-33 summarizes which impact estimation methods are likely to be most useful for different types of end-use customers, for *ex post* impact estimation and *ex ante* impact estimation. In any particular evaluation context, the methods that will be most effective will depend on a variety of factors, including specific evaluation goals, participant load characteristics, data availability, numbers of participating customers, and evaluation budget and timeframe.

Table 2-34 Usefulness of DR Impact Estimation Methods by End-Use Participant Type and Perspective

Impact Estimation Method	**Homogeneous Customer Group (Residential, Small Commercial/Industrial**		**Heterogeneous Customer Group, Each Customer with Low or Moderate Load Variability**		**Customers with Highly Variable Loads**	
	Ex Post	***Ex ante***	**Ex Post**	***Ex ante***	**Ex Post**	***Ex ante***

Individual Regression	Very useful	Useful with additional work	Useful	Useful with additional work	Possibly useful	Useful with additional work
Pooled Regression	Useful	Very useful	Not useful	Not useful	Not useful	Not useful
Match Day	Possibly useful	Possibly useful with additional work	Possibly useful	Possibly useful with additional work	Useful if match on customer condition	Useful if match on customer condition, with work
Experimental design simple difference	Very useful	Useful with additional work	Not useful	Not useful	Not useful	Not useful
Experimental design with modeling	Very useful	Very useful	Not useful	Not useful	Not useful	Not useful
End use metering with Duty Cycle Analysis	Very useful	Very useful	Potentially useful	Potentially useful	Potentially useful	Potentially useful
Customer engineering and site analysis	Not generally useful	Not generally useful	Potentially useful	Potentially useful	Potentially useful	Potentially useful
Composite Analysis	Potentially useful	Potentially useful	Not generally useful	Not generally useful	Not useful	Not useful

2.8.7 DR WHEN CONNECTED TO SYSTEM AGGREGATOR

Independent System Operators (ISO) and Regional Transmission Organizations (RTO) have strict protocols for electricity providers who are connected to their markets. At the time of authorship, Entergy New Orleans is not connected to and ISO/RTO but is geographically located in the Midcontinent Independent System Operator, Inc. (MISO) territory.

At the time of protocol authorship, Entergy New Orleans is a MISO member but does not participate in the MISO Demand Response market. However, should Entergy New Orleans decide to participate in the future, the following Business Practice Manual (BPM), developed by MISO, will provide background information, guidelines, business rules, and processes established by MISO for the operation and administration of MISO markets, provisions of transmission reliability services, and compliance with MISO settlements, billing, and accounting requirements.

All definitions in this document are as provided in the MISO Tariff, the NERC Glossary of Terms Used in Reliability Standards, or are as defined by this document.

Hyperlinks to all BPMs referenced appear in the ‘Bibliography & References’ section of this protocol, and a complete list of MISO Business Practice Manuals (BPMs) is available for reference through MISO’s website .

2.8.7.1 Introduction to the MISO Business Practice Manual

This introduction to the Midcontinent Independent System Operator, Inc. (MISO) Business Practices Manual (BPM) for Demand Response includes basic information about this BPM and the other MISO BPMs. The first section (2.8.7.1.1 *Purpose of MISO Business Practices Manuals* Introduction to the MISO Business Practice Manual) of this Introduction provides information about the MISO BPMs. The second section (2.8.7.1.2 *Purpose of this Business Practices Manual*) is an introduction to this BPM.

2.8.7.1.1 Purpose of MISO Business Practices Manuals

The BPMs developed by MISO provide background information, guidelines, business rules, and processes established by MISO for the operation and administration of MISO markets, provisions of transmission reliability services, and compliance with MISO settlements, billing, and accounting requirements. A complete list of MISO BPMs is available for reference through MISO's website. All definitions in this document are as provided in the MISO Tariff, the NERC Glossary of Terms Used in Reliability Standards, or are as defined by this document.

2.8.7.1.2 Purpose of this Business Practices Manual

This BPM: (1) provides Market Participants (MPs) with the information needed to understand the purpose and application of demand response within the MISO Region; (2) covers the rules, design, and operational elements governing the implementation of the various types of demand response within MISO's Day-Ahead and Real-Time Energy and Operating Reserve Markets; and, (3) describes how demand response can be accredited with Zonal Resource Credits and can be dispatched to interrupt their loads during system emergencies. Demand response used as a Non- Transmission Alternative is discussed separately in BPM-020: Transmission Planning.

MISO employs Security Constrained Unit Commitment (SCUC) and Security Constrained Economic Dispatch (SCED) algorithms to dispatch supply including Demand Response Resources, which simultaneously co-optimizes the Energy and Operating Reserve Markets. The Attachments to the Energy and Operating Reserves BPM explain these functions in greater detail.

This BPM benefits readers who want answers to the following questions:

- What are the roles of MISO and MISO's Market Participants in facilitating the participation of demand response in MISO Energy and Operating Reserve Markets?
- What are the basic concepts that one needs to know to understand the benefits to be derived from demand response?
- What activities must a Market Participant perform in order for its Demand Response Resources to participate in the Energy and Operating Reserve Markets?

2.8.7.2 Overview Of Demand Response

This section presents a high-level description of the role that demand response plays in MISO markets.

DR refers to the ability of a Market Participant to reduce its electric consumption in response to an instruction received from MISO. Market Participants can provide such demand response either with discretely interruptible or continuously controllable loads or with behind-the- meter generation. Market Participants are compensated by MISO for providing such load reductions, as described later in this BPM. MISO market structures provide the opportunity for MPs with demand response to participate either on the demand-side or the supply-side of its markets. For the demand-side, MPs have the ability to make consumption decisions based on the value of energy consumed compared to the market price, and this is discussed further in the BPM for Energy and Operating Reserve Markets. This BPM for Demand Response is devoted to the supply-side, where MPs can offer and monetize the flexibility of demand response to help MISO meet the power balance, meet its ancillary service needs and/or meet its capacity obligations.

2.8.7.2.1 Eligible Market Participants

Three types of entities who have been certified by MISO as Market Participants may provide demand response in MISO:

- Load Serving Entities (LSEs)
- Aggregators of Retail Customers (ARCs)
- End-use customers that have Market Participant status

If your entity is not a certified Market Participant, you must register and be certified as a MISO Market Participant prior to participation in any MISO Market. For more details on the registration processes, see Section 2.8.7.3 *Registration Options for Demand Response.*

2.8.7.2.2 Types of Demand Response Services

MISO employs demand response to:

- Reduce load in the Energy market (i.e., Economic Demand Response)
- Provide Regulating Service, Contingency Reserves (i.e., Operating Reserves Demand Response), or Ramp Capability Product (OR&RCP)
- Reduce demand during system Emergencies (i.e., Emergency Demand Response)
- Substitute for generating capacity (i.e., Planning Resources Demand Response)
- Substitute for transmission (i.e., Demand Response as a Non-Transmission Alternative)

Each of these services is further described below.

(i) Economic Demand Response

A Demand Response Resource (DRR) is a demand resource or behind-the-meter generation (“btmg”) resource that can respond to instructions from MISO. DRRs are the only demand resources that can “inject” Energy on an economic basis. Currently, the minimum size for DRRs to participate in MISO’s markets is one (1) MW.

There are two types of DRRs:

- A DRR – Type I is capable of supplying a fixed, pre-specified quantity of Energy, through physical load reduction, or behind-the-meter generation, to the Energy and Operating Reserve Market when instructed to do so by MISO
- A DRR – Type II is capable of supplying a range (continuum) of Energy through physical load reduction or behind-the-meter generation, to the Energy and Operating Reserve Market and is capable of complying with MISO’s Setpoint Instructions.

Market Participants may submit DRR Energy offers into the Day-Ahead Market and/or the Real Time Market. DRR offers submitted to these two markets are independent, i.e., the price-quantity schedules offered into one market are not linked to the schedules offered into the other market.

Market Participants with DRR offers that clear the market and that subsequently follow MISO instructions, within acceptable tolerance, are paid the Locational Marginal Prices (LMPs) for the Energy they provided to the market through their load reductions. In addition, if necessary, they are made whole to their offers if committed by MISO as part of MISO's Security Constrained Unit Commitment (SCUC) process. These offers can include Energy Offers, Shut-Down Offers and Hourly Curtailment Offers, as described below.

(ii) Operating Reserves and Ramp Capability Product (OR&RCP) Demand Response

OR&RCP Services take on several forms:

- Regulating Reserve
- Spinning Reserve
- Supplemental Reserve
- Ramp Capability Product (RCP)

Together, Spinning Reserve and Supplemental Reserve are also known as Contingency Reserve.

In addition to providing Energy, DRR-Type I and DRR-Type II resources that are technically qualified to do so may provide one or more forms of Operating Reserve Service. DRR-Type I Resources can provide either Energy or Contingency Reserve Service but cannot provide both simultaneously. DRR-Type II Resources may provide Energy and/or one or more Operating Reserve products (as well as the Ramp Capability Product) simultaneously, in a fashion similar to other Generation Resources. MISO uses its SCUC and Security Constrained Economic Dispatch (SCED) algorithms to determine which product a resource will provide in any particular time interval. Currently, the minimum size of DRRs capable of offering these services is one (1) MW.

The technical capabilities required to qualify for each service (see BPM-002) are most stringent for Regulating Service and least stringent for Supplemental Reserve. A DRR that is qualified to provide a more stringent service is generally qualified to provide all of the services with less stringent requirements. Due to its "on/off" nature, DRR-Type I is not allowed to provide Regulation Service or the Ramp Capability Product. Due to the frequency responsive nature of Regulation Service, DRR-Type II resources without telemetry are not allowed to provide Regulation Service. In addition, DRRs cleared for Spinning Reserve Service cannot exceed 40% (on a MW basis) of the market-wide total for cleared Spinning Reserve.

In addition to providing the information required for an Energy Offer, a DRR that is available to provide one or more Operating Reserve products must submit additional pricing information in its offer (e.g., a reserve availability offer). Using these data, MISO will determine whether to clear the DRR offer to provide Energy and/or one or more Operating Reserve services plus RCP. A DRR Type II may submit a price curve (up to 3 MW-price pairs) for each Operating Reserve or other reserve product. A DRR may also choose to submit a daily limit per resource for the amount of Regulation or Contingency Reserve that may be deployed during one Operating Day of the Real Time-Market.

(iii) Emergency Demand Response

Market Participants can also offer to reduce their gross loads specifically when MISO declares an Emergency event (e.g., NERC EEA2 or EEA3 events). MISO's Emergency Demand Response (EDR) Initiative allows, but does not require, EDR resources to indicate their willingness to provide demand response during such events (unless they are also claiming capacity credit as Planning Resources, in which case they must be available to reduce load during Emergency events, as discussed herein). A Market Participant's decision to offer as an EDR is in addition to the choice of creating a DRR and/or an LMR. Currently, the minimum size of these EDRs is one hundred (100 kW).

Each day a Market Participant can decide how much of each of its EDR resources to make available to MISO for EDR service the following day, and at what cost. In addition to providing hourly curtailment costs in its daily EDR offer, the Market Participant can also specify a one-time shutdown cost and a number of operational constraints for each EDR resource. When an Emergency event occurs, MISO will use the information in the EDR offers to decide the order in which to curtail the associated EDR resources, using SCED protocols. EDR offers cannot vary across the hours of the Operating Day.

The EDR Initiative, set forth in Schedule 30 of the MISO Tariff, provides Market Participants with the flexibility to shape their EDR offers based on their near-term circumstances while also providing them with opportunities to increase their operating profits through load curtailments when energy prices are high. In addition, EDR resources may simultaneously qualify as Planning Resources as discussed below.

(iv) Demand Response as a Planning Resource

Planning Resources fall into two potential categories (see Table 2-35): Capacity Resources and LMRs. DRR Type I or II can qualify for either category of Planning Resource, as presented above. Load Modifying Resources (LMRs) qualify as such when the Market Participant registers, and MISO accepts, those assets as LMRs. LMRs are either Demand Resources or Behind-the-Meter Generation (BTMG)4[134]. Registering as an LMR and clearing the Planning Resource Auction (or being committed through a Fixed Resource Adequacy Plan (FRAP)) commits the Market Participant in advance to using the resource to reduce the gross load on the system when instructed to do so by MISO during an Emergency event. Module E-1 of the MISO Tariff prescribes how LMRs are accredited as Planning Resources. Planning Resources have monetary value because they can be substituted for Generation Resources by an LSE in meeting its assigned Planning Reserve Margin Requirement (PRMR). Currently, the minimum size of these LMRs is one hundred (100 kW).

Table 2-35 Planning Resource Categories[135]

	Planning Resource			
	Capacity Resource		**Load Modifying Resource**	
	Generation and External Resource	**DR Resource**	**BTMG**	**DR**
Capacity verification	x	x	x	
Must offer	x			
GADS Data Entry	x		x	
DADs Data Entry		x		x

[134] If the MP registers behind-the-meter generation as an LMR, then its acronym is BTMG. If not registered as an LMR, but registered as another demand response instrument, then its acronym is btmg

[135] MISO BPM-026-r6. Demand Response Business Practices Manual (2021)

Must Respond to EOP	x (x4)	x (x4)	x	x

As shown in Figure 2-19, there are many options available for demand response registration. Note that not all these configurations have been used by MISO Market Participants, but they are available. The finer distinctions between registering as a Capacity Resource, an LMR, a DRR, or an EDR should be evaluated by the Market Participant prior to registering under any of these categories.

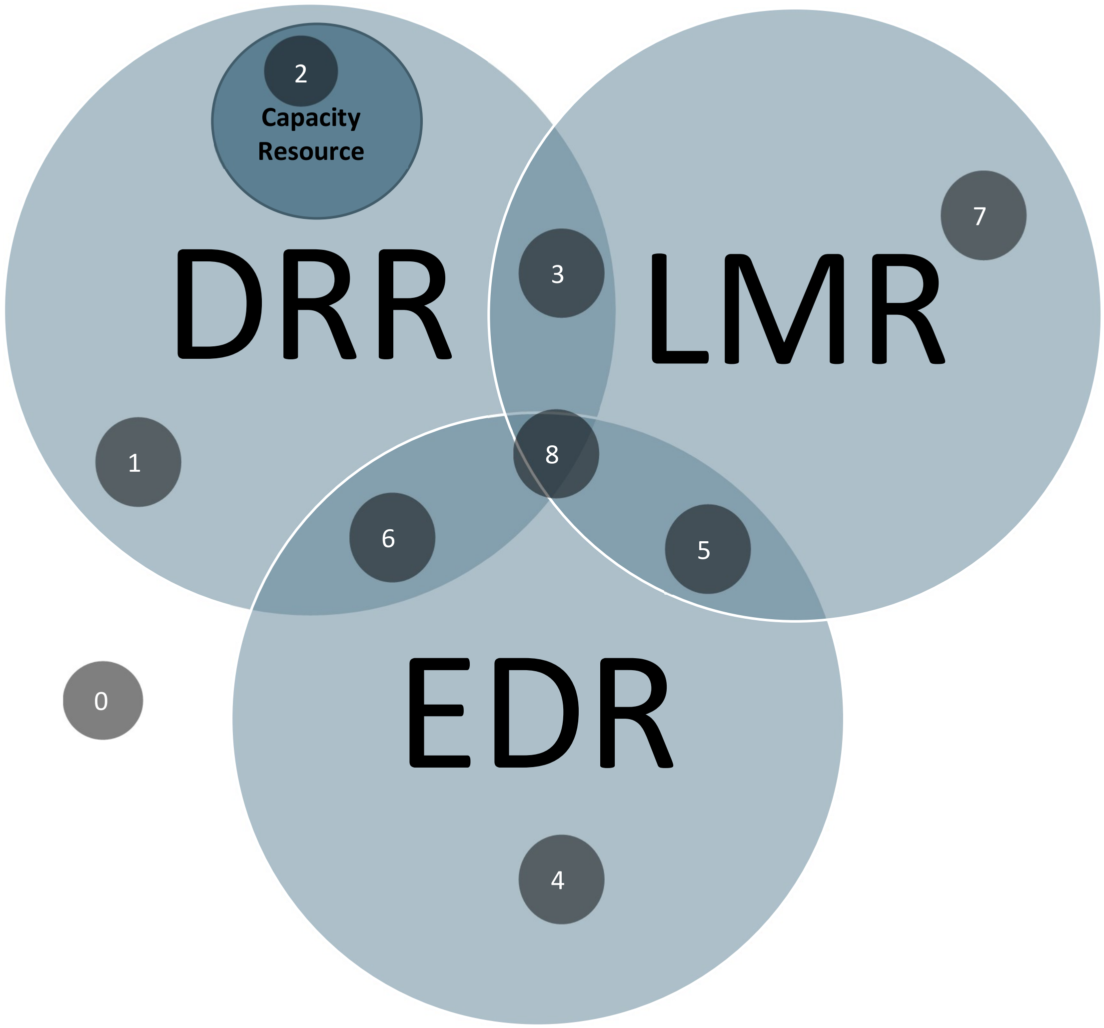

Figure 2-19 Demand Response Registration Options[136]

Notes:

- Note 0: Not MISO Registered; cannot participate
- Note 1: There is no DRR "must offer" requirement, since there are no capacity credits.
- Note 2: DRRs. "must offer" into the Energy & Ancillary Services markets.

[136] MISO BPM-026-r6. Demand Response Business Practices Manual (2021)

- Note 3: Asset registers as an LMR and receives capacity credits, and also registers as a DRR with options to offer into the Energy & Ancillary Services markets.
- Note 4: EDR Only. No capacity credits or "must offer" requirement.
- Note 5: LMR that optionally provides an EDR offer for emergency energy.
- Note 6: Similar to "1", but optional participation in emergencies
- Note 7: LMR only. Not involved in Energy and Ancillary Services markets.
- Note 8: Similar to "5" but can optionally participate in Energy & Ancillary Services markets.

Note that in Options 1 – 8, the entity must be a Certified MISO Market Participant in order to participate.

(v) Demand Response as a Non-Transmission Alternative

Consistent with Attachment FF of the Tariff, both transmission and Non-Transmission Alternatives (NTA) to resolve Transmission Issues will be considered on a comparable basis within the MISO transmission planning process. Non-transmission alternatives include contracted demand response, new or upgraded generators with executed interconnection agreements, and other non- transmission assets (e.g., energy storage not classified as a transmission asset, etc.). Additional details about this use for demand response are presented in Section 4.3.1.2 of BPM-020: Transmission Planning.

2.8.7.2.3 State and Other Retail Regulatory Requirements

In addition to MISO's own standards and requirements for demand response, the states or other retail regulatory entities within the MISO Region may also have various requirements and regulations that must be met regarding the use of demand response. MISO acknowledges the important role that state and other retail regulatory authorities play, in collaboration with FERC, and has developed its demand response initiatives to be supportive of these requirements.

For example, some state Relevant Electric Retail Regulatory Authorities (RERRA) currently do not allow ARCs to do business directly with retail customers subject to their jurisdiction. Such prohibitions may also be imposed by the RERRA having regulatory control over public power entities and cooperatives. Section B.8.7.3.2 below expands on this.

For further details, Market Participants are encouraged to review demand response registration provisions contained in the BPM for Market Registration (BPM-001), and the BPM for Resource Adequacy (BPM-011). Credit requirements for Market Participants with demand response are found in Attachment L of the Tariff; and modeling requirements are specified in the BPM for Network and Commercial Models (BPM-010).

2.8.7.3 Registration Options for Demand Response

Registration of demand resources requires knowledge of two key issues: what are the operational characteristics of the resource ("what is it capable of doing") and how much responsibility for market participation is the Market Participant willing to accept? There are various levels of market interaction available to demand resources; some of these may be beyond the capabilities of the resource (e.g. regulation service), while some may be more than the Market Participant is willing to assume (e.g. does not wish to voluntarily interrupt during certain time periods). Answers to these two key questions will usually provide the Market Participant with a clearer picture of how the resource should be registered with MISO.

Finally, while this section of this BPM is intended to aid Market Participants related to the registration of demand resources, please consult the BPMs for Market Registration (BPM-001) and for Resource Adequacy (BPM-011) for further details or contact your Client Services & Readiness representative.

2.8.7.3.1 Registration as a Market Participant

In order to ensure fair, efficient, and competitive markets, MISO requires all entities desiring to participate in the Open Access Transmission, Energy and Operating Reserve Markets to undergo Market Registration and Qualification processes, also described in section 38.2.2 of the MISO Tariff. Only valid legal entities not otherwise prohibited from market participation by FERC, or any appropriate regulatory authority, may register as a Market Participant.

Opportunities to join in the Open Access Transmission, Energy and Operating Reserve Markets for asset owning and non-asset owning MPs will be in accordance with Commercial Model or other applicable timelines[137], which allows new Applicants to be adequately informed and have their facilities properly modeled before they participate as MPs. To become a Market Participant, an Applicant must complete the Market Participant Qualification Process with MISO by completing the online application, submitting all sections, and required documents, completing the verification of assets by the quarterly Commercial Model deadline (as applicable), and completing the credit requirements as outlined in Section B.8.7.3.4 of the BPM.

To register as a Market Participant, all Applicants will use MISO's Online Registration tool. Applicants will be prompted to complete application sections based on intended market activities. The tool will direct Applicants to complete the applicable sections and accompanying legal documents. It is important to follow the directions carefully for each section as the Applicant's organizational structure and type of activities it wants to engage in will determine the Market Participant's rights and obligations under the MISO Tariff. All applicable forms and supporting documentation must be submitted in accordance with stated deadlines; failure to do so will delay processing of the application.

For full details on the process, please refer to the BPM for Market Registration (BPM-001).

(i) Demand Response Resources (DRRs)

Market Participants who wish to employ a demand resource in the Energy and Operating Reserve market must register their resource as a DRR. Such registration enables the resource to offer energy services, as well as providing any of the OR and RCP services for which the resource is qualified (capable).

The Market Participant may also decide to qualify the resource as a Capacity Resource; if so qualified, the MP accepts the "must offer" requirements associated with Capacity Resources and is also entitled to receive Zonal Resource Credits (ZRCs) commensurate with its ability to reduce load at MISO's peak. Note that a resource's maximum capability to reduce load may not be the same amount by which that resource is able to reduce load at MISO's peak. This distinction will be important to provide during registration. For example, an MP with a particular demand resource may be capable of reducing its load on the system by a maximum of 1.5 MW, but only capable of reducing its load by 1.0 MW at MISO's peak. The difference in these two values may be the result, for example, of the resource having its maximum operation at night or during the winter.

---

137 See BPM-001 Market Registration, Section 3.3: Commercial Model Timeline

As an alternative to registering as a Capacity Resource, a DRR could be registered as an LMR (Planning Resource). An LMR receives ZRCs and is obligated to respond to a MISO Emergency any time they are available during the Planning Year, but no less than five times during the Planning Year, consistent with the availability indicated in the Demand Side Resource Interface (DSRI) and their Scheduling Instructions.

Failure to respond during these Emergencies when sent Scheduling Instructions may result in financial penalties and/or potential disqualification from participation in the Planning Resource Auction. The distinctions described in the previous paragraph related to load reduction would still apply here. While a DRR may also be registered either as a Capacity Resource or as an LMR (or neither, if it so chooses), it may not be registered as both. Market Participants are urged to review the benefits, potential costs, and requirements of various options in order to select the most appropriate to their circumstances and desired operation.

1. DRR Registration

If the DRR was not registered as a part of the initial Market Participant application, a Market Participant may register its DRR in accordance with stated Commercial Model deadlines posted on MISO’s public website. The Market Participant must submit all required documentation to add such resource including, but not limited to:

- Attachment B – Change of Information Form
- Commercial Model Master Template
- Section XIX: Certificate Representation Relationship between Applicant/Market Participant and Owners of Demand Response Resource(s)

All documentation must be received by stated Commercial Model deadlines in order for the resource to be adequately modeled. The Market Participant submitting the registration request will also be required to confirm the requested change to the Commercial Model during the Asset Confirmation period. A member of the Client Services & Readiness team will notify Market Participants when the confirmation period has opened.

As part of the asset registration process for DRRs, Market Participants are required to submit two default offers, each consisting of 24-hourly parameters, for use in the Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market, respectively. These default offer parameters must include the data elements described in Section B.8.7.4 of the BPM.

To register a DRR that will also serve as a Capacity Resource or as an LMR, the Market Participant must also utilize the Module E Capacity Tracking (MECT) tool and comply with all registration deadlines as described in BPM-011 Resource Adequacy. For more information on the registration and qualification process for a DRR to serve as a Capacity Resource or as an LMR, please refer to BPM-011 Resource Adequacy.

(i) Load Modifying Resources (LMRs)

Registering demand response as a Load Modifying Resource commits the resource to respond to any MISO Emergency at least five times per Planning Year when called upon by MISO. In recognition of this responsibility, the resource is granted ZRCs in an amount commensurate with the amount of load reduction provided by the resource at the expected time of MISO’s annual peak demand. Given MISO’s current composition, the expected peak occurs during the period June through August during the hours from 2:00 pm through 6:00 pm. Market Participants must submit a variety of information at registration. While the following lists are intended to assist the Market Participant in understanding the required information, MPs are encouraged to review BPM-011 Resource Adequacy for details and contact Capacity Market Administration with any questions.

1. Demand Resource LMR

For a Demand Resource LMR, qualification and registration information include:

- The Demand Resource must be equal to or greater than 100 kW (grouping a number of smaller resources is allowed in meeting this standard).
- Submitting monthly availability (in megawatts) and notification time (in hours) for the upcoming Planning Year.
- Submitting the documentation listed below if the LMR is only available less than 6 months or requires a notification time greater than or equal to 6 hours. If requested by MISO, the documentation below should be available within five (5) Business Days if an LMR is available less than 9 months or requires a notification time greater than 2 hours:
- Attestation by a senior employee describing the physical capability of the LMR
- LMR operational characteristics or seasonal load output
- Timeline from notice to output (Notification Only)
- Regulatory or contractual limitations
- The Demand Resource must be available to be scheduled for a Demand reduction at the targeted Demand reduction level or by moving to a specified firm service level with no more than 12 Hours advance notice from MISO. For the 2022/2023 Planning Year, a Demand Resource with a notification time requirement greater than 6 hours but less than or equal to 12 hours and a minimum of 10 interruptions allowed during the Planning Year will receive 50% credit as a Planning Resource. For the 2022/2023 Planning Year, Demand Resources with notification time requirements greater than 6 hours but less than or equal to 12 hours with less than 10 interruptions allowed will receive no credit.
- Once Scheduling Instructions are given by MISO that require a Demand reduction, the Demand Resource must be capable of ramping down to meet the targeted Demand reduction level or achieve the firm service level by the Hour designated by MISO's Scheduling Instructions.
- Once the targeted level of Demand reduction or firm service level is achieved, the Demand Resource must be able to maintain the targeted level of Demand reduction or firm service level continuously for at least four (4) consecutive hours.
- The Demand Resource must be capable of being interrupted at least the first five (5) times during the Planning Year when called upon by MISO. For the 2022/2023 Planning Year, Demand Resources with a notification time requirement less than or equal to 6 hours will receive credit as a Planning Resource based on a multiplier of:
- 80% if 5 to 9 interruptions per Planning Year are allowed on the Demand Resource; or,
- 100% if 10 or more interruptions per Planning Year are allowed on the Demand Resource.
- Market Participants with Demand Resources can demonstrate a real power test for accreditation. The real power test of the Demand Resource may be from a MISO called event or a self-scheduled implementation in accordance with section 4.2.9.8 of BPM-011. If a Demand Resource test is not performed for accreditation, additional options outlined in BPM-011 may be utilized.
- If the MP with the Demand Resource does not conduct a real power test under MISO's Tariff (Section 69.A.3.5.j) and is thus not accredited via a real power test, the MP can choose to opt out with potential 3x performance penalties and a credit requirement. If the MP has a regulatory preclusion it can document, it will not be subject to higher penalties. If the MP opts out or has a regulatory exclusion, the MP may

provide operational data, or develop an alternative mechanism, subject to the approval of MISO, by which the demand reduction capability can be demonstrated, and the MP has to participate in at least one of the voluntary LMR drills MISO conducts.

- Unless the Demand Resource is unavailable as a result of maintenance requirements or for reasons of Force Majeure, when a Demand reduction is requested by MISO, the resultant reduction must be a reduction that would not have otherwise occurred within the next twenty-four (24) hour period. There shall be no penalties assessed to a Market Participant representing the entity that has designated the ZRCs from the LMR if the Demand Resource is unavailable for interruption as a result of maintenance requirements or for reasons of Force Majeure, or in the event the specified Demand reduction had already been accomplished for other reasons (e.g., economic considerations, self-scheduling at or above the credited level of Demand Resource, or local reliability concerns in accordance with instructions from the LBA).
- A Demand Resource for which curtailment is voluntary or optional during Emergency events declared by MISO pursuant to MISO's emergency operating procedures will not qualify as an LMR.
- Demand Resources that are offered into the Energy and Operating Reserve Markets as price sensitive Bids are nevertheless obligated to be interrupted during an Emergency pursuant to MISO's emergency operating procedures, regardless of the projected or actual Energy Market LMP.
- MISO will use the MECT tool to ensure that there can be only one MP using ZRCs from a Demand Resource.
- A Market Participant must provide written documentation to MISO from the RERRA having jurisdiction over the Market Participant, or from customers represented by the LMR Market Participant, with the amount and type of Demand Resource and the procedures for achieving the Demand reduction. For a Market Participant without state or other retail regulatory accreditation procedures for a Demand Resource, the Market Participant must secure verification from a third party auditor that is unaffiliated with the Market Participant to provide documentation of the Demand Resource's ability to reduce to the targeted Demand reduction level or to a specified firm service level when called upon by MISO, or provide past performance data that demonstrates such reduction capabilities.

(i) Behind the Meter Generation LMR

A Market Participant that possesses ownership or equivalent contractual rights in a Behind-the-Meter Generator (BTMG) can request accreditation as a BTMG resource by:

- Registering such resource(s) with MISO as documented in BPM-011 Resource Adequacy
- Demonstrating Generation Verification Test Capacity (GVTC) capability for each Planning Year on an annual basis as established in BPM-011 Resource Adequacy, by conducting a real power test or using operational data, and by submitting the GVTC results to MISO no later than October 31 prior to such Planning Year for existing accredited BTMG. All new BTMGs, or an existing accredited BTMG that has an increased installed capacity, shall submit their GVTC to MISO prior to qualification as established in BPM-011 Resource Adequacy.
- Submitting generator availability data (including, but not limited to, NERC GADS information) into a database provided by MISO and as established in BPM-011 Resource Adequacy. A BTMG greater than or equal to 10 MW (based on GVTC) shall provide MISO with generator availability data. A Market Participant

is not required to report generator availability data for a BTMG less than 10 MW if the Market Participant has never provided such data for that BTMG. A Market Participant that begins reporting generator availability data for such a BTMG must continue to report such data; and

- Confirming the BTMG can be available to provide energy with notice not to exceed 12 Hours.
- Submitting monthly availability (in megawatts) and notification time (in hours) for the upcoming Planning Year.
- Submitting the documentation listed below if the LMR is only available less than 6 months or requires a notification time greater than or equal to 6-hours. If requested by MISO, the documentation below should be available within five (5) Business Days if an LMR is available less than 9 months or requires a notification time greater than 2-hours: For the 2022/2023 Planning Year, a BTMG with a notification time requirement greater than 6 hours but less than or equal to 12 hours and a minimum of 10 interruptions allowed during the Planning Year will receive 50% credit as a Planning Resource. For the 2022/2023 Planning Year, BTMG with notification time requirements greater than 6 hours but less than or equal to 12 hours with less than 10 interruptions allowed will receive no credit.
- Attestation by a senior employee describing the physical capability of the LMR
- LMR operational characteristics or seasonal load output
- Timeline from notice to output (Notification Only)
- Regulatory or contractual limitations

(ii) LMR Registration

Each LMR must be registered with MISO in advance of receiving accreditation. Only Certified Market Participants may register the LMR, and this process is completed by accessing the Module E Capacity Tracking (MECT) tool through the secure Market Portal.

To qualify as a Planning Resource the LMR must meet all of the Tariff provisions, summarized in Section 2.8.7.3 *Registration Options for Demand Response.*

For more information on the process and deadlines associated with registering LMRs, refer to BPM-011 Resource Adequacy.

(iii) Emergency Demand Response Resources

A Market Participant within MISO's footprint may register an Emergency Demand Response (EDR) resource if it has the ability to cause a reduction in demand in response to receiving an EDR Dispatch Instruction from MISO because the Market Participant: (i) is the operator of a facility capable of reducing demand; (ii) is a Load Serving Entity (LSE) or Aggregator of Retail Customers (ARC) with a contract that entitles the Market Participant to reduce Load at such facility, or; (iii) has the ability to cause an increase in output from a btmg resource to enable a net demand reduction, in response to receiving an EDR Dispatch Instruction from MISO. Only a Market Participant is allowed to register an EDR resource making itself eligible to submit EDR offers to MISO to reduce demand during an emergency event.

The Market Participant must be able to receive an EDR Dispatch Instruction from MISO via Extensible Markup Language (XML). Additionally, the Market Participant must utilize metering equipment that meets the requirements established in the Tariff, including, but not limited to, the ability to provide integrated hourly kWh values on a Commercial Price Node (CPNode) basis. A Market Participant with a registered EDR resource may provide hourly kWh values for non-interval metered demand reductions (e.g., direct Load control) using the alternative Measurements and Verification Criteria provided in Attachment TT of the Tariff. Measurement of demand reductions will be made on an aggregated applicable CPNode basis to enable the Market Participant's demand reduction to be identified with an LMP; EDR offers can set LMP.

A Market Participant that intends to use a btmg resource for the purpose of reducing demand shall confirm to MISO in writing that: (i) it holds all necessary permits (including, but not limited to, environmental permits) applicable to the operation of the generation resource; (ii) it possesses rights to operate the generation resource that are equivalent to ownership of such unit; and (iii) the generation resource is not a designated Network Resource. Unless notified otherwise, MISO shall deem such representation applies each time the generation resource is used to reduce demand during an emergency event and that the generation resource is being operated in compliance with all applicable permits, including any emissions, run-time limits or other operational constraints that may be imposed by such permits. The Market Participant shall be solely liable for identification of, and compliance with, all such applicable permits.

If the generation resource designated by a Market Participant historically has operated during non-Emergency conditions, the Energy that can be offered under the EDR Initiative is the increase in output from a btmg resource to enable a net Demand reduction, in response to receiving an EDR Dispatch Instruction from MISO. Determination of such output shall be based on the EDR offer and the amount of load reduction provided, as described in the Measurement and Verification protocols.

A Market Participant with a registered EDR resource shall be required to identify if the Demand reduction can be variable (curtail to the firm service level) or alternatively provide a specific level of Demand reduction. Upon receipt of an EDR Dispatch Instruction, the Market Participant shall either: (i) curtail to the firm service level specified in their EDR offer or (ii) provide a specific level of Demand reduction as specified in their EDR offer. Market Participants electing the first option shall be required to identify an expected peak Load in their EDR offer, which can change daily.

The Market Participant is responsible for maintaining Demand reduction information, including the amount in MWh of reduced Demand during emergency events whenever the Market Participant responds to an EDR Dispatch Instruction from MISO. The Market Participant shall provide this information to MISO in accordance with the procedures specified in BPM-005 Market Settlements.

(iv) EDR Registration

Prior to participating in the EDR Initiative, a Market Participant must complete and submit all required EDR registration forms posted on MISO's public website (Markets and Operations > Market Participation > Supplemental Registration). An EDR Participant and its associated load asset or btmg asset must be defined in the EDR registration form. The required registration process includes:

- Submit a case through the Help Center at: https://help.misoenergy.org/
- Attach the following documents to the case:
- EDR Certification Form
- EDR Registration Form

Note the case must be submitted first and then the documents can be attached. Please refer to the MISO Help Center Online Guide located in the Learning Center at the following link: https://miso.csod.com/catalog

In addition to the above documentation, the following documentation is required for ARCs (due to the potential quantity of documents, please send these files to help@misoenergy.org):

- ARC EDR Physical Location Worksheet
- All registration forms
- Section XX Certificate Confirming Fulfillment of Requirements for Applicants Seeking to Participate as Aggregator of Retail Customers (ARC)

An EDR Participant shall verify in writing through the EDR Certification Form that it has received any required approvals from all applicable state regulatory agencies to enable the entity to participate in the EDR Initiative.

The aforementioned documentation must be received by MISO at least 30 days prior to the requested effective date of the EDR resource and the effective start date must be the first day of the month. MISO shall notify the Market Participant when it has met all required qualifications as set forth in Schedule 30, following which the Market Participant is eligible to submit EDR offers beginning on the first day of the month following its approval.

A Market Participant that wants an EDR resource to be accredited with Zonal Resource Credits under Module E-1 must separately register that resource as an LMR, as described in Section B.8.7.3.1 of the BPM.

For questions related to EDR registration, refer to BPM-001 Market Registration or contact a member of the Client Services & Readiness team.

2.8.7.3.2 Registration as an Aggregator of Retail Customers (ARCs)

By definition, an Aggregator of Retail Customers (ARC) is a Market Participant sponsoring one or more DRRs, LMRs, and/or EDRs provided by end-use customers that the ARC does not serve at retail. An ARC can, but need not, be an LSE sponsoring a DRR, LMR, or EDR that is the end-use customer of another LSE.

An entity may choose to participate as an ARC provided, they have met the registration requirements outlined in the BPM for Market Registration (BPM-001) and have received approvals from all required parties, including ensuring that their respective Relevant Electric Retail Regulatory Authority (RERRA) allows for ARC participation. LSEs can aggregate their own end- use customers subject to their retail regulatory authority approval; therefore, they need not register as ARCs to do so.

2.8.7.3.3 ARC Registration

An applicant will indicate its desire to register as an ARC during the Market Participant Application process. BPM-001 Market Registration contains complete information on the registration process. If the Market Participant did not register as an ARC during the initial Market Participant application, it may choose to submit required documentation in accordance with applicable timelines. If the Market Participant intends to register a DRR as an ARC, the Applicant needs to start the registration process at least 30 days prior to the Commercial Model deadline date to allow for registration and approvals (DRR Type I and DRR Type II). Additional information regarding the registration of LMRs as an ARC can be found in BPM-011 Resource Adequacy.

As a pre-requisite, the ARC must ensure it has followed registration procedures for its DRRs, LMRs, or EDRs, including the submission of all required documentation by stated deadlines. Applicants or Market Participants seeking to register as an ARC are required to complete the following document as proof that the entity meets applicable RERRA laws, regulations, or orders regarding participation in MISO’s Energy and Operating Reserve Markets (complying with Tariff 38.6): Certificate Confirming Fulfilment of Requirements for Applicants Seeking to Participate as Aggregator of Retail Customers (ARC), including a list of all RERRA areas that the ARC intends to operate in

An ARC can bundle multiple end-use loads to form an asset, but all loads must be located within a single LSE within an LBA. Each asset may be comprised of one or more Enrollments. Enrollments may be comprised of one or more physical or virtual locations. This applies for DRRs and EDRs. LMRs may only be aggregated up to a Load Zone CPNode level.

Additional data for each end-use load comprising the asset must be provided by applicable deadlines. Market Participants with DRR Type I and/or Type II resources will provide such data through the Demand Response Tool. Market Participants with EDRs will provide the information listed below by completing a physical location template. Market Participants with LMRs will provide the information listed below during registration in the MECT. The Applicant or Market Participant will provide information including, but not limited to, the following for each end-use load comprising the ARC's asset:

- The Local Balancing Authority Area where the end-use loads are located[138];
- The LSE serving each end-use load that the ARC will control;
- The Relevant Electric Retail Rate Authority (RERRA[139]) having jurisdiction over the LSE;
- Expected demand reductions of each registered DRR, LMR, or EDR resource;
- The Measurement & Verification methodology to be used for each identified demand resource;
- The names of relevant contact persons or entities, postal and e-mail addresses, and telephone numbers; and
- A list of end-use customer accounts that comprise the demand resources being registered, including names, addresses, and account numbers of such end-use customers.

In addition, the ARC must certify the following for each of its end-use customers:

- Where the utility serving the customer at retail distributed more than four (4) million MWh in the prior fiscal year.
- The ARC must certify that the laws, regulations, or order(s) of the RERRA do not preclude the end-use customer from participating directly in MISO's Energy and Operating Reserve Markets, providing Capacity or obtaining Zonal Resource Credits under Module E-1 of the Tariff, or being an EDR resource; or,
- Where the utility serving the customer at retail distributed four (4) million MWh or less in the prior fiscal year.
- The ARC must certify that the laws, regulations, or order(s) of the RERRA specifically permit the retail customer to participate directly in MISO's Energy and Operating Reserve Markets, providing Capacity under Module E-1 of the Tariff, or being an EDR resource.

The Market Participant registering as an ARC is required to provide the contact information of the RERRA via the submission of the section XX form. For DRR Type I and Type II registrations, a pull-down list of RERRAs is available in the Demand Response Tool; if the appropriate RERRA is not listed, the ARC will need to notify MISO

---

138 An ARC can bundle multiple end-use loads to form an asset, but all such loads must be located within a single LSE. In addition, a single end-use load can be a DRR Asset. An ARC may register more than one asset.

139 The RERRA will typically be a Commission, but it could also be the board of a public power entity or a rural electric cooperative.

(Market Settlements), and the RERRA will then be added so that the ARC can complete the registration. The ARC is responsible for initial and subsequent validation of the RERRA, notifying MISO of any changes.

Concurrent with MISO review of the application, the LBA and the LSE named by the ARC candidate will be notified, triggering concurrent review regarding the information presented by the ARC. The LBA and LSE have ten (10) business days from receipt of the submitted enrollment to “Confirm” or “Object to ” the enrollment. Inaction on the part of the LBA or LSE will not result in delay of application approval. For DRR Type I and Type II, the Demand Response Tool will list the applicable reasons for “objection” as well as providing a field for Comments (e.g., helpful details regarding the reasons for “objection”). For EDRs, the “objection” reasons are provided in the physical location template. For LMRs, if the “objection” occurs after the LMR registration deadline (March 1st), the ARC will be given one chance to correct the error or clarify the enrollment and if “objection” after the second attempt, the registration will be reviewed by MISO . If the ARC candidate asset is ultimately denied by MISO as a result of the above processes, any further dispute resolution of the resource application occurs through the Tariff’s dispute resolution procedures[140].

2.8.7.3.4 ARC Participation and Review Process

ARC participation is different from other participation in the markets administered by MISO for several reasons. This section attempts to summarize certain issues related to ARC participation. General issues discussed here include the potential for double-counting, communication protocols related to information sharing between ARCs, LBAs, LSEs, and MISO, and re- constitution of load for settlement.

With regard to double-counting, ARC registration requirements include physical addresses and other information which may then be cross-checked by MISO, the LSE, and the LBA with other demand resources registered in MISO Markets. If apparent double counting occurs between MPs during the registration process, MISO will accept end-use customers in a demand resource into a MISO Market on a first-come first-serve basis. LBAs are requested to review and provide important location details (e.g., EPNodes) based on end-use customer addresses and other information and are thus made aware of ARC resources within their service areas. LSEs are requested to review if the end-use customer(s) is already included in the LMR, DRR or EDR for that LSE, if the end-use customer(s) is served by the LSE, account numbers, demand reduction capabilities for assets registering within their service territories and validating and/or providing the CPNode to represent the enrollment.

(i) LSE Responsibilities for EDRs registered by ARCs

Items for review include:

- Correct LSE is listed
- CPNode is owned by the LSE and is still active and not terminated
- Customer account number
- Customer meter number
- Physical location address (Note abbreviations and shortened versions of the street address are acceptable)
- No duplicate account numbers

The following are the confirm/object reasons: Please note other reasons could be included dependent upon changes in the EDR registration process.

---

[140] MISO Tariff Attachment HH: Dispute Resolution Procedures

- Confirm
- Object – the customer is already registered as part of a LMR, DRR, or EDR for the LSE
- Object – the customer is not served by the LSE
- Object – duplicate account number
- Object – the LSE CPNode provided for this location/customer is incorrect
- If MISO does not receive confirm/object within ten business days, the registrations are auto approved unless the approval is subject to RERRA review with respect to a utility with sales equal to or less than 4 million MWhs/fiscal year, in which case failure of the RERRA to confirm within ten business days will result in auto rejection.

With respect to information access for LBAs, the Tariff provides that the LBAs will participate with MISO in reviewing the composition of CPNodes. LBAs will have access to the electrical location and magnitude of resources in an ARC's portfolio of resources in order to perform operational planning studies. Further, LBAs will be notified of ARC demand reduction offers that have been cleared in the day-ahead and real-time markets in order to perform reliability assessments and planning roles in the day-ahead and real-time horizon.

(ii) LBA Responsibilities for EDRs registered by ARCs

Items for review include:

- Correct LBA is listed
- CPNode is still active and not terminated
- Customer account number
- Customer meter number
- Physical location address (Note abbreviations and shortened versions of the street address are acceptable)
- No duplicate account numbers

The following are the confirm/object reasons: Please note other reasons could be included dependent upon changes in the EDR registration process.

- Confirm
- Object – invalid location information
- Object – duplicate account number
- Object – invalid customer account information
- If MISO does not receive confirm/object within ten business days, the registrations are auto approved unless the approval is subject to RERRA review with respect to a utility with sales equal to or less than 4 million MWhs/fiscal year, in which case failure of the RERRA to confirm within ten business days will result in auto rejection.

To the extent that MISO is required to disclose information specific to ARC demand reduction, MISO will need to follow the Disclosure of Certain Confidential Market Participant Data to Balancing Authorities and Transmission Operators provisions set forth in section 38.9.1(A) of Module C of the Energy and Operating Reserve Markets Tariff.

LSEs will have access to all pertinent metering, settlements, and Measurement & Verification (M&V) information associated with the operation of an ARC in an LSE's zone upon submission of requested meter data. Upon submission of settlement data by the ARC, the LSE has ten (10) business days to complete its review and confirm or object to the settlement. If objected by the LSE, the ARC then has ten (10) business days in which to resubmit or dispute the objection. If resubmitted, the LSE then has five (5) business days to review. This process continues, including dispute resolution, until the settlement is approved or denied by MISO, or expires[141]. Also as part of the settlement process, LSEs will have access to data on Actual Energy Injections associated with DRRs (and LMRs/EDRs), within seven (7) days of the Operating Day, so that LSEs can verify ARC-related charges. LSEs will also be notified of cleared ARC load reduction offers in real-time through settlement data.

With specific regard to DRR participation and RERRA approvals:

- MISO will not accept offers from new DRRs until after the ten-day deadline and the Commercial Model has been loaded to production
- MISO will automatically accept a DRR's registration following the ten-day deadline, unless the RERRA objects and unless the approval is subject to RERRA review with respect to a utility with sales equal to or less than 4 million MWhs/fiscal year, in which case failure of the RERRA to confirm within ten business days will result in auto rejection;
- RERRAs can reject a DRR's registration at any time, including after the ten-day notice period, and the demand asset will be promptly removed from participating in MISO's markets; and,
- If an otherwise prohibited end-use customer is registered in a DRR, or an end-use customer becomes non-compliant after having registered with MISO, then MISO will not allow the customer to participate in its markets.

MISO shall review the participation of an ARC in the Energy and Operating Reserve Market when the ARC's settlements submitted under section 38.6 of the Tariff are successfully disputed more than ten percent (10%) of the time by a relevant LSE. The ten (10) percent threshold is based on disputes made by a relevant LSE, irrespective of the RERRA, against an ARC and its representing end-use customers served by the relevant LSE for failure to actually perform as indicated during a given demand response event. This threshold will be addressed quarterly, based on the ARC's rolling average performance with regard to demand response events.

MISO shall have thirty (30) days to conduct a review pursuant to this section of the Tariff. MISO shall refer the matter to the RERRA and may refer the matter to the Independent Market Monitor, if the review indicates the relevant ARC and/or LSE is engaging in activity that is inconsistent with the Energy and Operating Reserve Market Tariff.

2.8.7.3.5 Resource Testing

Prior to participation, each demand resource and/or btmg unit that the Market Participant is proposing to use must document its ability to interrupt load within a prescribed time limit when instructed to do so. The prescribed time limit will depend on the particular service the resource is being qualified to provide. See Section 2.8.7.7 *Resource Testing* in this BPM for more details on Resource Testing. Additional requirements related to LMR testing may be found in the BPM for Resource Adequacy (BPM-011).

2.8.7.3.6 Credit Requirements

To participate in the MISO Markets, all Market Participants must have an approved credit application and must have established a Total Credit Limit with MISO Credit Department in accordance with MISO Credit Policy.

[141]If a settlement is not confirmed within 103 calendar days of the event, it will expire.

Additional details on what is required in the credit application can be found in BPM-001 Market Registration or found in Attachment L of the MISO Tariff.

(i) Changes to Registration

Once a Market Participant is certified, changes may occur in the information originally provided, as specified in the Tariff. Depending on the desired change and the type of demand resource in question, the Market Participant may need to submit documentation or use one of MISO's tools that supports the registration and maintenance of such information. Changes to registrations must follow the applicable timelines.

For questions regarding changes to the following demand response options, please contact the Client Services & Readiness team:

- DRR (Type I or Type II)
- EDR Resource
- ARC Participation

For questions regarding changes to LMRs, please contact the Capacity Market Administration team.

2.8.7.4 Economic Energy, Operating Reserves And Ramp Capability Product

The provision of economic energy is a service different from the provision of operating reserve and other Ancillary Services. Section 2.8.7.2.2 *Types of Demand Response Services* of the BPM refers to the former as, "Economic Demand Response" and to the latter as "Operating Reserve Demand Response." However, the two services are intimately linked through the algorithms MISO uses to schedule the future outputs of DRRs, for example, the Security Constrained Unit Commitment (SCUC) and the Security Constrained Economic Dispatch (SCED) algorithms. Working together, the SCUC and SCED "co-optimize" (i.e., maximize the market benefits derived from) the provision of these services by simultaneously determining which service (or services where qualified), and how much, each DRR should provide in each forthcoming hour of the day. See BPM-002 Attachment A for further insight into the SCUC and SCED optimization algorithms. In light of this interrelationship and because the provision of the two services shares much in common, this section concurrently addresses both Economic Demand Response and Operating Reserves Demand Response.

2.8.7.4.1 Demand Response Characteristics

As stated earlier, DRR-Type I and DRR-Type II resources are the only resources eligible to provide Economic Demand Response in MISO markets.

- A DRR-Type I is defined in Module A of the Tariff as a resource owned by a single Load Serving Entity or ARC within the MISO Balancing Authority Area and that (i) is registered to participate in the Energy and Operating Reserve Markets, (ii) that is capable of supplying a specific quantity of Energy, Contingency Reserve or Capacity, at the choice of the Market Participant, to the Energy and Operating Reserve Market through Behind the Meter Generation and/or controllable Load, (iii) is capable of complying with the Transmission Provider's instructions and (iv) has the appropriate metering equipment installed. Each Demand Response Resource – Type I will be modeled as a Commercial Pricing Node consisting of defined Elemental Pricing Nodes maintained and approved by the Transmission Provider that comprise injections of customer demand response within a single Local Balancing Authority Area for the purposes of scheduling, reporting Actual Energy Injections, and settling Energy and Contingency Reserve transactions.

The Demand Response Resource – Type I can be modeled as aggregations of whole or portions of Elemental Pricing Nodes. Given the appropriate qualification, Demand Response Resource-Type I Resources can provide the following products: Energy, Contingency Reserve, and capacity under Module E.

- A DRR-Type II is defined in Module A of the Tariff a resource owned by a single Load Serving Entity or ARC within the MISO Balancing Authority Area and that (i) is registered to participate in the Energy and Operating Reserve Markets, (ii) is capable of supplying a range of Energy and/or Operating Reserve, at the choice of the Market Participant, to the Energy and Operating Reserve Market through Behind The Meter generation and/or controllable Load, (iii) is capable of complying with Transmission Provider's Setpoint Instructions and (iv) has the appropriate metering equipment installed. Such Resources will be modeled and/or otherwise treated in a manner comparable as Generation Resources and must comply with the same Applicable Reliability Standards as Generation Resources. Given the appropriate qualification, Demand Response Resource-Type II Resources can provide the following products: Energy, Operating Reserve, Up Ramp Capability, Down Ramp Capability, and/or capacity under Module E-1.

To comply with the MISO settlements process, the individual EPNodes comprising a DRR must be EPNodes associated with one Load Serving Entity (LSE).

The two types of DRRs differ primarily with respect to their flexibility in responding to dispatch instructions. A DRR-Type I resource has only two output states (either "on" or "off") whereas a DRR-Type II resource can deliver output over a continuous range of values.

(i) Modeling of DRR-Type I

No special modeling of a DRR-Type I is required in the MISO Network Model, where a DRR-Type I capable load is modeled as regular load. Commercial modeling of DRR-Type I is done using a "DRRNODE1" CPNode, which is similar to the Load Zone CPNode. More information can be found in Section 4.2.3 of the Network and Commercial Models BPM 010.

(ii) Modeling of DRR-Type II

Because a DRR-Type II may consist of both behind-the-meter generators and controllable load, special modeling is required to account for the DRR-Type II properly as a Resource. For Network Model purposes, the load and generator combination is represented by a single equivalent generator. The Commercial Model representation of a DRR-Type II is similar to that of modeling a traditional Generator, in which a single EPNode-CPNode relationship is used. More information can be found in Section 4.2.4 of the Network and Commercial Models BPM 010[142].

(iii) CPNode LMP Determination

The MISO settlement system pays MPs for their energy injections and charges MPs for their energy withdrawals using LMPs corresponding to their respective CPNodes. For each Operating Day, the Day-Ahead / Real-Time (DART) system calculates the LMPs at each EPNode for the Day Ahead Market and again for the Real Time Market. For resources that inject into a single EPNode or loads that withdraw from a single EPNode, their respective CPNode LMPs are simply their respective EPNode LMPs. However, DRR-Type I resources may consist of aggregations of suitable loads located at different EPNodes. In such cases, the hourly LMPs at each CPNode are calculated as a weighted average of the respective hourly LMPs at the EPNodes, where the weighting factors

[142] End-use customer assets can be aggregated as long as all assets originate electrically from a single EPNode.

are the respective weighting factors based on the Target Demand Reductions that the Market Participant sponsoring the DRR submitted when the resource was registered. The calculation is described as follows:

$$CPNLMP_h = \sum_{i=0}^{k} (Weighting\ Factor_i \times EPNLMP_h)$$

Where:

h indexes each of the 24 hours in the Operating Day and

i indexes each of the EPNodes comprising the resource's CPNode.

2.8.7.4.2 Qualifications to Provide Energy

Both types of DRRs are qualified to provide Energy to the market. However, a DRR-Type I is only capable of delivering two levels of output: either zero or its Targeted Demand Reduction. In contrast, a DRR-Type II can deliver varying levels of output spanning a continuum and is also capable of following MISO 5-minute Setpoint Instructions. Because a DRR-Type II is treated as if it were a traditional generator, it must be capable of providing telemetered output data.

2.8.7.4.3 Qualifications to Provide Operating Reserves and RCP

To provide Operating Reserves and/or other Ancillary Services including the Ramp Capability Product, a DRR must be able to deliver energy to the grid within a prescribed time limit specific to the Operating Reserve product offered and must satisfy all other requirements set forth in the Energy and Operating Reserve Market Tariff. Table 2-36 displays these time limits based on reliability standards adhered to by MISO.

Notes for the table below.

- Note 1: Must provide both REG UP and REG DOWN service.
- Note 2: Must respond to AGC instructions within four seconds.
- Note 3: Must be capable of automatically responding to frequency deviations.
- Note 4: DRR-Type I resources only need to provide five-minute interval data within 5 days after a contingency event.
- Note 5: Must be physically located within MISO footprint.

Table 2-36 Operating Reserve Response Time Requirements[143]

Product	Maximum Allowed Response Time	Minimum Continuous Duration	Data Telemetry	Notes
Regulation	4 Seconds	60 Minutes	2 Seconds	1, 2, 3
DRR-Type II				
Spinning Reserve	10 Minutes	60 Minutes		4, 5
DRR-Type I			None	
DRR-Type II			10 Seconds	
Supplemental Reserve	10 Minutes	60 Minutes		4, 5
DRR-Type I			None	
DRR-Type II			10 Seconds	

(i) Regulation

Only DRR-Type II resources can provide Regulation Service because this service requires near- continuous changes in output over a range of values. In addition, the resource must meet the qualifications for providing Regulation service, including the following:

- Fully deployable in both the regulation-up and regulation-down directions
- Capable of automatically responding to and mitigating frequency deviations via speed governor or similar device
- Capable of responding to Automatic Generation Control (AGC) signals within 4 seconds and telemetering its output data at 2-12 second periodicity
- Capable of providing the Regulation Service for a minimum continuous duration of sixty minutes or for the maximum duration specified by Applicable Reliability Standard.

2.8.7.4.4 Spinning Reserve

Both types of DRR resources are eligible to register to provide Spinning Reserve Service. In addition, these resources must be:

- Capable of deploying 100% of their cleared Spinning Reserve (including Spinning Reserve cleared to meet Supplemental Reserve Requirements) within the 10-minute Contingency Reserve Deployment Period
- Capable of sustaining 100% of their cleared Spinning Reserve as energy for a continuous duration of 60 minutes or the maximum duration specified by Applicable Reliability Standards
- Capable of automatically responding to and mitigating frequency deviations if required by Applicable Reliability Standards
- Capable of providing telemetered output data that can be scanned every 2-12 seconds periodicity (except for DRRs-Type I, which need only provide five-minute interval data no later than 5 days after they reduce load in response to a contingency event)

[143] *Ibid.*

- Physically located within the Market Footprint
- Any resource that is qualified to provide Regulating Reserve is also qualified to provide Spinning Reserve. A DRR Type-II registered as a Regulation Qualified Resource must also be registered in the Energy and Operating Reserve Markets as a Spin Qualified Resource and as a Supplemental Qualified Resource. Registration is necessary to allow cleared on-line Regulation Qualified Resources to supply Spinning and/or Supplemental Reserve through substitution of such Resources for Spin Qualified Resources. Currently, DRRs can only clear up to forty (40) percent of the spinning reserve requirement, measured in MWs. A special type of DRR Type I called a Batch-Load Demand Resource (BLDR), (as described in the Baseline Adjustment Examples section of Volume 3, Appendix B), can provide spinning reserve if a Spin Qualified Resource.

(i) Supplemental Reserve

Both types of DRR can provide Supplemental Reserve Service if the resource:

- Is capable of deploying 100% of its cleared Supplemental Reserve within the 10- minute Contingency Reserve Deployment Period
- Is capable of deploying 100% of their cleared Supplemental Reserve for a continuous duration of 60 minutes, or the maximum duration specified by Applicable Reliability Standards
- Has a Minimum Down Time of less than or equal to three hours if a Quick-Start Resource
- Is capable of providing telemetered output data that can be scanned every 2-12 seconds periodicity (except for DRRs-Type I, which need only provide five-minute interval data no later than 5 days after they reduce load in response to a contingency event)
- Is physically located within the market footprint

Any resource that is qualified to provide Spinning Reserve is also qualified to provide Supplemental Reserve. Any Resource registered as a Spin Qualified Resource must also be registered in the Energy and Operating Reserve Markets as a Supplemental Qualified Resource to allow cleared Spin Qualified Resources to supply Supplemental Reserve through substitution of such Resources for Supplemental Qualified Resources. A special type of DRR Type I called a Batch-Load Demand Resource (BLDR), described in Section 2.8.8 *Batch Load Demand Response* below, can provide supplemental reserve if a Supplemental Qualified Resource.

(ii) Ramp Capability Product

Only DRR Type-II resources are eligible to provide the Ramp Capability Product. The Ramp Capability Product is cleared in the Day-Ahead or Real-Time Energy and Operating Reserve Markets to reserve ramp capability to respond to net load variations and includes the following features:

- The Up Ramp Capability and Down Ramp Capability requirements are designed to model both the expected net energy demand change and additional uncertain variation across all market processes and across different system operational conditions at a system level (zonal values will be calculated).
- The contribution of a resource to the ramp capability constraint is limited by its operating limits and its ramp rate over the modeled deployment time. No Market Participant offer price is needed. Market Participants will be able to indicate their offered dispatch status as either “Economic” or “Not Participating”.

- Ramp capability is not explicitly “deployed.” Rather Ramp Capability prepositions resources so that adequate ramp is available in subsequent dispatch intervals. Ramp Capability Requirement Demand Curve will enforce this constraint as a soft constraint.

See BPM-002 Sections 3.4 and 4.2.1.4 for additional Ramp Capability information.

2.8.7.4.5 DRR Offers

MISO maintains a Day-Ahead Schedule Offer and a Real-Time Schedule Offer for each DRR- Type I and DRR-Type II resource. These are standing Offers that are maintained for each market (DA and RT) independent of the other. Initially the standing Offers are established at the time the DRR is registered with MISO and may be updated by the sponsoring Market Participant. Updates may be designated as updating the Day-Ahead Schedule Offer only, the Real-Time Schedule Offer only, or both.

Starting in July 2016, the Real Time Offer Override Enhancement (RTOE) capability went live. RTOE allows the Market Participant to programmatically request overrides of resource capability offers in real time, through the Market Portal’s DART MUI or XML. Overrides are grouped in nine independent sets. Complete sets must be submitted when requesting an override (see Table 2-37). Market Participant overrides will be valid for the current market hour and next market hour. Market Participant override termination date/time will be adjusted if the underlying offer is updated subsequent to the override request, termination date/time will be set to least of a) existing termination date; or b) start of updated schedule market hour.

Table 2-37 Real Time Offer Override Enhancement (RTOE) Sets[144]

Set	GEN/DRRII/EAR Parameters	SER Parameters	DRRI Parameters
Run Times	Notification Time		Notification time
Offer / Unit Limits	Eco Min, Eco Max, Reg Min, Reg Max, Emergency Min, Emergency Max	Reg Min, Reg Max	Target Demand Reduction MW
Offline Response	Offline Resource Limit		
Ramp Rates	RR Up, RR Down, Reg RR (bi-directional)	Reg RR	
Self-Schedule MW	Energy & Regulation MW, Spinning Reserve	Regulation MW	Spinning Reserve MW, Supplemental Reserve MW
	MW, Online & Offline Supp MW		
Dispatch Status	Energy Dispatch Status, Reg status, Spinning Reserve Status, Online	Regulation	Spinning Reserve, Supplemental Reserve
	Supp Status, Offline Supp Status, Ramp Capability		
Commit Status	Energy Commit Status	Energy Commit Status	Energy Commit Status
Offline Control	Off Control Flag, EEE Flag	Off Control Flag, EEE Flag	Off Control Flag, EEE Flag

[144] *Ibid.*

<table>
<tr><td rowspan="2">Fast Ramp Resource</td><td rowspan="2">Fast Ramp Resource Flag</td><td>Fast Ramp Resource Flag, Neutral Zone</td><td rowspan="2"></td></tr>
<tr><td>Lower Limit, Neutral Zone Upper Limit</td></tr>
</table>

MISO uses DRR offers as inputs to the SCUC and SCED (Real-Time Unit Dispatch System only uses SCED). Such offers may be submitted for the Day-Ahead and Real-Time Energy and Operating Reserve Markets. The contents of these offers are briefly described next. Detailed descriptions of the data elements comprising DRR offers can be found in the BPM for the Energy and Operating Reserve Markets (BPM-002).

(i) DRR-Type I

The tables below identify the data elements comprising a DRR-Type I offer.

Notes for Table 2-38 are as follows:

- Note 1: If qualified to provide the service.
- Note 2: The Targeted Demand Reduction is valid for the indicated hour. A DRR-Type I resource is capable of delivering this full reduction or no reduction, i.e., intermediate values are infeasible Note 3: Up to 3 MW/Price pairs may be submitted.
- Note *: Offer parameters can be overwritten in Real-Time Market using Real-Time Offer Override (RTOE). Override is effective next dispatch interval.

Table 2-38 DRR-Type I Economic Data Summary[145]

Data Element	**Units**	**DAM Offer**	**RTM Offer**	**Notes**
Energy Offer	$/MWh	Hourly	Hourly	2
Hourly Curtailment Offer	$/Hr	Hourly	Hourly	2
Shut-Down Offer	$	Daily	Daily	2
Spinning Reserve Offer	$/MW	Hourly	Hourly	1,2,3
Supplemental Reserve Offer	$/MW	Hourly	Hourly	1,2,3
Self-Scheduled Spinning Reserve	MW	Hourly	Hourly*	1
Self-Scheduled Supplemental Reserve	MW	Hourly	Hourly*	1

Notes for Table 2-39 are as follows:

- Note 1: If qualified.
- Note 2: The Targeted Demand Reduction is valid for the indicated hour. A DRR-Type I resource is capable of delivering this full reduction or no reduction, i.e., intermediate values are infeasible.
- Note 3: Default Offers are used if no values are submitted for the day.
- Note *: Offer parameters can be overwritten in Real-Time Market using Real-Time Offer Override (RTOE). Override is effective next dispatch interval.

[145] https://efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=936258568

Table 2-39 DRR-Type I Operating Parameter Data Summary[146]

Data Element	Units	DAM Offer	RTM Offer	Notes
Targeted Demand Reduction Level	MW	Hourly	Hourly*	2,3
Minimum Interruption Duration	hh:mm	Daily	Daily	3
Maximum Interruption Duration	hh:mm	Daily	Daily	3
Minimum Non-Interruption Interval	hh:mm	Daily	Daily	3
Shutdown Time	hh:mm	Hourly	Hourly*	3
Shutdown Notification Time	hh:mm	Hourly	Hourly*	3
Energy Commitment Status	Select	Hourly	Hourly	
Spinning Reserve Dispatch Status	Select	Hourly	Hourly*	1
Supplemental Reserve Dispatch Status	Select	Hourly		1
Maximum Daily Contingency Reserve Deployment	MWh	N/A	Daily	1

(ii) DRR-Type II

Because DRR-Type II resources canprovide a greater range of output to the markets, their Offers are more complex than DRR-Type I Offers.

Table 2-40 through

[146] https://efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=936258568

Table 2-42 identify the DRR-Type II offer data elements.

Notes for the following table:

- Note 1: If qualified.
- Note 2: If not Spin Qualified.
- Note 3: Quick-Start Resources only
- Note 4: Default Offers are used if no values are submitted for Energy and Operating Reserve Markets.
- Note 5: Can take the form of a Block Offer or a Slope Offer. See BPM-011 for further information.
- Note 6: For a DRR-Type II, "No Load Offer" is the hourly price for maintaining a readiness to reduce load.
- Note 7: For a DRR-Type II, its "Startup Offer" is the daily price for being available to reduce load.
- Note 8: Up to 3 MW/Price pairs may be submitted.
- Note *: Offer parameters can be overwritten in Real-Time Market using Real-Time Offer Override (RTOE). Override is effective next dispatch interval.

Table 2-40 DRR-Type II Economic Data Summary[147]

Data Element	**Units**	**DAM Offer**	**RTM Offer**	**Notes**
Energy Offer Curve	MW,$/MWh	Hourly	Hourly	5
No-Load Offer	$/Hr	Hourly	Hourly	4,6
Regulating Reserve Capacity Offer	$/MWh	Hourly	Hourly	1,5,8
Regulating Reserve Mileage Offer	$/MW	Hourly	Hourly	1
Spinning Reserve Offer	$/MWh	Hourly	Hourly	1,5,8
On-Line Supplemental Reserve Offer	$/MWh	Hourly	Hourly	1,2,5,8
Off-Line Supplemental Reserve Offer	$/MWh	Hourly	Hourly	3,5,8
Hot Start-Up Offer	$	Daily	Daily	4,7
Intermediate Start-Up Offer	$	Daily	Daily	4,7
Cold Start-Up Offer	$	Daily	Daily	4,7

[147] https://efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=936258568

Self-Scheduled Regulation	MW	Hourly	Hourly*	1
Self-Scheduled Spinning Reserve	MW	Hourly		1
Self-Scheduled On-Line Supplemental Reserve	MW	Hourly		1,2
Self-Scheduled Off-Line Supplemental Reserve	MW	Hourly		3
Self-Scheduled Energy	MW	Hourly		
Fast Ramping Resource Flag	True/False	N/A	Hourly	

Notes for the following tables:

- Note 1: Default Offers are used if no values are submitted for Energy and Operating Reserve Markets
- Note 2: Hourly Ramp Rate is used in Day-Ahead and RAC
- Note 3: Ramp Rates may be submitted by MPs at any time and remain fixed until changed by MPs
- Note 4: Only applicable to Quick-Start Resources
- Note 5: Not applicable to Dispatchable Intermittent Resources in the Real-Time Market
- Note 6: Not applicable to Dispatchable Intermittent Resources
- Note 7: Only applicable to Dispatchable Intermittent Resources
- Note 8: Participant-limited to the level achieved during last deployment or test of Offline Supplemental Reserves issued by MISO
- Note 9: Only applicable to DRR-Type II Resources in Real-Time Market
- Note *: Offer parameters can be overwritten in Real-Time Market using Real-Time Offer Override (RTOE). Override is effective next dispatch interval.

Table 2-41 DRR-Type II Commitment Operating Parameter Data Summary[148]

Generation and DRR-Type II Offer Data	Units	Day-Ahead Schedule Offer	Real-Time Schedule Offer	Notes
Hot Notification Time	hh:mm	Hourly	Hourly*	
Hot Start-Up Time	hh:mm	Hourly	Hourly	
Hot to Intermediate Time	hh:mm	Daily	Daily	
Intermediate Notification Time	hh:mm	Hourly	Hourly	
Intermediate Start-Up Time	hh:mm	Hourly	Hourly	
Hot to Cold Time	hh:mm	Daily	Daily	
Cold Notification Time	hh:mm	Hourly	Hourly	
Cold Start-Up Time	hh:mm	Hourly	Hourly	
Maximum Daily Starts	Integer	Daily	Daily	
Maximum Daily Energy	MWh	Daily	Daily	
Minimum Run Time	hh:mm	Daily	Daily	
Maximum Run Time	hh:mm	Daily	Daily	
Minimum Down Time	hh:mm	Daily	Daily	

[148] https://efis.psc.mo.gov/mpsc/commoncomponents/viewdocument.asp?DocId=936258568

Commitment Status	Select	Hourly	Hourly	1
Max Daily Regulation Up Deployment	MWh	NA	Daily	9
Max Daily Regulation Down Deployment	MWh	NA	Daily	9
Max Daily Contingency Reserve Deployment	MWh	NA	Daily	9

Table 2-42 DRR-Type II Dispatch Operating Parameter Data Summary[149]

Generation and DRR-Type II Offer Data	Units	Day-Ahead Schedule Offer	Real-Time Schedule Offer	Notes
Hourly Economic Minimum Limit	MW	Hourly	Hourly*	1
Hourly Economic Maximum Limit	MW	Hourly	Hourly*	1,5
Hourly Regulation Minimum Limit	MW	Hourly	Hourly*	1,6
Hourly Regulation Maximum Limit	MW	Hourly	Hourly*	1,6
Hourly Emergency Minimum Limit	MW	Hourly	Hourly*	1
Hourly Emergency Maximum Limit	MW	Hourly	Hourly*	1,5
Maximum Off-Line Response Limit	MW	Hourly	Hourly*	1,4,6,8
Energy Dispatch Status	Select	Hourly	Hourly*	1
Regulating Reserve Dispatch Status	Select	Hourly	Hourly*	1,6
Spinning Reserve Dispatch Status	Select	Hourly	Hourly*	1,6
On-line Supplemental Reserve Dispatch Status	Select	Hourly	Hourly*	1,6
Off-line Supplemental Reserve Dispatch Status	Select	Hourly	Hourly*	1,4,6
Hourly Single-Directional-Down Ramp Rate	MW/min	N/A	Hourly*	1,3
Hourly Single-Directional-Up Ramp Rate	MW/min	N/A	Hourly*	1,3

[149] *Ibid.*

Hourly Bi-Directional Ramp Rate	MW/min	N/A	Hourly*	1,3
Hourly Ramp Rate	MW/min	Hourly	Hourly	1,2,3
Single-Directional-Down Ramp Rate Curve	MW/min	N/A	Hourly	3
Single-Directional-Up Ramp Rate Curve	MW/min	N/A	Hourly	3
Bi-Directional Ramp Rate Curve	MW/min	N/A	Hourly	3
Combined Cycle Status	Select	Daily	Daily	
Forecast Maximum Limit	MW	N/A	Rolling 5-Min	7
Ramp Capability Dispatch Status	Select	Hourly	Hourly*	

2.8.7.4.6 Commitment and Dispatch

MISO uses two optimization algorithms, SCUC and SCED, to optimally schedule Resources in a least cost manner to meet the energy balance in its Day-Ahead and Real-Time Markets. Security Constrained Unit Commitment (SCUC) optimally commits Resources in a least cost manner considering Start Up (Shutdown) Offers and No Load (Hourly Curtailment) Offers. Security Constrained Economic Dispatch (SCED) optimally dispatches Resources to operating levels to meet Day-Ahead or Real-Time needs. Both algorithms are employed to simultaneously clear Supply Offers and Demand Bids for each time interval, efficiently allocate transmission capacity to Day-Ahead or Real-Time Schedules by resolving transmission congestion and commit and dispatch Resources at least-cost to meet the Energy and Congestion Management requirements throughout the Operating Day.

(i) DRR-Type I

The figure below shows the operation timeline for DRR Type I.

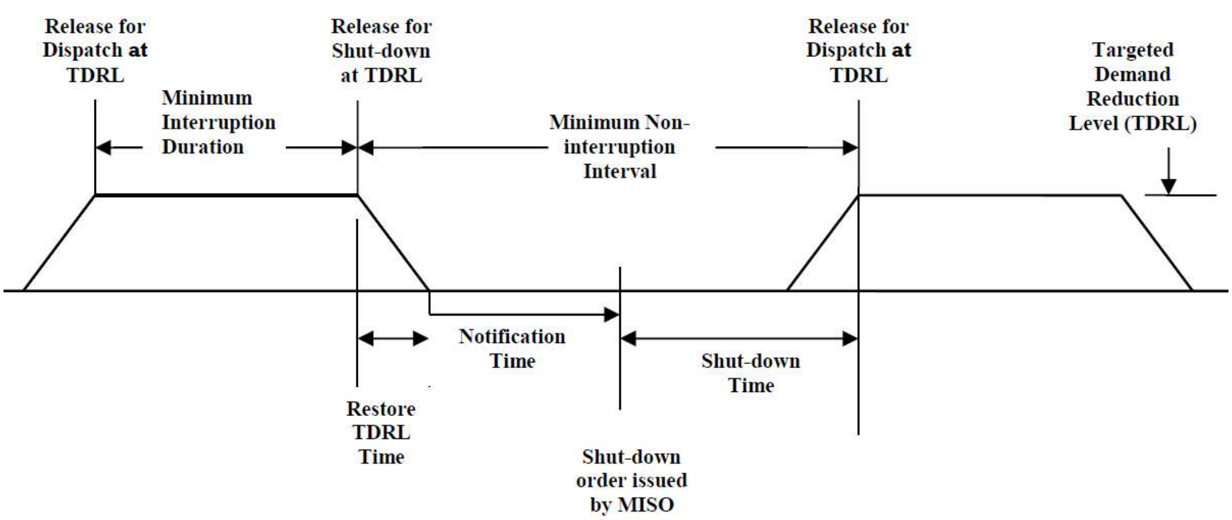

Figure 2-20 DRR-Type I Operation Timeline

(ii) DRR-Type I Commitment Status

The table summarizes how DRR-Type I operating parameters are used in MISO's Day-Ahead Energy and Operating Reserve Market and Reliability Assessment Commitment (RAC) process to commit and economically dispatch these resources.

Table 2-43 DRR-Type I Commitment and Dispatch

Parameter	Validation	Use
Shut-Down Notification Time	The Shut-Down Notification Time parameter is submitted as part of the Day-Ahead Schedule Offer and Real-Time Schedule Offer. These times are accepted in hh:mm format. The default value is 00:00. This value cannot exceed 23:59.	The Shut-Down Notification Time is used in evaluating the commitment in the Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market. This parameter, in conjunction with the associated Shut-Down Time, establishes the time required to shut down the Resource at the Targeted Demand Reduction Level.
Shut-Down Time	The Shut-Down Time parameter is submitted as part of the Day-Ahead Schedule Offer and Real-Time Schedule Offer. This time is accepted in hh:mm format.	The Shut-Down Time is used in evaluating commitment in the Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market. This parameter, in conjunction with the associated Shut-Down Notification Time, establishes the time required to shut down the Resource at the Targeted Demand Reduction Level.
Minimum Interruption Duration	The Minimum Interruption Duration is submitted as part of the Day-Ahead Schedule Offer and Real-Time Schedule Offer. This time is accepted in hh:mm format.	MISO schedule commitments in the Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market are for at least as many consecutive hours as specified by Minimum Interruption Duration. Commitment times may be for greater than the Minimum Interruption Duration if a DRR -Type I is economic for additional hours.
Minimum Non-Interruption Interval	The Minimum Non-Interruption Interval is submitted as part of the Day-Ahead Schedule Offer and Real-Time Schedule Offer. This time is accepted in hh:mm format. The default value is 00:00.	The Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market commitments respect the Minimum Non-Interruption Interval in determining when a DRR -Type I is available for shut down.
Maximum Interruption Duration	The Maximum Interruption Duration is submitted as part of the Day-Ahead Schedule Offer and Real-Time Schedule Offer. This time is accepted in hh:mm format. The default value is 99:99.	The Maximum Interruption Duration restricts the number of consecutive hours a DRR -Type I can be committed during the Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market.
Contingency Reserve Status	The Contingency Reserve Status is submitted as part of the Day-Ahead Schedule Offer and Real-Time Schedule Offer. Valid entries for Contingency Reserve Status are “online” and “offline”.	The Contingency Reserve Status determines whether the DRR – Type I will be considered to clear and deploy Spinning Reserves, or whether it will be considered to clear and deploy Supplemental Reserves. See Sections 0 and 8.2.9 for more information on the Contingency Reserve Status.

Maximum Daily Contingency Reserve Deployment	The Maximum Daily Contingency Reserve Deployment is submitted as part of the Real-Time Schedule Offer, in MWh.	The Maximum Daily Contingency Reserve is the maximum MWh a Resource is able to deploy as Contingency Reserve over a 24 hour Operating Day of the Real-Time Energy and Operating Reserve Market.

Both a Day-Ahead Schedule Offer and Real-Time Schedule Offer have an associated DRR-Type I commitment status. The commitment status impacts the decisions made in unit commitment. The three commitment status options are:

- Not Participating – Designates the DRR-Type I is not available for Energy commitment in the Energy and Operating Reserve Markets for that Hour but could be available for Contingency Reserve clearing depending on the Spinning Reserve or Supplemental Reserve Dispatch Status.
- Emergency – Designates the DRR-Type I is available for commitment for Energy in Emergency situations only.
- Economic – Designates the DRR-Type I is available for commitment for Energy by MISO.

For a DRR – Type I that is a designated Capacity Resource, the Not Participating Commitment Status is only applicable if that Resource is unavailable due to a forced or planned outage or other physical operating restrictions.

The single value commitment status can vary by hour in the Day-Ahead Schedule Offer or Real- Time Schedule Offer and will override the default status. The default status is set during asset registration. If the MISO SCUC algorithm commits the DRR Type I resource, then because of the on/off property of this asset, it is cleared for energy by definition.

(iii) DRR-Type I Offer Dispatch Status

Dispatch Status for a DRR-Type I can be selected on an hourly basis for Spinning Reserve and Supplemental Reserve (if it is a Spin Qualified Resource), or for Supplemental Reserve (if it is a Supplemental Qualified Resource but not a Spin Qualified Resource).

Spinning Reserve or Supplemental Reserve Dispatch Status selections made in combination with Commitment Status selections allow a DRR-Type I to choose whether or not they can be committed for Energy only or dispatched for Spinning Reserve or Supplemental Reserve only, as applicable, under both normal and Emergency conditions. Valid DRR-Type I Dispatch Status selections are: Economic, Self-Schedule, Emergency, Not Qualified or Not Participating. For a DRR-Type I that is a designated Capacity Resource and is qualified to provide Spinning Reserve and/or Supplemental Reserve, the Not Participating Spinning Reserve Dispatch Status or Supplemental Reserve Dispatch Status is only applicable if such Resource is unavailable due to a forced or planned outage or other physical operating restrictions.

The table below shows the valid Dispatch Status and Commit Status selection combinations to achieve the desired results.

Table 2-44 Valid DRR-Type I Commit and Dispatch Status Combinations

		Normal Operations	**Emergency Operations 1**

Commit Status	Spin or Supp Dispatch Status	Energy Only	Spin/ Supp Reserve Only	Either	None	Energy Only	Spin/ Supp Reserve Only	Either	None
Economic	Economic			X				X	
Economic	Not Participating	X				X			
Economic	Not Qualified	X				X			
Economic	Self-Schedule			X				X	
Economic	Emergency	X						X	
Not Participating	Economic		X				X		
Not Participating	Not Participating				X				X
Not Participating	Not Qualified				X				X
Not Participating	Self-Schedule		X				X		
Not Participating	Emergency						X		
Emergency	Economic			X				X	
Emergency	Not Participating				X	X			
Emergency	Not Qualified				X	X			
Emergency	Self-Schedule			X				X	
Emergency	Emergency				X			X	

DRR-Type I Dispatch status may be selected as part of the Day-Ahead and Real-Time Schedule Offer and will override the default status. The default status value is set during asset registration. For a DRR Type I that is a Spin Qualified Resource, if the MP elects 'not participating' for its Commit Status and either 'economic' or 'self-schedule' for its dispatch status, then the DRR Type I resource can be cleared for Spinning Reserve but the MP will not be guaranteed recovery of any Shut Down Offers because the resource has not been committed by MISO through its SCUC algorithm.

(iv) DRR-Type I Self-Schedule

DRR-Type I resources can only submit Self-Schedules for Energy, Spinning Reserve or Supplemental Reserve in amounts less than or equal to their Targeted Demand Reduction Levels (BPM-002 section 4.2.4.3.4). Submitting a Self-Schedule for Spinning Reserve or Supplemental Reserve will generally ensure that the DRR-Type I resource clears for Contingency Reserve provided that the DRR-Type I has not been committed for Energy. If the Self-Schedule MW value is less than the Targeted Demand Reduction Level, the Resource may clear Spinning Reserve or Supplemental Reserve above the Self-Schedule MW amount, based upon the DRR- Type I Spinning Reserve

Offer or Supplemental Reserve Offer, on an economic basis as part of the Energy and Operating Reserve Markets clearing process. A Self-Schedule is a price taker up to Self-Schedule MW level.

MISO will reduce Self-Schedules if such schedules cannot be physically implemented based upon the submitted Targeted Demand Reduction Level. Additionally, MISO may reduce accepted Self- Schedules as necessary to manage transmission constraints, maintain Operating Reserve requirements, satisfy Energy demand and/or maintain reliable operating conditions. In no case will MISO violate the DRR-Type I operating parameters; consequently, it will either accept the Self-Schedule or de-commit the DRR-Type I resource.

(v) DRR-Type II

Figure 2-21 below presents an Operational Timeline for DRR-Type II resource commitment and dispatch.

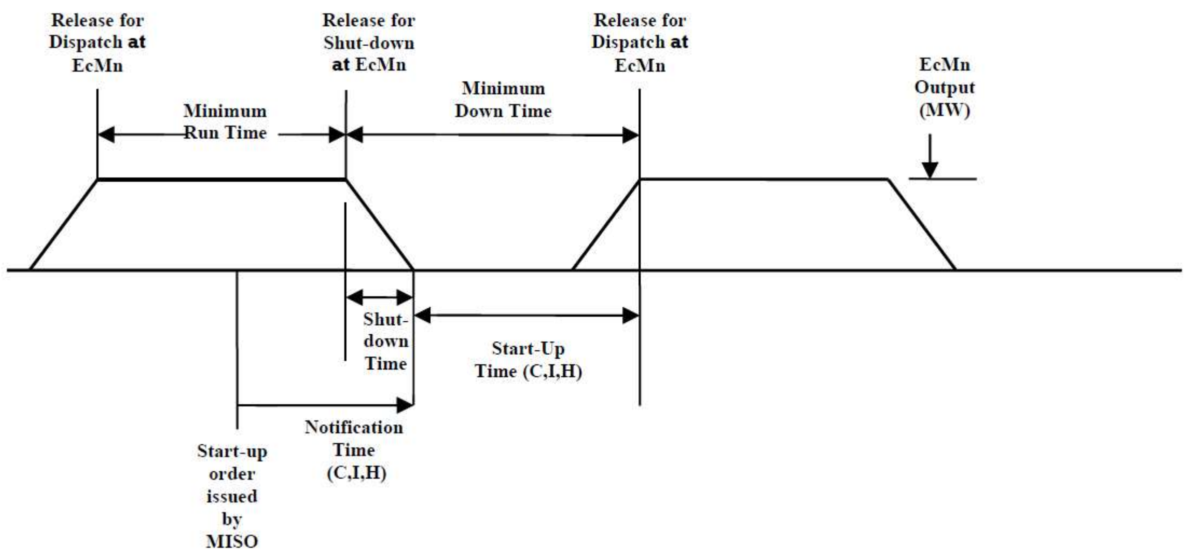

Figure 2-21 DRR-Type II Operation Timeline

The table below summarizes how DRR-Type II operating parameters are used in MISO's Day- Ahead and Real-Time Energy and Operating Reserve Market and Reliability Assessment Commitment ("RAC") processes to commit and economically dispatch DRR Type II resources. Section 4.2.3 of BPM – 002 further describes commitment and dispatch of DRR Type II, similar to generation resources.

Table 2-45 DRR-Type II Commitment and Dispatch

Parameter	Use	Format and Validation
Start-up Notification Time	The Start-up Notification Time is used in evaluating the commitment in the Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market. This parameter, along with the associated Start-up Time, establishes the time required for the resource to begin following dispatch instructions to vary its load.	The Start-up Notification Time parameter is submitted as part of the Day-Ahead and Real-Time Schedule Offer. These times are accepted in hh:mm format. These values must be less than or equal to 23:59.
Start-up Time	See Above	The Start-Up Time parameter is submitted as part of the Day-Ahead and Real-Time Schedule Offer. These times are accepted in hh:mm format.

Hourly Economic Minimum Limit	The Hourly Economic Minimum Limit designates the minimum Energy output, in MW, from the Resource under non-Emergency conditions. This value may vary from hour to hour in the Day-Ahead Schedule Offer and Real-Time Schedule Offer. The Overall Economic Minimum Limit affects both commitment and dispatch in both the Day-Ahead and Real-Time Energy and Operating Reserve Markets. Energy and Operating Reserve Market dispatch is from Hourly Economic Minimum Limit to Hourly Economic Maximum Limit under normal conditions.	The Hourly Economic Minimum Limit may be submitted to override the default Offer, for both the Day-Ahead Schedule Offer and Real-Time Schedule Offer. The data value accepted may be to the tenth of a MW. This value is expected to be negative, indicating the amount of baseline load when no Energy is cleared.
Hourly Economic Maximum Limit	The Hourly Economic Maximum Limit designates the maximum Energy available, in MW, from the Resource under non-Emergency conditions. This value may vary from hour to hour through submission in the Day-Ahead Schedule Offer and Real-Time Schedule Offer. The Overall Economic Maximum Limit affects both commitment and dispatch in both the Day-Ahead and Real-Time Energy and Operating Reserve Markets. Energy and Operating Reserve Market dispatch is from Hourly Economic Minimum Limit to Hourly Economic Maximum Limit under normal conditions	The Hourly Economic Maximum Limit may be submitted to override the default Offer as part of the Day-Ahead Schedule Offer and/or Real-Time Schedule Offer. The data value accepted may be to the tenth of a MW.
Hourly Regulation Minimum Limit	The Hourly Regulation Minimum Limit designates the minimum operating level, in MW, at which the Resource can operate while scheduled to potentially provide Regulating Reserves. This value may vary from hour to hour through submission in the Day-Ahead Schedule Offer and Real-Time Schedule Offer. The Hourly Regulation Minimum Limit does not affect commitment but may affect Energy dispatch in both the Day-Ahead and Real-Time Energy and Operating Reserve Markets.	The Hourly Regulation Minimum Limit may be submitted to override the default offer as part of the Day-Ahead Schedule Offer and/or Real-Time Schedule Offer. The data value accepted may be to the tenth of a MW.
Hourly Regulation Maximum Limit	The Hourly Regulation Maximum Limit designates the maximum operating level, in MW, at which the Resource can operate while scheduled to potentially provide Regulating Reserves. This value may vary from hour to hour through submission in the Day-ahead Offer and Real-Time Schedule Offer. The Hourly Regulation Maximum Limit does not affect commitment but may affect Energy dispatch in both the Day-Ahead and Real-Time Energy and Operating Reserve Market.	The Hourly Regulation Maximum Limit may be submitted to override the default Offer as part of the Day-Ahead Schedule Offer and/or Real-Time Schedule Offer. The data value accepted may be to the tenth of a MW.
Hourly Emergency Minimum Limit	The Hourly Emergency Minimum Limit designates the lowest level of energy, in MW; the Resource can produce and maintain a stable level of operation under Emergency conditions.	The Hourly Emergency Minimum Limit may be submitted to override the default Offer as part of the Day-Ahead Schedule Offer and/or Real-Time Schedule Offer. The data value accepted may be to the tenth of a MW.
Hourly Emergency Maximum Limit	The Hourly Emergency Maximum Limit designates the highest level of Energy, in MW; the Resource can produce and maintain a stable level of operation under Emergency conditions.	The Hourly Emergency Maximum Limit may be submitted to override the Default Offer as part of the Day-Ahead Schedule Offer and/or Real-Time Schedule Offer. The data value accepted may be to the tenth of a MW.
Minimum Run Time	MISO scheduled commitments in the Day- Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market are for at least as many consecutive hours as specified by the Minimum Run Time. Commitment times may be for greater than the Minimum Run Time if a Resource is economic for additional hours.	The Minimum Run Time is submitted as part of the Day-Ahead and Real-Time Schedule Offer. This time is accepted in hh:mm format.
Minimum Down Time	The Day-Ahead Energy and Operating Reserve Market and the Real-Time Energy and Operating Reserve Market commitments respect the Minimum Down Time in determining when a unit is available for Start-up.	The Minimum Down Time is submitted as part of the Day-Ahead and Real-Time Schedule Offer. This time is accepted in hh:mm format.

Maximum Run Time	The Maximum Run time restricts the number of hours a unit can be run during the Day- Ahead Energy and Operating Reserve Market or during a study period for the Real-Time Energy and Operating Reserve Market.	The Maximum Run Time is submitted as part of the Day-Ahead and Real-Time Schedule Offer. This time is accepted in hh:mm format.
Maximum Daily Starts	The Maximum Daily Starts are the maximum number of times a unit may receive a Start-Up per day during the Day-Ahead Energy and Operating Reserve Market or during a study period of the Real-Time Energy and Operating Reserve Market.	The Maximum Daily Starts are submitted as part of the Day-Ahead and Real-Time Schedule Offer. These times are accepted in integer number of times.
Maximum Daily Energy	The Maximum Daily Energy is the maximum MWh a Resource is able to supply over a 24 hour period during the Day-Ahead Energy and Operating Reserve Market or during a study period of the Real-Time Energy and Operating Reserve Market.	The Maximum Daily Energy is submitted as part of the Day-Ahead and Real-Time Schedule Offer, in MWh.
Maximum Daily Contingency Reserve Deployment	The Maximum Daily Contingency Reserve restricts the amount of contingency reserve that may be deployed on a DRR-Type II in the Real-Time Energy and Operating Reserve Market. It is not used in the Day-Ahead Market or RAC process.	The Maximum Daily Contingency Reserve Deployment limit is submitted as part of the Real-Time Schedule Offer. The format is MWh.
Maximum Daily Regulation Up Deployment	The Maximum Regulation Up Deployment restricts the amount of Regulating Reserve Up that may be deployed on a DRR-Type II in the Real-Time Energy and Operating Reserve Market. It is not used in the Day-Ahead Market or RAC process.	The Maximum Daily Regulation Up Deployment limit is submitted as part of the Real-Time Schedule Offer. The format is MWh.
Maximum Daily Regulation Down Deployment	The Maximum Daily Regulation Down Deployment restricts the amount of Regulating Reserve Down that may be deployed on a DRR-Type II in the Real-Time Energy and Operating Reserve Market. It is not used in the Day-Ahead Market or RAC process.	The Maximum Daily Regulation Down Deployment limit is submitted as part of the Real-Time Schedule Offer. The format is MWh.

2.8.7.4.7 Market Price Determination

This section briefly describes MISO market clearing processes that determine prices in the Day- Ahead Energy and Operating Reserve Markets and the Real-Time Energy and Operating Reserve Markets.

(i) Day-Ahead Markets

Offers for Energy and Operating Reserve submitted to the Day-Ahead Energy and Operating Reserve Markets are simultaneously cleared for each hour of the following Operating Day using SCUC and SCED computer-based algorithms to satisfy the Energy Demand Bids and Operating Reserve requirements of that Operating Day.

The Day-Ahead market clearing process produces hourly ex-ante Locational Marginal Energy Prices (LMPs) at each EPNode and hourly ex-ante Market Clearing Prices (MCPs) at each CPNode for Regulating Reserve, Spinning Reserve, Supplemental Reserve, and the Ramp Capability Product. The pricing algorithm has been enhanced with the Extended Locational Marginal Pricing (“ELMP”) mechanism that allows the cost of committing Fast Start Resources, and the Energy cost of Fast Start Resources dispatched at limits to set prices. It also produces hourly schedules for Energy Demand, Energy supply, Regulating Reserve, Spinning Reserve and Supplemental Reserve, Up and Down Ramp Capability for each Resource that was offered into the Day-Ahead Market. If the 40% constraint on the amount of DRRs that clear Spinning Reserve in the Day-Ahead Market binds, then the MCP for cleared DRRs will differ from the MCP for cleared Generation Resources.

(ii) Real Time Markets

Offers for Energy and Operating Reserve submitted to the Real-Time Energy and Operating Reserve Markets are simultaneously cleared every five minutes using the SCED computer-based algorithm to satisfy the forecasted 5-minute Energy Demand and Operating Reserve requirements of the Real-Time Markets based on actual operating conditions, as captured by MISO’s State Estimator. Similar to the Day-Ahead, the ELMP mechanism allows the cost of committing Fast Start Resources (“FSR”), the Energy cost of Fast Start Resources dispatched at limits and Emergency Demand Response Resources to set price. ELMP also provides the mechanism to introduce emergency pricing, in an ex-post manner, to prevent inefficient price depression during system or local area shortage conditions when MISO utilizes Emergency Resources, including the Emergency range of available resources, Emergency Demand Response Resources, Load Modifying Resources, External Resources that are qualified as Planning Resources or Emergency Energy purchases.

The Real-Time market clearing process produces five-minute ex-ante LMPs for Energy along with five-minute ex-ante MCP values for Regulating Reserve, Spinning Reserve, Supplemental Reserve, and Ramp Capability Product, and five-minute Dispatch Targets for each Resource operating in the Real-Time markets.

The SCED operating in real-time is supported by a Reliability Assessment Commitment (RAC) process that identifies in advance of Real-Time dispatch the need for additional resources to ensure that sufficient capacity will be online to meet Real-Time operating conditions. The RAC process utilizes the same SCUC algorithm employed in the Day-Ahead Markets to minimize the cost of committing the capacity needed to meet forecasted Energy Demand, confirmed Energy Interchange Schedule Exports, and forecasted Operating Reserve requirements.

The RAC process identifies the need for committing additional Resources after the clearing of the Day-Ahead Energy and Operating Reserve Market, after posting the Day-Ahead Markets results but before the start of the Operating day, or anytime during the Operating Day, as required.

Under LMP, DRR Type IIs within their limits can set price. Under ELMP, FSRs and EDRs can set price. Any DRRs can set MCPs for products they are qualified to provide. The BPM for Energy and Operating Reserve Markets (BPM-002), provides more detailed descriptions of how the Day-Ahead and Real-Time Energy and Operating Reserve Markets operate. If the 40% constraint on the amount of DRRs that clear Spinning Reserve in the Real-Time Market binds, then the MCP for cleared DRRs will differ from the MCP for cleared Generation Resources.

2.8.7.4.8 DRR Performance Assessment

Because it is impossible to directly measure the energy that a DRR resource would have consumed in the absence of the dispatch instruction to reduce load, its Demand reductions will be imputed through comparisons between the DRR's Consumption Baseline and its actual hourly metered consumption. Tariff Attachment TT provides detailed M&V criteria.

(i) Consumption Baseline

The selection, development and application of appropriate Consumption Baselines are part of the Measurement and Verification process.

The specific Baseline adopted depends, in part, on the specific product being delivered:

- Regulating Reserve service
- Energy
- Contingency Reserve service
- Capacity

2.8.7.4.9 Regulating Reserve Service

As stated earlier, only DRR-Type II resources are eligible to provide Regulating Reserve service. The Consumption Baseline used to estimate the amount of Regulating Reserve delivered by a DRR-Type II in any 5-minute Dispatch Interval uses the same measurement approach as used by generation resources providing this service.

2.8.7.4.10 Contingency Reserve Service

Contingency Reserve consists of Spinning Reserve and Supplemental Reserve. The Consumption Baselines are identical for both of these reserve products but are different for DRR- Type I and DRR-Type II resources.

A DRR-Type II providing Contingency Reserve service must provide telemetered demand data, scanned at 10-second intervals, to MISO. When a contingency event occurs, the DRR-Type II resource's Consumption Baseline is its telemetered average demand in the 10-second interval just prior to the start of the contingency event. The amount of contingency reserve deployed is then measured by the difference between its Consumption Baseline value and its telemetered demand in the 10-second interval occurring exactly 10 minutes after the start of the event.

The Consumption Baseline for a DRR-Type I resource is different because this resource is not required to provide telemetered data. The measurement and verification of Demand Response Type I Resource output is captured and calculated in the Demand Response Tool. The DRR- Type I Consumption Baseline is its metered demand for the 5-minute interval immediately preceding the start of the contingency event. The amount of contingency reserve deployed is then measured by the difference between its metered demand for the 5-minute interval ending 10 minutes after the start of the contingency event. BLDR Resources (section B.8.7.9 below) have a different measurement for assessing deployment. To the extent that an event starts or ends within a 5-minute interval reading, MISO requires that the Market Participant sponsoring the resource provide the actual load values for a DRR: (a) at the start of the event; (b) at 5 minutes into the event; and, (c) at 10 minutes into the event. The Market Participant should be prepared to provide supporting calculations based on the interval meter readings.

2.8.7.4.11 Energy

Four different generic Consumption Baselines exist for DRRs delivering the energy product:

- Metered Generation
- Calculated Baseline
- Direct Load Control
- Custom Baseline

2.8.7.4.12 Metered Generation

This type of Consumption Baseline only applies to behind-the-meter generation (btmg).For a btmg resource, the Consumption Baseline is the resource's actual metered generation over the hour beginning two hours prior to the hour in which the DRR is initially instructed to reduce load. The DRR's deemed demand reduction in response to a dispatch instruction in any hour is the difference between its metered output and its Consumption Baseline.

2.8.7.4.13 Calculated Baseline

This type of Consumption Baseline only applies to demand resources that actually reduce load. For a demand resource the Consumption Baseline is a profile of hourly demand (for the load behind the DRR asset) based on an averaged sample of historical data which may be adjusted for factors that reflect specific, on-the-day conditions, such as temperature. Unless the Market Participant sponsoring the DRR submits an alternative design for MISO approval, the default Consumption Baseline will be designed as follows:

- Separate hourly demand profiles will be determined for non-holiday weekdays and for weekends/holidays
- The “weekday” hourly profile will be based on the average of the ten (10), but not less than five (5), most recent weekdays that are not holidays or other non-standard “event” days
- The “weekend/holiday” hourly profile will be based on the average of the four (4), but not less than two (2), most recent weekend days or holidays that are not “event” days
- An “event” day is one during which there was, for the resource in question, a real- time energy or ancillary services dispatch, or a scheduled outage
- The maximum look-back window will be limited to 45 days
- If the 45-day window contains insufficient days to meet the minimum number of days described above, the profiles will be constructed based on the available days within the 45-day window that qualify, supplemented by the largest (MW) matching “event” day(s) values for that resource within that same window as necessary to obtain the minimum number of values.

The Market Participant sponsoring a DRR will have the option (at registration) to accept the unadjusted Consumption Baseline or to modify it by applying one of the following adjustment mechanisms:

2.8.7.4.14 Symmetric Multiplicative Adjustment (SMA)

- Adjusts each baseline hourly value (MW) during the event up or down by the ratio of the sum of hourly demands for the three hours beginning four hours prior to the event and (b) the sum of those same three hourly baseline demands.
- The adjustment is limited to a change in any individual baseline hour of plus or minus 20 percent.
- If multiple events occur during the same day, the SMA is calculated only for the first event, but applied to all events that day.

2.8.7.4.15 Weather Sensitive Adjustment (WSA)

- Adjusts each baseline hourly value (MW) up or down by a Weather Adjustment Factor
- The Weather Adjustment Factor is determined by a mathematical relationship derived through a regression analysis that considers the DRR load and historical hourly temperature data.

If the Market Participant sponsoring a DRR wishes to select either of the Adjustments described above or one of the non-default Consumption Baselines, the Market Participant must submit appropriate documentation to MISO for approval. Documentation must be credible and replicable analysis that supports the use of the applicable adjustment. The WSA baseline approach requires a complete, rigorous and defensible study or report that shows the complete statistical methods and analysis used to determine the Weather Adjustment Factor. The SMA baseline approach requires three (3) months of hourly data to be submitted with analysis used to justify the approach. Submitted documentation will be shared with the applicable LSE.

Example calculations of Calculated Baselines are provided in Appendix B: Examples for Existing Baseline Methods for Settlement and Examples of Baseline Adjustments. In addition, Calculated Baselines will not be adjusted for events beginning prior to 5:00 am Eastern Standard Time.

2.8.7.4.16 Direct Load Control

This type of baseline only applies to direct load control (DLC) programs consisting of many small, distributed resources that are not interval metered; consequently, only DRR-Type I resources are eligible.

A DLC Consumption Baseline will be statistically estimated from hourly metered demand data. MISO must approve the specific statistical methodology to be employed before the Market Participant can utilize a DLC Consumption Baseline. The input provided for the DLC Consumption Baseline becomes the performance (demand reduction) for that resource during an Event.

2.8.7.4.17 Custom Baseline

The Market Participant sponsoring a DRR may develop a custom Consumption Baseline if none of the three standard baselines described above would produce reasonable estimates of the resource’s demand reductions. MISO must approve of the specific methodology to be employed before the Market Participant can utilize such a baseline. For custom Consumption Baselines, the input provided becomes the Consumption Baseline that will be subtracted from metered amounts to determine performance (demand reduction).

2.8.7.4.18 Capacity

The Consumption Baseline employed to determine a DRR’s compliance with an instruction to reduce load during an emergency condition will be the same employed to estimate its delivered energy during normal conditions, i.e., those described in the preceding sections. The performance of Demand Resources in their role as Planning Resources is addressed in Section 2.8.7.6 *Demand Response as a Planning Resource.*

2.8.7.4.19 Metering

All MPs sponsoring DRRs are responsible for providing meter data appropriate to the services being provided. Revenue quality metering and telemetry equipment is required for DRR-Type II in order to support Regulation Reserve requirements. A DRR comprised of btmg must directly meter such generation. All DRRs must possess telemetry capabilities commensurate with the services to be provided. See MISO Tariff Module C section 38.2.5.e for additional detail on metering requirements. In addition, aggregated resources have specific metering requirements, detailed further in the section below on Meter Data Submission Types.

2.8.7.4.20 Meter Data File Formats

This section defines the details of the meter data that must be supplied by Market Participants for uploading settlement and compliance data into the Demand Response Tool system. Settlement and compliance data submitted by Market Participants will be available, through the system, to the LSE.

Two file formats are supported for submission of meter data: daily and interval. The daily file must always have 24 hour-ending (HE) values. The interval file must have sufficient data for the load reduction period and must match the hour, minute, and second of the required intervals. For an enrollment that contains more than one registered location, one set of entries should be provided for each registered location unless otherwise specified. Enrollments that contain virtual locations should provide one set of entries.

The file to be submitted must be of type “.xls”. The format of each file is described in the tables below.

Table 2-46 Daily File Format

Column Header Name	Type	Definition	Example
Enrollment	Text preceded "R" or "r"	DRT-generated ID for an Enrollment	R9999

Unique ID	Text	LBA account number assigned to the location	12345
Date	Date format mm/dd/yyyy	Date for which load is submitted	6/15/2021
UOM	Text	Type of meter data submitted	Compliance
UOM	Text	Units of Measurement for meter data, which must always be value-kW (represents integrated energy consumption over the interval)	kW
Type	Text	Type of meter data submitted	See meter data submission types listed below
HE1 through HE24	Integer	Meter value for each hour	100, 83, 89, 93, 99

Table 2-47 Interval File Format

Column Header Name	Type	Definition	Example
Enrollment	Text preceded "R" or "r"	DRT-generated ID for an Enrollment	R9999
Unique ID	Text preceded "R" or "r"	LBA account number assigned to the location	12345
Date	Date format mm/dd/yyyy hh:mm:ss	Beginning Date and Time for consumption over the interval	Compliance
Type	Text	Type of meter data submitted	Compliance
UOM	Text	Units of Measurement for meter data, which must always be value-kW (represents integrated energy consumption over the interval)	kW
Value	Integer	Meter value of the interval	100

2.8.7.4.21 Meter Data Submission Types

This section describes the meter data types that can be submitted for each type of Enrollment program. Various types of meter data are supported:

- Hourly Load: Hourly load data used for economic energy settlements. This information will be used to calculate the baseline and to determine the actual load during an economic Event. In the case of an aggregate enrollment, the load must be provided for each registered location of the aggregate.
- Compliance: Five (5) minute interval data used for compliance. This information will be used to calculate the baseline and to determine the actual load during an Ancillary Service Event. In the case of an aggregate enrollment, the load must be provided for each registered location of the aggregate. Used only for Interval Reading.

- HourlyCBL: Baselines are calculated outside of the DRT system (designated as “Manual” baseline on the enrollment) by the participant. The data submission also requires HourlyLoad for each HourlyCBL provided. In the case of an aggregate enrollment, the aggregate baseline should be provided as a submission for one of the registered locations.
- HourlyGen: Generation meter data will be used to determine the quantity of load reduction. In the case of an aggregate enrollment, there must be generation values for each registered location.
- HourlyDLC: Hourly load reduction based on a statistical sample approved by MISO, the number of active sites controlled, and weather conditions during the event

2.8.7.4.22 Market Settlements

The payments made for DRR performance are treated differently for those sponsored by LSEs serving them at retail than for those sponsored by ARCs. Each treatment is described in the following sections. The reader is cautioned that these descriptions are intended to provide settlement information only in the most conceptual terms. Please consult the BPM for Market Settlements (BPM-005), and associated attachments MS-OP-029 Market Settlements Calculation Guide and MS-OP-031 Post Operating Processor Calculation Guide, for the controlling language, descriptions, and formulas.

Settlements are further complicated by the fact that, while demand resources are compensated at LMP at all times, the cost allocation to pay for such services differ according to a comparison of the LMP with the Net Benefits Price Threshold (NBPT). The NBPT is a single value applicable for an entire month and is posted no later than the 15th of the prior month. See a tab under https://www.misoenergy.org/markets-and-operations/settlements/market-settlements/ for more information. When the LMP equals or exceeds the NPBT, charges for the energy provided are recovered from all other real-time “buyers”[150] within the Reserve Zones that benefit; when the LMP falls short of the NPBT, then the LSE serving the load behind the DRR is charged. This can lead to a variety of possible settlement conditions, and the primary ones are described in the following sections.

(i) LSE-Sponsored DRRs

Currently, most DRRs are sponsored by their LSE. In many states, this arrangement is dictated by state regulatory policy, commission rules, etc. In return for some incentives provided by the LSE, the retail customer may agree to not consume some of the energy it is entitled to purchase through its retail tariff. Since the DRR is sponsored by its LSE, there are fewer net settlement issues.

Day-Ahead Energy and Operating Reserve Market Settlements – For DRR Energy that is cleared into the Day-Ahead Energy and Operating Reserve Market, LMP will be paid to the MP with the DRR by purchasers in the day-ahead market. From a settlement perspective, DRR Energy is indistinguishable from energy provided by other resources.

For a DRR providing Operating Reserve, the MP will be credited for the Day-Ahead cleared Regulation Amounts, Spinning Reserve Amounts, and Supplemental Reserve Amounts multiplied by the applicable Day-Ahead hourly

[150] A real-time “buyer” is a Market Participant who purchases power in real-time without an offsetting purchase in the day-ahead market. For example, if an LSE schedules 100 MWh in day-ahead and consumes 105 MWh in real-time, it would be a real-time “buyer” of 5 MWh. Note that resources may also be real-time “buyers” in order to cover day-ahead positions not fully provided in real-time.

MCPs. MCP will be paid to the MP with the DRR by purchasers in the day-ahead market of the same Reserve Zone.

Real-Time Energy and Operating Reserve Market Settlements – In the Real-Time Energy and Operating Reserve Market, each LSE will be credited (or charged) for Energy based upon the incremental difference between its real-time energy transactions and its Day-Ahead scheduled energy transactions multiplied by the applicable Real-Time LMPs.

The LSE with the DRR will be unaffected when the LMP is below the NBPT, as the credit for the DRR reduction is exactly offset by an identical charge for that same amount of energy. From the LSE's viewpoint, it simply buys less net energy. For example, the LSE might schedule the purchase of 100 MWh (including the amount the DRR would have used) in day-ahead. In real-time, other usage is as-predicted, except for the DRR that "provides" 5 MWh (its load reduction). In this case, the LSE would simply receive payment for the 5 MWh (the net position). However, the MP with the DRR may also receive a "make whole" credit equal to that needed to fully recover the DRR's Production Cost if the LMP revenues do not recoup such costs and the DRR was committed by MISO through the SCUC process. Production Cost is the sum of the DRR's Shutdown Offer(s) plus the sum of its hourly Curtailment Offers plus the sum of its hourly Energy Offers.[151]

When the LMP equals or exceeds the NBPT, then the LSE will be credited for the full amount of the DRR reduction, while only being charged its pro-rata share across all buyers of the Reserve Zones that benefit. Ignoring this relatively small charge, the LSE effectively benefits in two ways: first, it will only be charged for the amount of energy actually consumed (95 MWh in the example above); second, it will in addition receive a credit for the 5 MWh reduction.

For Operating Reserve, the MP will be settled based upon the incremental difference between the DRR's Real-Time cleared Operating Reserve and its Day-Ahead scheduled Operating Reserve multiplied by the applicable RT MCPs. For DRRs not committed by MISO as part of its SCUC process, clearing Spin Reserve Service and deployed during a Contingency Reserve Deployment (CRD) event, credits will be entirely based on the applicable LMP at its CPNode. No make whole payments will be made for the MWs deployed during the dispatch intervals for the CRD event, regardless of hourly curtailment offers exceeding LMPs. By eliminating these make- whole payments, the Market Participant is allowed to add the expected cost of deployment in excess of expected Market revenues (net cost of deployment) to its Spinning Reserve Offer through a probabilistic cost adder by multiplying the Market Participant's expected possibility of deployment by the net cost of deployment. Incorporating deployment risk into the Spinning Reserve Offer will more accurately reflect the cost of selecting and deploying these resources during a CRD Event, which provides for better alignment with Market-based procurement of Spinning Reserves.

Additional charges related to system reliability, asset performance, Operating Reserve and the distribution of system losses are also settled in the Real-Time Energy and Operating Reserve Market.

[151] The terms, "ShutDown Offer" and "Curtailment Offer," when applied to a DRR mean, respectively, its price to be available to initiate load reduction when instructed, and its hourly price to maintain its load reduction, when instructed. Note that a DRR may incur multiple shutdown costs if it is released from commitment, then recommitted at a later time.

(ii) ARC-Sponsored DRRs

The settlement procedure for ARCs works in the same way described above for LSEs, except that the MP receiving payments or charges related to the DRR is the ARC, not the LSE. For certain market charges (e.g., Revenue Neutrality Uplift), the LSE's Real-Time energy purchases will be adjusted to reflect the RT energy reductions of the DRR.

2.8.7.5 Emergency Demand Response

The Emergency Demand Response Initiative is established in Schedule 30 of the Tariff and is designed to encourage Market Participants that have demand response capabilities available to them to offer those resources to MISO for use during North American Electric Reliability (NERC) Energy Emergency Alert 2 ("EEA2") or Energy Emergency Alert 3 ("EEA3") events. EDR resources are only dispatched during such events in response to dispatch instructions from MISO. LMRs are eligible to provide EDR service but must include a one-to-one relationship between the registration of an LMR and an EDR.

In addition to encouraging demand response participation, the EDR Initiative provides information MISO needs to commit and dispatch available EDR resources in economic merit order, i.e. by first curtailing those loads that customers value the least (or dispatching btmg with the lowest production costs) and progressively curtailing loads of increasing value (or btmg with increasing production costs) until the target level of demand reduction has been achieved. Such an efficient dispatch will minimize Market Participants' total costs of responding to Emergency Events.

When an EEA2 or EEA3 Event is imminent, MISO will develop a schedule of EDR Dispatch Instructions based on the information provided in the EDR offers for that Operating Day. After the Event has been declared, MISO will send EDR Dispatch Instructions to the affected MPs who will then be solely responsible for compliance using the EDR resources they offered.

After the Emergency Event ends, the responses of each EDR resource to its EDR Dispatch Instructions will be measured by comparing the resource's metered hourly loads (or net output of btmg) with its Consumption Baseline (or its generation baseline for btmg). The methodology used to determine Consumption Baselines is discussed below.

Each Market Participant will be compensated for the net demand response reductions its EDR resource delivered in response to their EDR Dispatch Instructions, but not for excess reductions, and will be exempt from related RSG Charges. Any Market Participant whose EDR resources do not fully comply with their respective EDR Dispatch Instructions will be assessed a penalty as described later.

MISO will recover the total payments made to Market Participants with dispatched EDRs in any Hour, net of any noncompliance penalties collected for that Hour, from the LSEs located in the Local Balancing Authority Area(s) where the Emergency Event(s) occurred in that Hour. Thus, these payments will be recovered from the parties that benefit most from the demand reductions that gave rise to the payments.

2.8.7.5.1 EDR Characteristics

A Market Participant may participate in the EDR Initiative if it controls a resource that can either: reduce Loads (either by reducing demand by a fixed number of MW or by curtailing use to a fixed target amount) in response to a request from MISO; or increase the outputs of btmg resources beyond what they would normally produce, in response to receiving EDR Dispatch Instructions from MISO. In addition, the Market Participant must be able to receive EDR Dispatch Instructions from MISO via an Extensible Markup Language (XML) interface, as more fully described in section. Lastly, a Market Participant must be able to provide integrated hourly energy consumption data on a CPNode basis.

2.8.7.5.2 EDR Offers

When an EDR resource is first registered, the Market Participant sponsoring it will submit a default EDR Offer, which will remain valid until updated. MPs may submit updated offers at any time prior to DA Market Close for application to the following Operating Day. All Offers are applicable to every hour of the day and will remain valid until modified or revoked by the Market Participant. Updated Offers may take the form of a declaration that the EDR resource will be unavailable for interruption until a new Offer is submitted.

Table 2-48presents the information that a valid EDR Offer must contain. If any of these data elements are missing in the Offer submittal, MISO will substitute the corresponding data elements from the previous Offer.

Notes for the table below:

- Note 1. Reductions must be expressed in increments of 0.1 MWh per hour.
- Note 2. Enter either Maximum Demand Reduction or Reduction to Firm Load Level – not both.
- Note 3. Curtailment Price cannot exceed $3,500 per MWh.

Table 2-48 EDR Resource Offer Data

Data Element	Unit	Note
Maximum Demand Reduction	MW	1,2
Reduction to Firm Load Level	MW	1,2
Curtailment Price	$/MWh	3
Shutdown Cost	$	
Advance Notification	hh:mm	
Interval when Reduction is Available	hh:mm to hh:mm	
Minimum Down Time	hh:mm	
Maximum Down Time	hh:mm	
Daily Availability	Yes/No	
Any Temporary Limitations	Text Field	

EDR resources are not subject to the usual must-offer obligations because participation in the EDR Initiative is voluntary. However, any EDR resources that also qualify as LMRs under Module E-1 of the Tariff will have a must-offer obligation during MISO-declared Emergencies and thus cannot declare the portion of load that is an LMR as unavailable for curtailment.

2.8.7.5.3 Commitment and Dispatch

On a day when an Energy Emergency Alert (EEA 2) is anticipated, MISO will use the data in EDR Offers valid for that day to develop EDR Dispatch Instructions that minimize customers' total collective costs of achieving the load reductions needed to offset the supply resource shortfall. Typically, this will produce EDR Dispatch Instructions that call on EDR resources in order of their increasing EDR Production Costs, which consists of the EDR resource's Curtailment Cost (dispatch price multiplied by expected MWh curtailed) and its Shutdown Cost (or its one-time Startup Cost for btmg). However, because the EDR curtailment schedules are based on constrained optimizations that account for EDR resource inflexibilities and other operating constraints, including their location on the transmission grid, they may not reflect simple, monotonic rankings of EDR Production Costs.

Each Dispatch Instruction will include the following information:

- Hour the demand reduction is to commence
- Amount of demand reduction or the firm load level to be achieved
- Schedule of incremental changes to the reduction level, if any
- Duration of each demand reduction level

Dispatch Instructions and all other communications between MISO and Market Participants with EDRs will be via XML interface.

2.8.7.5.4 EDR Performance Assessment

As with other forms of demand response, an EDR resource's demand reductions must necessarily be imputed through comparisons between its metered hourly consumption and its Consumption Baseline.

2.8.7.5.5 Consumption Baseline

The Consumption Baseline is the actual usage of the facility containing the EDR resource in the Hour prior to the start of the instructed demand reduction.

For EDR resources that are under direct load control, the Market Participant must provide: a description of the direct Load control system, a description of Load Research data used in the measurement and verification analysis, a description of the methodology used to produce the estimate, and a description of all source information for the variables used in the analysis.

2.8.7.5.6 Metering

All MPs sponsoring EDR resources are responsible for providing meter data for the Hour prior to the start of the reduction and for every Hour in which the reduction occurred. This can be done through a third-party Meter Data and Management Agent (MDMA). MDMAs must provide meter data to MISO prior to noon EST of the 53rd day after the Operating Date. Along with a record of its meter readings, Market Participants utilizing on-site generation must also provide a written statement from the Market Participant certifying that the Demand reductions were made in response to MISO's EDR Dispatch Instructions and that they would not otherwise have occurred.

2.8.7.5.7 EDR Market Settlement

Market Participants with a registered EDR are compensated at the higher of the revenues resulting from hourly LMPs (i.e., applying the hourly Real-Time LMPs at each EDR resource's CPNode to the resource's instructed hourly demand reductions), or the EDR resource's Production Costs for the total period of reduction. EDR Production Costs are defined as the shutdown cost plus the lesser of the amount of hourly Demand reduction or the hourly Dispatch

Instruction, multiplied by the EDR Curtailment Price applicable to the period of actual Demand reduction.

To qualify for compensation an EDR resource must comply with MISO's EDR Dispatch Instructions. If an EDR resource reduces its Demand by an amount that exceeds the reduction level specified in the EDR Dispatch

Instruction, it will only be compensated for the amount specified in the MISO Dispatch Instruction. However, the MP will not be subjected to RSG charges for its excessive reductions.

Payments made in excess of market revenue will be funded on pro rata basis via Load Ratio Share to Market Participants in the Local Balancing Authority Area(s) where the Emergency event occurred.

Meter data is required within 53 days following the Operating Date of the Emergency event. Settlement will occur on the relevant applicable settlement statement after submission of meter data.

2.8.7.5.8 Penalty for Underperformance

An EDR resource that reduces Demand in any Hour by less than the amount specified in the EDR Dispatch Instruction will be fully compensated if the reduction is not less than the Demand Reduction Tolerance level (which is set equal to 95 percent of the EDR Dispatch Instruction amount) for that Hour. An EDR resource that reduces demand by less than the Demand Reduction Tolerance level will be charged an amount equal to the Demand Reduction Shortfall multiplied by the Real-Time LMP of the load zone in which the EDR resource is located. The Demand Reduction Shortfall is equal to the Demand Reduction Tolerance minus the actual Demand reduction, or zero, whichever amount is greater. Failure to reduce demand at a level higher than the Demand Reduction Tolerance level will also result in a loss of guaranteed cost recovery.

Revenue collected from the underperformance penalty will be distributed pro rata via Load Ratio Share to Market Participants in the Balancing Authority Area(s) where the Emergency event occurred.

2.8.7.6 Demand Response as a Planning Resource

Module E-1 of the Tariff defines a Load Modifying Resource (LMR) as a Demand Resource or BTMG that satisfies the requirements for being a Planning Resource. An LMR is not required to be a Network Resource[152]. An LMR need only be available for interruption during Emergency Events. The Emergency Operating Procedures (e.g., SO-P-EOP00-002 and SO-P-EOP-00-004) describe how and when LMRs will be called during an Emergency Event.

LMRs may also qualify as Emergency Demand Response (EDR) resources by meeting the requirements in Schedule 30 of the Tariff. LMRs may also participate in Planning Resource Auctions as briefly described later in this BPM. More detailed information regarding LMR participation under Module E-1 is contained in BPM-011 Resource Adequacy.

Each LMR must be registered, reviewed, and approved annually by MISO in advance of receiving capacity accreditation as a Planning Resource. Only Market Participants may register LMR and this process is completed by accessing the Module E Capacity Tracking (MECT) tool through the secure Market Portal.

2.8.7.6.1 Utilization of LMR Capacity

LMR capacity has value because it can be used to meet the Planning Reserve Margin Requirement (PRMR) of an LSE. The Market Participant registering the LMR (either the LSE or an ARC) may choose to treat an LMR as a Planning Resource for conversion into Zonal Resource Credits (ZRCs). When such treatment is requested (and

[152] Excess BTMG – the unforced capacity of an LMR BTMG in excess of an LSE's PRMR, can participate in the PRA as long as it demonstrates deliverability since it represents a net injection onto the transmission system Deliverability can be demonstrated by being granted commensurate Transmission Service or Interconnection Service. LMR DRs have no deliverability requirement. See the BPM for Resource Adequacy.

accepted) the LMR's accredited capacity will be entered into the MECT and the Market Participant can use these ZRCs to meet its PRMR, offer them into the PRA or trade these ZRCs with other MPs.

2.8.7.6.2 LMR Performance Assessment

Following an Emergency Event in which a LMR was instructed to curtail its load, the Market Participant that registered the LMR will collect data needed to perform the calculations comparing the LMR's actual load with a Consumption Baseline adopted at the time of registration and subsequently updated as needed. The Market Participant will certify the results of this analysis and submit them to MISO via the Demand Response Tool (DRT). MISO will use these results to determine if the LMR reduced by the targeted MW level (or to a specified firm service level if applicable), when called upon to do so by MISO. Additional details are available in Tariff Attachment TT. Each LMR will be evaluated on its individual performance and not in aggregate across a Market Participant's portfolio.

(i) Consumption Baseline

The Consumption Baseline for a DR will be the expected value of the DR's average hourly load, rounded to the nearest kWh, for each of the 24 hours in a day. A Consumption Baseline is required for each DR that is included in an LSE's Resource Plan. A default Consumption Baseline will be calculated for each hour in a day, as being the simple averages of hourly meter data from the ten business days prior to an Emergency Event. See attachment TT of the Tariff for additional details. The default baseline procedure will be used unless the Market Participant proposes an alternative Consumption Baseline procedure at the time it registers the DR, and it is accepted by MISO. For an LMR that agrees to reduce load to a specified level, its demand reduction will be the difference between its Consumption Baseline and the specified level.

Following an Emergency Event in which the LMR resource was deployed, the Market Participant that registered it shall collect and provide the hourly meter data to calculate the resource's Consumption Baseline in the Demand Response Tool and submit them to MISO within 53 days from the time the resource was deployed. MISO will review these metering data to verify that the Demand Resource reduced load by the targeted MW level, or to a specified firm service level, when called upon by MISO.

(ii) Metering

BTMG consisting of one or more generating units that have been identified by MISO must have metering equipment for operational security purposes. BTMG consisting of multiple generating units at a single site that have been identified by MISO must have metering equipment but may be metered as a single unit, in which case they will be treated as a single unit for purposes of LMR performance evaluation.

2.8.7.6.3 LMR Settlements

LMRs interact with MISO Settlements process in two ways:

- MPs trading ZRCs associated with an LMR that clears in a Planning Resource Auction are paid or charged based on the market clearing prices as established in the auction.
- If an LMR does not meet the Measurement and Verification protocol selected during registration (reduced by the targeted MW level or to a specified firm service level if applicable) during Emergency Events, the Market Participant that registered it may be penalized.

(i) Planning Resource Auction Settlements

MISO will settle each Planning Resource Auction (PRA) by charging the applicable Auction Clearing Price (ACP) for that Planning Year to MPs with PRMR and crediting the applicable ACP to MPs with cleared ZRC offers. The invoice credit will be available through the Market Portal daily during the Planning Year.

(ii) Penalty for Nonperformance

Unless the LMR is unavailable as the result of maintenance or for reasons of Force Majeure, the Market Participant representing the LMR will be penalized when the LMR fails to perform as instructed during an Emergency Event. See Tariff, section 69A.3.9 of Module E-1. However, no penalties will be assessed if an LMR is unavailable for interruption due to its Load being off the Transmission System for external reasons, or if the targeted Demand reduction had already been achieved for other reasons (e.g., economic considerations or local reliability concerns). MISO will credit the proceeds of LMR penalties to only those MPs representing the LSEs in the LBA area(s) that experienced the Emergency that triggered the use of an LMR. Such revenues shall be distributed on a Load Ratio Share basis. An LMR, unavailable or unresponsive for reasons other than exempted by MISO, could be disqualified from participation for the rest of the Planning Year. Disqualification results in removal of ACP payments. In addition, the MP will be charged the ACP for the remainder of the Planning Year, and proceeds will be redistributed pro rata based on the LSE's PRMR in the LRZ. Additional details can be found in Section 69A.3.9 of Module E-1.

2.8.7.6.4 LMRs that dual-register as EDRs

Resources that register both as an LMR and an EDR have the following characteristics: a one-to- one relationship must occur between the registration of an LMR and an EDR; and the exact same end-use accounts must make up the defined LMR and EDR. At the current time, separate registration processes are required to dual-register the resource. All the requirements and characteristics specified in Section 2.8.7.3.1 *Registration as a Market Participant* above under LMRs must be met. For example, the joint LMR/EDR resource must meet the specified availability and notification times, and minimum run times as registered under LMRs in Section B.8.7.3.1. ARCs that dual-register resources must meet the requirements, as specified in Section *2.8.7.3.2 Registration as an Aggregator of Retail Customers (ARCs)* above, separately for both the EDR and LMR registrations. In addition, by registering as an EDR, the Market Participant can submit EDR Offers and must be able to receive EDR Dispatch Instructions via XML. Commitment and Dispatch will occur as specified as part of the SO-P-EOP00-002 and SO-P-EOP-00-004Emergency process. There can be only one selected Consumption Baseline for a dual-registered LMR/EDR resource. Payment for performance is based as specified under the EDR Initiative; any shortfall charges are based on the LMR paradigm. LMRs should not report their availability in the DSRI for days when they have active EDR Offers. It is the responsibility of the Market Participant to ensure there is no double counting of MWs offered across the dual registration types. Double counted MWs may be subject to underperformance penalties.

2.8.7.7 Resource Testing

To participate in MISO markets each resource must demonstrate its ability to interrupt load within a prescribed time limit after being instructed to do so. The prescribed time limit will depend on the particular service the resource is being qualified to provide.

2.8.7.7.1 DRR-Type I

DRR resources must provide information similar to what is provided by generating resources, including submission of data through the GADS or DADS, as appropriate. Annual testing and verification are required. Details may be found in the BPM for Energy and Operating Reserve Markets (BPM-002).

2.8.7.7.2 DRR-Type II

See DRR Type I above.

2.8.7.7.3 EDR Resources

There are no ex-ante resource testing requirements applicable to EDR resources unless the resource is dual registered as an LMR; such resources are measured and verified during the Emergency Events to which they respond.

2.8.7.7.4 Load Modifying Resources

(i) Demand Resources

Market Participants with Demand Resources should demonstrate a real power test for capacity accreditation. The real power test of a Demand Resource may be from a MISO called event or a self-scheduled implementation in accordance with section 4.2.9.8 of BPM-011 Resource Adequacy. If a Demand Resource test is not performed for accreditation, additional options outlined in BPM-011 Resource Adequacy section 4.2.9 may be utilized.

(ii) BTMG

BTMG capacity accreditation generally follows the same documentation requirements of generating resources. BTMG greater than 10 MW must submit performance and event data to GADS as well as an annual Generation Verification Test Capacity (GVTC). BTMG below this limit are only required to submit an annual GVTC and can accept the class average EFORd assigned to the unit type by MISO. Additional details regarding BTMG testing requirements may be found in BPM-011 Resource Adequacy Section 4.2.8.

2.8.7.8 Credit Requirements

MISO's Credit Policy requires all Market Participants to have an approved credit application and an established Total Credit Limit with MISO Credit Department. Attachment L of the Tariff describes in detail how MISO will determine a Market Participant's Total Credit Limit requirement as well as the procedures it will follow in evaluating the Market Participant's creditworthiness. It also contains all of the requisite forms and describes the procedures for a Market Participant to follow to establish its Total Credit Limit. Attachment L of the Tariff is available on MISO website: https://docs.misoenergy.org/legalcontent/Attachment_L_-_Credit_Policy.pdf

The remainder of this section briefly describes how MISO will determine the increase to a Market Participant's Total Credit Limit requirement contributed by the product offered by a given demand resource.

2.8.7.8.1 Economic Energy

As a supplier, the credit requirements for a new MP with a DRR are based on two factors: the amount of energy (MWh) that can be produced from the resource in an hour, and the historical average day-ahead LMP for the appropriate CP Node. The formula used to determine the credit requirement (in dollars) is the product of: (a) the maximum MWh value just described; (b) 600 hours, times; (c) the historical average LMP for the preceding three-month period, times; (d) 5%. For example, if a given DRR resource could produce 1 MWh in an hour and the historical average LMP was $30/MWh, the credit requirement would be equal to 1 × 600 × 30 × 5%, or $900. This credit requirement is reduced for ARCs as specified in Attachment L of the Tariff. See Attachment L for credit conditions for existing certified MPs.E

2.8.7.8.2 Operating Reserve Services

Credit requirements for Operating Reserve Services are included in the credit requirements for Economic Energy. No additional credit requirements are applicable.

2.8.7.8.3 Emergency Demand Response

There are no additional credit requirements related to the provision of EDR service. Please see Attachment L of the Tariff for general credit requirements.

2.8.7.8.4 Planning Resources

There are no additional credit requirements related to the offer of Planning Resources (LMR) unless a Market Participant with a Demand Resource has waived its obligation to conduct a real power test per Tariff Section 69.A.3.5.j as described above in Section 2.8.7.3.1 *Registration as a Market Participant* of this document.

2.8.8 BATCH LOAD DEMAND RESPONSE

This section describes the business rules governing Contingency Reserves provided by Batch- Load Demand Response (BLDR) resources.

A BLDR resource is a load caused by a cyclical production process. During most of its duty cycle the load consumes energy at some nominal level but periodically reduces load for a short interval, typically less than 10 minutes. The following figure illustrates the actual consumption pattern of one such load in the MISO footprint.

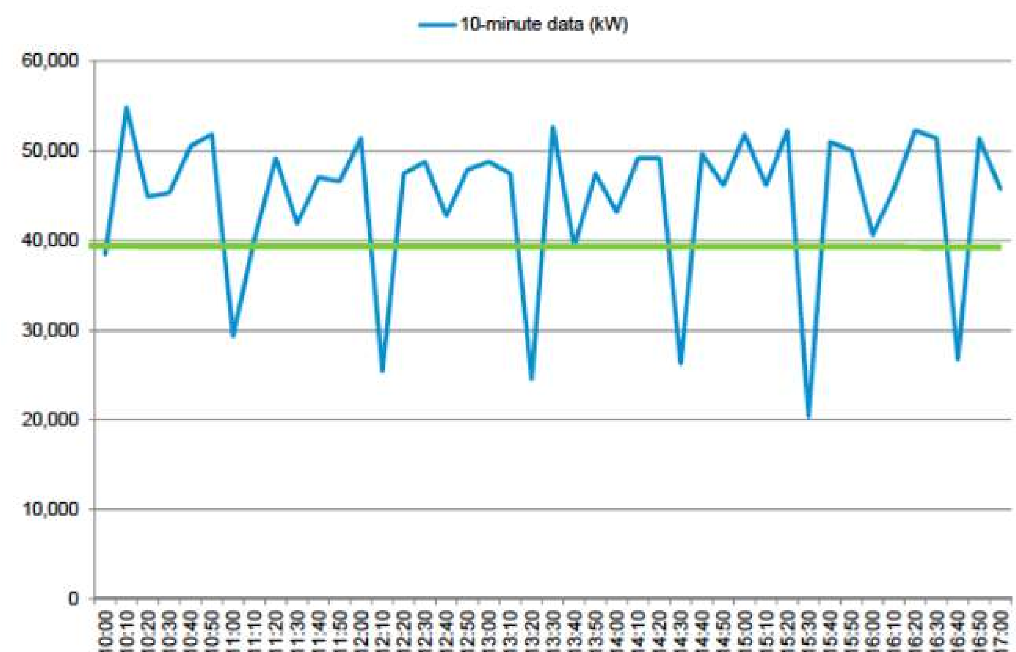

Figure 2-22 Batch Load Consumption Pattern of One MISO Market Participant

The MISO Tariff currently requires Contingency Reserve products to fully deploy their cleared Contingency Reserve within the 10-minute period (Contingency Reserve Deployment Period) following receipt of a MISO deployment instruction, as prescribed in ERO Standard BAL 002-0. In contrast with other Contingency Reserve assets, a BLDR resource may be capable of releasing little or no energy within the mandatory Contingency Reserve Deployment Period if it receives the MISO dispatch instruction while its load is not at the "top" of the cycle, as illustrated in the figure below. Nonetheless, by remaining at the bottom of its cycle, the BLDR resource helps MISO in meeting the BAL standard by not exacerbating the ACE deviation, which it would do if it resumed operations of its batch load process. This latter effect must be weighed when evaluating the resource that, most of the time, could release significant amounts of energy to assist MISO in responding to a contingency event.

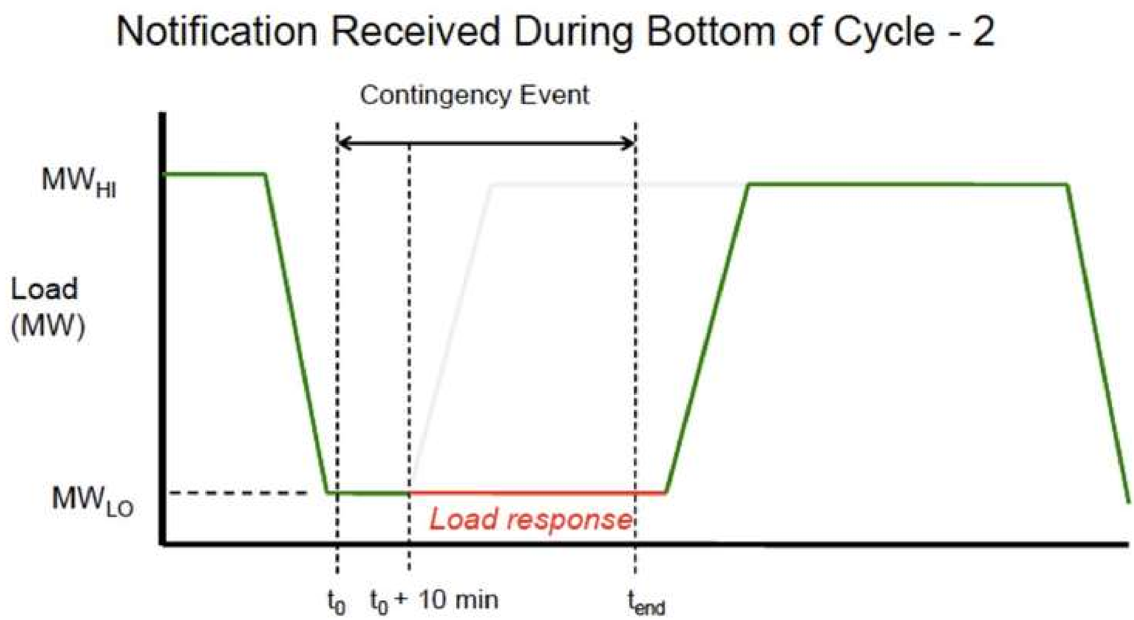

Figure 2-23 Operating Reserve Energy to Deliver

Because it is impossible to know *ex ante* where a BLDR resource will be when instructed to deploy its Contingency Reserve, the best that can be done is to credit the resource with the expected value of the amount of that reserve. Each BLDR resource will be responsible for estimating this expected value at the start of each day during which it offers Contingency Reserve into the market and will offer no more Contingency Reserve than that expected value amount.

When a BLDR resource is instructed to deploy its Contingency Reserve, it is also obligated to maintain its energy reduction until the contingency event ends.

The inability of any single BLDR resource to fully deliver the expected value of its Contingency Reserve will not impact MISO if the resource represents a small portion of total cleared Contingency Reserve. This will also be true if no single BLDR resource represents a large portion of a portfolio of BLDR resources whose duty cycles are relatively uncorrelated in time, because the diversification effect will drive the portfolio's performance toward deploying an amount of Contingency Reserve that approximates the expected value of its total cleared Contingency Reserve.

2.8.8.1.1 Registration and Scheduling Contingency Reserve from BLDR Resources

Market Participants desiring to register a DRR Type I as a BLDR resource will have to so indicate during the registration process. Qualifications for using this M&V approach include submittal of the most recent three months of 5’ interval data and the asset being at its low duty cycle no longer than 10’. MISO will review the submitted data to decide whether to approve this M&V method. This baseline method is only available for contingency reserve assessment and not energy. After registration, the mP must retain a rolling three months of 5’ interval data that MISO can audit at any time.

At the current time, there is a 40% cap on using Demand Response Resources for provision of Spinning Reserve service. BLDR procurement will be included in the 40% cap imposed in the Spinning Reserve market. If the total amount of Operating Reserve provided by all BLDR resources is small (e.g., less than 10 percent) the selection of BLDR is a non-issue. As Table 2-35 shows, a large BLDR load might offer about 30 MW of Spinning Reserve to the market, which is well below the 40 percent limit; however, if many BLDR loads begin offering Spinning Reserve, the total amount could easily get capped at the 40 percent limit.

On the positive side, aggregating BLDR loads will have the combined effect of diversifying away the likelihood that all, or most, loads are at the bottom of their cycles when called to deploy their energy. As the number of BLDR loads increases their combined response to a contingency event, the combined response ability will approach a normal distribution whose expected value of the total Contingency Reserve that will be deployed is equal to the sum of the expected values of the Contingency Reserve that will be deployed by each BLDR resource.

MISO’s current Tariff allows the system operators to adjust the amounts of Contingency Reserve they procure based on contemporaneous system conditions. The current business rule is to place BLDR procurement under the 40% cap that currently exists in the spin market. The Supplemental Reserve market has no such caps.

2.8.8.1.2 Measurement and Verification of BLDR Contingency Reserve

BLDR resources are a special category of DRR Type – I resources. The current Tariff requires all DRR Type-I resources to provide 5-minute interval data to MISO no later than five (5) days after the end of the contingency event. This data must span the period starting five (5) minutes prior to when the contingency event began and ending at least 60 minutes later.

The DRR-Type I Consumption Baseline is its metered demand for the five (5)-minute interval immediately preceding the start of the contingency event. The amount of Contingency Reserve deployed is then measured by the difference between its Consumption Baseline value and its metered demand for the five (5)-minute interval ending ten (10) minutes after the start of the contingency event. If this M&V methodology is applied to a BLDR resource that is at the bottom of its duty cycle when it receives MISO’s deployment instruction, the resource will be in noncompliance and will have little incentive to suspend its cyclical production process. Suspending production provides value to MISO because it assists in controlling the ACE. In addition, it could also bring about an earlier end to the contingency event. For these reasons, a separate M&V methodology is needed for BLDR resources.

If the resource ramps down to its minimum demand and remains at that level until the end of the contingency event, it will be in full compliance. To make this assessment, MISO requires a snapshot of the BLDR resource’s normal consumption pattern. In such cases, the resource’s eligible amount of deployed Contingency Reserve will be the smaller of the difference between the resource's demand for the five (5)-minute interval immediately preceding the end of the contingency event and: 1) its demand for the five (5)-minute interval beginning ten (10) minutes immediately following the end of the event; or, 2) the 50% trimmed mean of the five (5)-minute intervals for the three (3) hours immediately following the Contingency Reserve Deployment Period.

The Market Participant sponsoring a BLDR may develop a custom Consumption Baseline if the above approach would not produce reasonable estimates of the resource's demand reductions.

Example:

Demand 10-min. following end of dispatch =	43 MW
Demand during dispatch =	30 MW
Calculated response =	43 – 30 = 13 MW
50% trimmed mean for the next 3 hours =	45 MW
Demand during dispatch =	30 MW
Calculated response =	45 – 30 = 15 MW

13 MW < 15 MW; 13 MW response

How this would work: The "trimmed" mean removes X% of the largest and smallest values from the data. This has the effect of reducing the impact of "outliers". As part of an M&V protocol, this technique would remove the cyclic lows and any dispatch down. This method would also remove the highest values that are probably not reflective of typical demand. See illustration below.

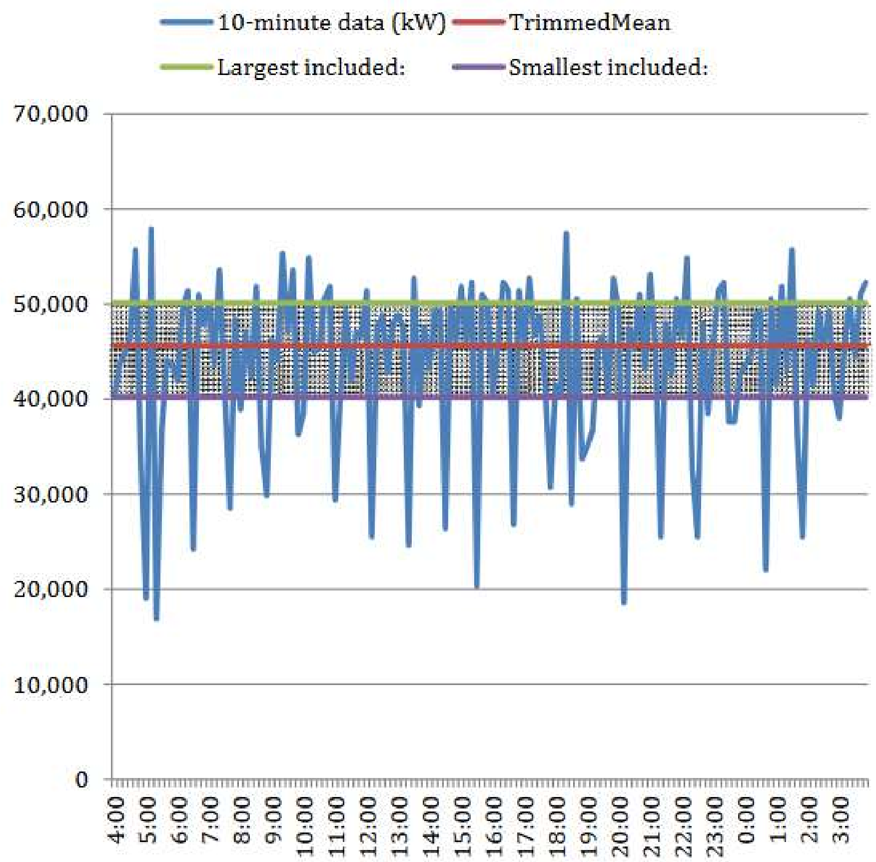

Figure 2-24 Consumption Trimming

2.8.8.1.3 Compensating BLDR Resources for Contingency Reserve

A BLDR resource will receive compensation comparable to that which a DRR Type – I resource would receive for its capacity and energy. The current Tariff compensates a DRR Type – I resource for being available to provide Contingency Reserve through capacity payments. In addition, when the resource reduces its load as instructed, it received energy payments for its foregone energy. This same compensation will apply to BLDR resources.

2.8.8.1.4 Underperformance Penalties

The underperformance penalties that currently apply to DRR Type – I resources will apply to BLDR resources when they are out of compliance. However, as stated earlier, a BLDR resource cannot control where it will be in the duty cycle when it receives a deployment instruction, so it should not be deemed out of compliance based solely on the amount of Contingency Reserve capacity it actually delivers. The resource will only be out of compliance if it fails to take action to shut down or resumes its batch load operation during the Contingency Reserve Deployment Period after receiving a deployment instruction. Table 2-35 and Figure 2-19 illustrate this situation. Although the resource deploys no Contingency Reserve energy, it is in compliance because none was available for deployment when the resource received the instruction to deploy. The resource would also be in compliance if some Contingency Reserve was available and all of that was deployed.

2.9 Protocols and Guidance for Establishing Quality Assurance / Quality Control for Programs

Continuous improvement in the operation of energy efficiency and demand response programs requires that procedures for quality assurance and quality control be put in place and applied continuously in real time.

- Quality Assurance (QA) are standards to promote consistency and minimize errors are developed and applied during the planning and design of a program.
- Quality Control (QC) activities are conducted continuously in real time to ensure that programs are being implemented and operated according to set quality standards.

2.9.1 QA/QC PROTOCOL 1: APPROACH TO QUALITY ASSURANCE

Quality Assurance activities occur throughout a program's lifecycle to ensure that program processes are aligned with objectives, that risk is avoided, and that efficiency is being promoted. QA activities are used to ensure that program rules and requirements are documented and current, that participating contractors and trade allies are properly licensed and trained and maintain high quality standards in interactions with customers, and that data are accurate and sufficient for analyzing energy savings analysis.

Examples of QA activities include the following:

- Developing program logic models and process maps that document the goals, processes, and expected outcomes associated with key activities in each program;
- Implementing training protocols that describe training procedures and requirements for key program stakeholders, such as CSPs and trade allies;
- Applying rigorous screening and qualifying protocols to CSPs, trade allies, and field staff that interact directly with customers;
- Documenting data collection protocols, including data and customer information needed to track activities and calculate savings for each program; and
- Summarizing CSPs' gross energy savings calculation methods that are reported at the measure or project level to support consistency and accuracy across programs.

Information on processes used with a program can be organized through preparation of a "program logic model" . In broadest terms, a logic model shows how resources are used in activities to produce outputs that yield outcomes. The logic model for a program should provide a clear description of the processes used with that program to provide energy efficiency services and / or products to customers participating in the program. Essentially, developing the logic model should show what the processes for a program are supposed to do, with whom and why. In particular, the program logic model should:

- Identify the group(s) involved with the program;
- Identify the resources being allocated to the program;
- Describe those activities or action steps that are being used to achieve outcomes;
- Define the outcomes or objectives for a program, where outcomes are those changes or benefits that result from activities; and
- Determine whether the objectives are being achieved.

While a program logic model shows the structure and practices desired and expected for a program, quality assurance procedures are used to identify and identify standards that eliminate variations or defects in program processes that may cause appropriate quality to not be achieved. A framework for assuring that quality requirements are being met is provided by the Plan-Do-Check-Act cycle paradigm that was popularized by Deming and that is the basis for ISO 9001: 2015, the international standard that specifies requirements for a quality management system. As summarized in Table 2-49, the PDCA cycle provides a four-step method for continuous quality improvement.

Table 2-49 Steps in PDCA Cycle for Quality Assurance

Step	**Activity**
Plan	Establish objectives for quality and determine processes or changes in processes that are required to deliver desired quality. Determine specific levels of quality or measurable results to be achieved
Do	Develop and test process and / or changes in processes.
Check	Monitor and evaluate processes or changes to determine whether quality is meeting predetermined objectives. To extent possible, use objective measurements or tests to determine whether quality goals are being met, rather than using subjective evaluation of quality.
Act	Implement actions that are necessary to achieve desired improvements in quality
If appropriate, repeat, beginning with new objectives being planned.	

Guidance on using the ISO 9001: 2015 standard and the PDCA cycle to develop and implement an effective quality assurance and management system can be found in a handbook published in 2016 by the International Standards Organization: ISO 9001: 2015 for Small Enterprises: What to do?

2.9.2 QA/QC PROTOCOL 2: PROCEDURES FOR QUALITY CONTROL

Quality control procedures should be applied continuously in real time to ensure that program activities adhere to the standards set through the QA work and conform to performance expectations at the program and portfolio levels. QC activities address operational procedures, data and records, and measure installation. Examples of QC procedures include the following:

- Ongoing tracking of program activities and costs.
- Reviewing all data and records to confirm that the proper data are collected consistently, resources are allocated appropriately, and program performance can be measured accurately.
- Conducting follow-up calls to participants to evaluate their satisfaction with the rendered services and to identify opportunities to improve the effectiveness of energy efficiency programs.

As shown in Table 2-50, quality control activities occur during both pre-implementation and post-implementation phases of a program.

Table 2-50 Quality Control Activities During Pre-Implementation and Post-Implementation Phases of a Program

Quality Control during Pre-Implementation
Documentation review: Program documentation should be examined to ensure that it is complete and that it provides all essential information for achieving and verifying savings.
Site pre-inspection and interviews: Site inspections may be conducted to verify preexisting conditions, quantities of measures, key operating parameters, equipment performance, and baseline assumptions in the measure documentation.
Measured data collection: Addresses uncertainties regarding performance of measures or to confirm validity of assumptions used in the baseline analysis. May include spot measurements, data trending (via data loggers or building control systems), or other data collection conducted before measures are implemented.
Quality Control during Post-Implementation
Documentation retention: Program-required documentation should be reviewed to ensure completeness and accuracy. All energy savings-related documentation should be retained for future savings validation or evaluation efforts.
Site post-inspection and interviews: Site inspections and interviews may be conducted to verify that measures were installed and commissioned and operate as intended.
Measured data collection: Data may be collected post-implementation to verify key operating parameters of measures or to meet requirements of an M&V plan.

Evaluation of programs by a TPE can also contribute to quality control of a program. In particular, quality control can be facilitated by having implementation and EM&V contractors coordinate and integrate their activities. Examples of how M&V activities can be coordinated and integrated with implementation activities include the following:

- Pre-installation review: This involves implementation and M&V contractor teams performing pre-installation review of measures and projects prior to a utility reserving incentive funding.
- Project-Specific M&V Plans: This involves implementation and M&V contractor teams coordinating to provide project-specific M&V plans for select projects to ensure the implementation contractor has a full understanding of the M&V approach for these projects prior to the projects being completed and incentivized.
- Coordinated joint site visits: This involves implementation and M&V contractor teams coordinating to conduct joint site visits for select projects. Joint visits reduce the impact on customers and allow data to be collected concurrently, reducing conflicting information collected during separate site visits.
- Project-Specific M&V Reports: This involves sharing project-specific M&V reports with implementation contractors prior to final program level analysis.

Sharing analysis files, energy models, engineering spreadsheets, etc. maintains transparency and allows all calculations used in determining evaluated verified energy savings to be reviewed by all parties.

The TPE should also conduct quality control for the evaluation work. Examples of areas where quality control should be exercised for evaluation work include the following.

Quality control assessment of evaluation plans:

- Analytical methods used to estimate savings
- Baseline determination
- Researchable questions
- Sampling approaches and segmentation or stratification (if appropriate)
- Data collection instruments and topics
- Mapping inputs and outputs for computation of effects
- Logical narrative

Quality control assessment of data procurement:

- Review of options for real time data collection
- Use of appropriate data collection procedures for sampling, collection, processing, attrition, bias, etc.
- How to best use data tracking systems to serve needs of both program implementation and evaluation

Quality control of evaluation reporting:

- Consistency of reporting with the corresponding plan and with best practices
- Cogency and clarity of reporting documentation
- Critical assessment of conclusions and recommendations
- Thoroughness of documentation of methods and results in reports

2.10 Protocol and Guidance for Updating the TRM

This protocol addresses the updating of the Technical Reference Manual (TRM). The protocol provides for periodically reviewing and, if appropriate, updating the content of the TRM. For many measures, updating may need to occur only when codes and standards affecting the specific measure change. Areas to focus on for major updating include:

- Making changes to existing measures, data, and calculations when significant changes are justified, typically because of changing baselines or availability of more current, applicable evaluation studies for updating values.
- Including new measures that are determined to be priorities in the TRM.

The focus of the updating should be on areas of high impact in the Energy Smart portfolio (e.g., duct sealing) and of potential future high impact measures (e.g., ductless mini-split HVAC systems).

A study of an existing or new measure is warranted when the following guidelines are met.

- Measures should be flagged for further review if they exceed 1% of portfolio savings. In such instances, it should be determined whether:
- Primary data have been collected in Energy Smart evaluations to support the deemed savings.
- The data is sufficiently recent to support its continued use.
- If data collection to support a deemed savings revision is cost-effective or cost-feasible, given the implementation and EM&V budgets for Energy Smart programs.
- Measures that are not over the high-impact threshold should be considered for impact or market assessment studies if:
- Stakeholders (the Council and their Advisors, ENO, implementers, interveners, the EM&V contractor, and/or other appropriate parties) conclude a measure is of strategic importance to future program implementation efforts; or
- A measure is high-impact within an important market sub-segment (such as low-income multifamily or municipal government).

Future implementation of Energy Smart programs may include measures that are not in the current version on the TRM. The treatment of these measures in the implementation and evaluation process will differ situationally.

- Many measures in the commercial and industrial segment are custom measures for which deemed savings are inappropriate. These measures will be validated individually based on IPMVP protocols.
- Direct load control (DLC) or load management (LM) programs curtail peak loads through installation of control devices on specific systems (DLC) or through voluntary self-curtailment (LM). These programs are not appropriate for inclusion in a TRM and should have their performance validated annually.

The TRM should be updated each year through a two-stage process.

- In the first quarter of each calendar year, a technical forum will be held in which stakeholders may suggest measure additions or updates. This will inform the scope of TRM additions and/or updates to be completed that calendar year.

- Based on this scope, the EM&V contractor will develop the updates, and submit these for comment in July. The results of these comments will be discussed in a second technical conference in August, with the TRM updates finalized in October.

Measures that may be appropriate for the TRM but that are not included in the then-current version should be brought forward in the first-quarter technical conference when possible. If a measure is brought forward by program implementers or other stakeholders, the EM&V contractor may work with the appropriate stakeholders in finalizing an ad hoc measure whitepaper for use until the measure can be formalized in a TRM update. It is at the discretion of the EM&V contractor to determine if primary data collection is warranted before allowing deemed savings for measures through this whitepaper process.

Updating of the TRM should be accomplished using data and tools that are the “best available” (i.e., accurate, relevant, and current). In particular, TRM updates should be based on EM&V studies that are conducted regularly.

The ongoing annual updating process will provide assessments of the reliability of deemed savings values, deemed calculations, and deemed variables and factors. Such assessments may not necessarily result in changes to the TRM. However, the reviews should assess whether the use of the “best (currently) available” data regarding baseline assumptions remains accurate or needs updating (e.g., because of changing code requirements or changes in market practices).