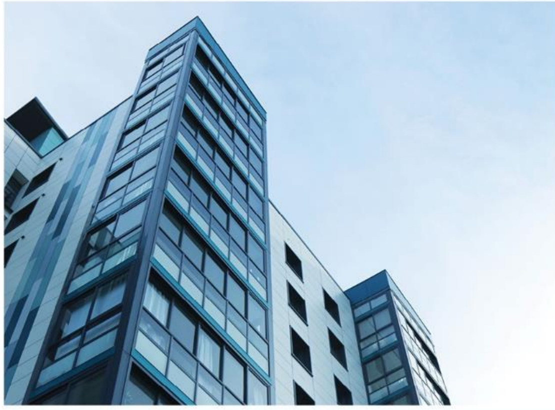
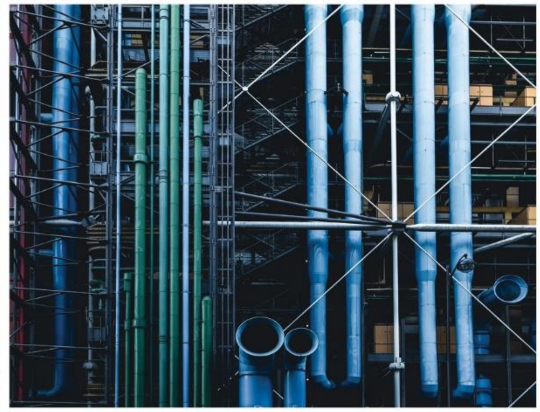


Cross-State C&I Active Demand Reduction Initiative Summer 2019 Evaluation Report

prepared for

Eversource, National Grid, and Unitil

ERS Corporate Headquarters
120 Water St., Suite 350
North Andover, MA 01845
978-521-2550
Fax: 978-521-4588

April 15, 2020

Contents

1 Executive Summary

Eversource, National Grid, and Unitil (the Program Administrators, or PAs) administer commercial and industrial (C&I) active demand reduction (ADR) initiatives (commonly known as demand response) in Massachusetts, New Hampshire, and Connecticut.

The PA ADR initiative goals are as follows:

- Reduce load during the system peak hour to capture system benefits (includes benefits for all rate payers) and participant benefits. The system benefits include reduced costs of avoided capacity, capacity demand reduction induced price effect (DRIPE), and avoided transmission and distribution investments. The participant benefits include a financial incentive from the PA to reduce load and ICAP charge reduction.
- Reduce carbon emissions on peak demand days. On the days with the highest demand, the marginal generating units are usually inefficient, simple-cycle gas peaker plants. Reduced operation of these plants would result in a beneficial emissions impact.
- Develop load shedding capabilities to assist the utilities in dealing with constrained circuits in the future, if the need arises.

These initiatives offer incentives to customers who respond to calls for peak demand reduction on summer weekdays. The PAs contracted ERS and DNV GL to evaluate the ADR initiatives. This evaluation focuses on the C&I interruptible and targeted battery storage projects implemented in the summer of 2019. The following PAs and states are included in this evaluation study and report:

- Eversource MA
- Eversource CT
- Eversource NH
- National Grid MA
- Unitil MA
- Unitil NH

Eversource and National Grid also offer incentives for daily dispatch battery storage projects; however, the daily dispatch projects were evaluated separately and are not included in this report. The Massachusetts and Connecticut winter 2019-20 initiative will also be the subject of separate reporting once that season is complete.

This is the first year of full-scale operation for all three PAs. National Grid had demand response demonstrations in 2017 and 2018, and Eversource had demonstrations in 2018 and

2019. It is important to note that Eversource's demonstration projects were technology-specific studies and did not have the same program design as the ADR initiative. In New Hampshire, the 2019 season is a pilot for Eversource and Unitil.

The PAs offered an incentive of $35/kW-summer for facilities that reduced their demand during peak demand events called by the PA. Under the targeted battery storage initiative, Eversource offered an incentive of $100/kW-summer for customers who used battery storage to reduce load during events called by the PA. The PAs contracted Curtailment Service Providers (CSPs) to recruit customers, notify customers of events, and pay incentives. Customers were only paid to the extent that they actually reduced demand. There are no direct penalties for non-performance, but non-performance reduced incentives proportionally. Table 1-1 shows the Summer 2019 initiative demand reduction goals by state and PA as well as the aggregate reported demand reductions (includes all customers including those without sufficient data for evaluation) and participation.

Table 1-1. Summer 2019 C&I Active Demand Reduction Initiative Goals and Participation

		Demand Reduction (MW)		Number of Curtailment Service Providers	Number of Participants	
Program Administrator	**State**	**Target kW**	**Reported**		**Accounts**	**Customers**
Eversource	MA	30.0	23.2	3	150	56
	CT	6.0	13.0	3	96	20
	NH	5.0	5.9	3	40	9
	Subtotal	**41.0**	**42.1**	**3**	**286**	**74**
National Grid	**MA**	**62.0**	**74.0**	**4**	**364**	**147**
Unitil	MA	0.9	0.9	1	3	3
	NH	1.8	1.3	1	7	4
	Subtotal	**2.7**	**2.2**	**1**	**10**	**7**
Total		**105.7**	**118.3**	**7**	**660**	**228**
All	MA	92.9	98.1	7	517	206
Eversource	CT	6.0	13.0	3	96	20
Eversource + Unitil	NH	6.8	7.2	3	47	13
Total		**105.7**	**118.3**	**7**	**660**	**239**

1.1 Study Objectives

The primary objectives of the evaluation are to independently assess program initiative impact and identify process improvement opportunities. Impact is measured as both the average demand reduction during specified events and during the annual peak installed capacity (ICAP) hour. Load reduction is based on the comparison of measured load against four different alternative/baseline load scenarios. The evaluation also attempts to understand the overlap between the PA ADR initiatives and the ISO-NE Forward Capacity Market (FCM) and provide input on other opportunities for peak demand management.

1.2 Evaluation Methodology, Framework, and Performance Metrics

For the process evaluation, the team reviewed initiative materials and conducted in-depth phone interviews with PA program managers, ISO-NE Price Responsive Demand (PRD) program managers, and CSPs. Discussion focused on barriers, overlap with ISO-NE programs, satisfaction, and opportunities. The team also conducted a mixed-mode (online and phone) participant survey during November and December 2019 with a sample of 61 customers to meet the target of 10% relative precision at 90% confidence.

For C&I interruptible interventions, the impact evaluation was based on account-level analysis of interval data of all available participants. To evaluate demand reduction impact, the counterfactual performance (i.e., baseline behavior, or the hypothetical behavior of customers absent the initiative) must be estimated. The evaluation team calculated baseline load based on five different scenarios:

1. **Evaluated-Unadjusted: Unadjusted 10-of-10 baseline.** This baseline constructs a pool of the 10 most recent eligible days occurring prior to the day for which load is being estimated. For the ADR initiative, eligible days include non-holiday weekdays on which a demand response event did not occur. The baseline shape is calculated as the rolling average of load in each interval across the 10 most recent eligible days.
2. **Evaluated-Symmetric: Adjusted 10-of-10 baseline with symmetric, additive adjustment.** This rolling baseline shifts the unadjusted 10-of-10 to meet observed load during pre-event hours (adjustment period) using a symmetric additive adjustment. Additive (as opposed to multiplicative) means that the magnitude of the shift for all intervals equals the difference between observed load and the unadjusted baseline during the adjustment period. Symmetric (as opposed to asymmetric) means that the adjustment may shift the unadjusted baseline upward or downward. Generally, however, the resulting baseline may not be less than zero (with the exception of some assets that have behind-the-meter generation)[1] and the associated load reduction may not be greater than the maximum load during the baseline period.
3. **Evaluated-Asymmetric: Adjusted 10-of-10 baseline with asymmetric adjustment.** This rolling baseline shifts the unadjusted 10-of-10 to meet observed load during pre-event hours (adjustment period) using an asymmetric additive adjustment. "Asymmetric"

[1] For assets with behind-the-meter generation that often provide net supply to the grid, a baseline floor of zero can overstate load reduction. Alternatively, the asset's interconnection agreement may be used to establish the baseline floor.

means that adjustments only shift the unadjusted baseline upward but never downward. One advantage to excluding baseline decreases is to avoid penalizing customers for early response to an event that decreases load during the adjustment period. This basis is used by the PAs for settlement and payment.

4. **Evaluated-Forecast: Adjusted 10-of-10 baseline with symmetric, additive adjustment, and adjusted to account for likelihood of shutdowns**. This baseline uses the Evaluated-Symmetric model and then adjusts it to account for the likelihood of unreported shutdowns. This metric satisfied a key goal of this evaluation, which is to produce an estimate of expected future load reduction.
5. **Evaluated-Regression: Regression baseline.** This baseline fits a regression model to an individual customer's load data across the entire season. The regression specification describes load for each hour of the day as a function of cooling degree-days (CDD), weekends and holidays, calendar month, and event day terms.

Table 1-2 shows the summary of performance metrics. These metrics are the baselines that were analyzed. The load reduction is the difference between these baselines and the actual observed load during the event hours.

Table 1-2. Summary of Performance Metrics for C&I Interruptible

Enrolled Capacity	Asset-level expected load reduction. Generally estimated by CSP.
Reported-Asymmetric	Load reduction claimed, reported by program implementers.
Evaluated-Validation	Evaluation team's attempt to replicate reported asymmetric results. Baseline calculation details (e.g., lookback days) vary by PA.
Evaluated-Unadjusted	Rolling average of load in each interval across the 10 most recent eligible days.
Evaluated-Asymmetric	Shifts the unadjusted baseline upward to meet observed load during pre-event adjustment period. Unadjusted baseline is used in lieu of downward adjustments.
Evaluated-Symmetric	Shifts the unadjusted baseline upward or downward to meet observed load during pre-event adjustment period
Evaluated-Forecast	Modification of evaluated symmetric results by accounting for probability of unreported shutdowns.
Evaluated-Regression	Site-level model of load across the summer. Specification describes load as a function of cooling degree-days, weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit.

For targeted battery storage projects, the impact evaluation was based on battery meter data. Event demand reduction is equal to the battery load (charge/discharge). The battery load is negative when discharging and positive when charging. The evaluators also calculated the seasonal battery efficiency, which is the ratio of the total battery discharge divided by the total battery charge over the season.

1.3 Process Evaluation Findings

The process evaluation team interviewed PA implementation staff and CSPs to identify successes and challenges in initiative delivery during the 2019 summer season. The team also surveyed participants to gauge barriers to entry and customer satisfaction.[2] Customers reported general satisfaction with all aspects of the ADR initiatives except for the event payments.

The key challenges in program administration are summarized below:

Eversource:

1. **Administration:** The program staff used multiple systems to track 2019 application submittals, send notifications, and monitor events, which they report was a challenge. Eversource is designing a distributed energy resource management system (DERMS) platform to enable all program administration for all states to be conducted within one platform.
2. **Settlement and payment:** The decentralized program management systems in use made the settlement calculation and payment process difficult for Eversource's implementation team. Eversource staff expects that the new centralized DERMS platform will improve the payment processing for the next summer season. Regarding event payments, over half (61%) of Eversource customers were unsure how to rate their satisfaction (i.e., gave an answer of "don't know"). Note that this does not indicate dissatisfaction; the evaluation team recommends monitoring customer satisfaction with event payments.
3. **Program application:** Three CSPs reported that Eversource changed the Terms and Conditions (T&C) on the customer application several times throughout the sales season (January – May). At times, multiple versions of the application (with different T&Cs) were in circulation. These CSPs reported having a constant back-and-forth with Eversource, even well into the summer, to negotiate the T&Cs and clarify program rules.

National Grid:

1. **Settlement and payment:** As with Eversource, payment processing remains challenging. Program marketing materials state that incentives will be paid out in October. When the evaluation team interviewed National Grid's staff in December of 2019, there were still several payments that had not yet been made. National Grid staff explained that the delay in payments was largely because National Grid's procurement protocols had been revised. This meant that National Grid staff had to have CSPs re-sign contracts, NDAs, and ISAs in order to process ADR initiative payments. Additionally, staff explained that each summer season, a small percentage (less than 5%) of customers

[2] The team administered customer surveys in November and December of 2019.

experience metering or data issues that result in delayed settlement and payment. With the increase in program enrollment over the past three summers, it has become more time-consuming for staff to resolve these data issues.

2. **Recruitment:** Recruitment for the 2019 summer season was more difficult for the PA than it had been over the first two years of the program. A significant percentage of the PA's largest commercial and industrial customers have already signed up for the program. The National Grid staff noted they had to increase their sales efforts in 2019 to achieve the same amount of MW reduction that had been reached in prior years when less resources were spent on recruiting. As can be expected with any growing PA offering, the PA anticipates this need for increased promotion to continue.

Unitil:

Unitil's MA initiative and NH pilot were small. (Unitil had only seven enrolled participants.) Unitil experienced few challenges apart from missing their NH kW delivered goal. Participants were mainly recruited by one CSP for the summer 2019 season. To meet their summer 2020 demand reduction targets, Unitil staff are considering several changes to their promotional efforts, including developing content on their website and working with an additional CSP.

1.4 Impact Evaluation Findings

The evaluators calculated the average load reduction during the events called by the PAs. The events were called as follows:

- 7/19 from 4:00 to 7:00 p.m. (Eversource only)
- 7/30 from 3:00 to 6:00 p.m. (Eversource, National Grid, and Unitil)
- 8/19 from 4:00 to 7:00 p.m. (Eversource only)

The sections below show the findings and key takeaways from the impact evaluation.

C&I Interruptible Findings

The evaluation team recommends using a symmetrically adjusted baseline (called Evaluated-Symmetric in the tables below) as the most appropriate estimate of event period load reduction for the 2019 summer season. The symmetrically adjusted baseline, with additional adjustment for likelihood of unreported shutdowns (called Evaluated-Forecast below), is the best estimate of load reduction for future years. The symmetrically adjusted baseline reduces biases for sites with variable load due to weather or other production factors. The symmetrically adjusted baseline methodology is the most commonly used baseline approach and is used by ISO-NE. The Baselines section of the Impact Evaluation Methodology and Framework describes the baselines and their advantages and disadvantages in detail. The C&I Interruptible Impact Evaluation Findings and Integrated Impact and Process Evaluation Findings substantiate the evaluator's baseline recommendation for the 2019 summer season. Table 1-3 provides Eversource's average load reduction estimates during event hours. The ICAP hour occurred July 30 from 5 to 6 p.m.

Table 1-3. Eversource Impact Summary

	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
Result	**MA**		**NH**		**CT**	
Enrolled Capacity	37,248	36,683	5,905	5,905	12,519	12,519
Reported-Asymmetric	22,261	N/A	5,156	N/A	13,085	N/A
Evaluated-Validation	21,528	N/A	5,786	N/A	12,558	N/A
Evaluated-Unadjusted	6,561	7,371	3,202	2,689	7,951	7,275
Evaluated-Asymmetric	19,912	21,760	5,661	5,981	12,158	11,564
Evaluated-Symmetric	17,432	20,558	5,147	5,947	11,462	10,921
Evaluated-Forecast	16,523	19,380	4,953	5,674	11,330	10,779
Evaluated-Regression	12,749	15,385	4,446	3,823	9,594	9,824
Accounts	**161**	**151**	**40**	**40**	**96**	**96**

Table 1-4 provides the summary of National Grid's load reduction estimates for Massachusetts. National Grid called a single event on July 30, the ICAP day.

Table 1-4. National Grid Impact Summary – Massachusetts

Result	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
Enrolled Capacity	93,134	93,134
Reported-Asymmetric	71,428	N/A
Evaluated-Validation	71,611	N/A
Evaluated-Unadjusted	42,461	36,090
Evaluated-Asymmetric	69,561	63,190
Evaluated-Symmetric	58,464	52,173
Evaluated-Forecast	57,264	51,266
Evaluated-Regression	48,752	42,538
Accounts	**357**	**357**

As with National Grid, Unitil called a single event on July 30. Results are in Table 1-5.

Table 1-5. Unitil Impact Summary

Result	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
	MA		NH	
Enrolled Capacity	950	950	1,600	1,600
Reported-Asymmetric	853	N/A	1,299	N/A
Evaluated-Validation	843	N/A	1,363	N/A
Evaluated-Unadjusted	587	601	1,036	784
Evaluated-Asymmetric	843	856	1,363	1,111
Evaluated-Symmetric	775	788	1,185	980
Evaluated-Forecast	775	788	1,153	949
Evaluated-Regression	668	687	588	341
Accounts	**3**	**3**	**7**	**7**

Table 1-6 shows the state-level impact summary.

Table 1-6. State-Level Impact Summary

Result	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
	MA		NH		CT	
Enrolled Capacity	131,332	130,767	7,505	7,505	12,519	12,519
Reported-Asymmetric	94,543	N/A	6,455	N/A	13,085	N/A
Evaluated-Validation	93,982	N/A	7,149	N/A	12,558	N/A
Evaluated-Unadjusted	49,609	44,062	4,238	3,473	7,951	7,275
Evaluated-Asymmetric	90,316	85,805	7,024	7,092	12,158	11,564
Evaluated-Symmetric	76,671	73,519	6,332	6,927	11,462	10,921
Evaluated-Forecast	74,562	71,435	6,106	6,623	11,330	10,779
Evaluated-Regression	62,169	58,610	5,035	4,164	9,594	9,824
Program Administrators	**3**	**3**	**2**	**2**	**1**	**1**

Based on the above results, the evaluators calculated the following performance ratios. They are defined as follows:

- Enrollment Ratio: This ratio is the reported asymmetric load reduction to the CSP reported enrolled capacity. This ratio provides insight into what percentage of the reported enrolled capacity was achieved, based on the program baseline and calculation methodology. This ratio is particularly meaningful for planning and sales purposes.
- Asymmetric Ratio: This ratio is the evaluated asymmetric load reduction to the reported asymmetric load reduction. This is an apples-to-apples comparison of the same baseline

methodology between the PAs and evaluators; however, this metric identifies the impact that different calculation rules between the PAs and evaluators has on load reduction.

- Retrospective Realization Rate: This ratio is the evaluated symmetric load reduction to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline is the most appropriate measure of retrospective load reduction for the 2019 summer season. This ratio shows how the choice of baseline adjustment and calculation methodologies impacts the load reduction estimates. The evaluators recommend using this realization rate to calculate the symmetric load reductions at the end of future seasons if there are no evaluations conducted.
- Prospective Realization Rate: This ratio is the evaluated symmetric load reduction with an adjustment for unreported shutdowns to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline accounting for unreported shutdowns is the most appropriate measure of prospective load reduction for future seasons. This ratio provides insight into the magnitude of reductions that could be achieved during future seasons as a function of the validated load reduction estimates. The prospective realization rate should only be used as an ex-ante estimate of future performance for planning purposes and not retrospectively.

Table 1-7. Summary of Performance Ratios

PA and State	Enrollment Ratio (Reported Asymmetric/ Enrolled Capacity)	Asymmetric Ratio (Evaluated Asymmetric/ Reported Asymmetric)	Retrospective Realization Rate (Evaluated Symmetric/ Reported Asymmetric)	Prospective Realization Rate (Evaluated Forecast/ Reported Asymmetric)
Eversource MA	59.8%	89.4%	78.3%	74.2%
Eversource NH	87.3%	109.8%	99.8%	96.1%
Eversource CT	104.5%	92.9%	87.6%	86.6%
National Grid MA	76.7%	97.4%	81.9%	80.2%
Unitil MA	89.8%	98.8%	90.8%	90.8%
Unitil NH	81.2%	104.9%	91.2%	88.7%
MA – All 3 PAs	72.0%	95.5%	81.1%	78.9%
NH – Eversource, Unitil	86.0%	108.8%	98.1%	94.6%
CT – Eversource	104.5%	92.9%	87.6%	86.6%

Targeted Battery Storage Findings

There were two targeted battery storage systems enrolled in the initiative during the summer 2019 season. Battery 1 was deployed at a manufacturing facility in MA while Battery 2 was deployed at a university in MA. Both customers are Eversource customers and were called three times during the summer season. They were also called for a voluntary weekend event, in which both systems were dispatched.

Battery 1 was dispatched in two different ways throughout the summer season: daily dispatch and targeted dispatch. The battery was committed to dispatch 500 kW on a daily basis from 4 p.m. to 7 p.m. on non-holiday weekdays. The targeted events and the daily dispatch schedule overlapped on all three event days. Hence, the evaluators attributed the first 500 kW of battery discharge to daily dispatch and the remaining kW reduction, if available, to the targeted event. On the ICAP day, July 30, an Eversource event was called from 3 to 6 p.m. Both battery systems were successfully dispatched during the ICAP hour (5 to 6 p.m.); however, neither system was dispatched from 3 to 4 p.m. In fact, Battery 1 was charging for that hour.

Table 1-8. Battery 1 Results

	PA Event Average Reduction (kW)	Weekend PA Event Average Reduction (kW)	ISO-NE ICAP Hour Reduction (kW)	Net Energy Impact (kWh)
Committed	400	N/A	N/A	N/A
Reported	283	347	800	N/A
Evaluated	295	703	800	84,122*
Hours	**9**	**3**	**1**	

N/A = Not available

*The ratio of discharging energy to charging energy for the 2019 summer season was found to be 0.674.

Table 1-9. Battery 2 Results

	PA Event Average Reduction (kW)	Weekend PA Event Average Reduction (kW)	ISO-NE ICAP Hour Reduction (kW)	Net Energy Impact (kWh)
Reported	1,033	N/A	1,290	7,000
Evaluated	1,033	1,162	1,290	N/A*

*The ratio of discharging energy to charging energy for the 2019 summer season was reported to be 0.897. Net energy evaluation was not possible due to absence of beginning and end states of charge and charging data for the days preceding the first event day.

The performance of the two batteries resulted in a realization rate of 1.01, meaning that the evaluated performance was 1% higher than the reported performance.

1.5 Integrated Process and Impact Evaluation Findings

Challenges to Reliability: The ADR initiative and prior demand demonstrations provide substantial evidence that it is reasonable to expect PA load reduction targets to be met. Despite variability in load reduction across states and events, the CSPs and PAs have successfully recruited and managed resources, identified the annual system peak hour, and met overarching PA load reduction targets. However, the variability in load reduction across the limited number of event days and hours reduces confidence in the reliability of DR resources in the future. A better understanding of the dimensions of variability inherent in any DR program mitigates these concerns and suggests opportunities for continued reliability improvements.

Shutdown Days: Customers are failing to report shutdown days to the PAs even though the initiative rules include a shutdown day allowance. The intent of this rule was to ensure that customer performance would not be negatively impacted if they had scheduled shutdowns. The shutdown day rule could save customers from a lower than expected event performance if events were called on a shutdown day. Also, shutdown days could have an impact on customer payments if they fell within their baseline period and went unreported.

Pre-Cooling, Gaming, and Snapback: The impact evaluation investigated whether there was evidence of pre-event load increases that could be explained either by pre-cooling, load shifting, or gaming.[3] Pre-cooling and load shifting are acceptable strategies for participation in the ADR initiative; however, acceptable load shifting strategies can be difficult to distinguish from gaming. Post-event, the impact evaluation investigated whether there was evidence of post-event load increases that could be explained by snapback. None of the load shapes point toward pre-event activity or snapback. The process evaluation investigated whether there was evidence of gaming through the customer surveys, where customers were asked if building operational adjustments were made in the hours leading up to an event. A quarter of respondents said yes but described taking action to *reduce* load prior to events (e.g., begin shutdown of slow-ramping equipment) to ensure that they could curtail adequately during events. These findings are described in more detail with illustrations in the body of this report (Evidence of Pre-Cooling, Gaming, and Snapback).

Summary of PA ADR Initiatives and ISO-NE Overlap Findings

- Scenarios in which PA ADR initiative events and ISO-NE scarcity conditions overlap or are called coincidently are rare, as scarcity conditions occur because of a supply constraint (at the transmission level) while PA ADR events are called in response to mitigate load during the system peak hour.
- ISO-NE staff concerns regarding PA ADR initiative overlap are:
  - Participation in PA ADR events could result in eroding the ISO-NE baseline calculation and same-day adjustment for performance, or vice versa.

[3] Gaming refers to manipulative load increases in the hours leading up to an event, specifically during the baseline adjustment period, to inflate the settlement baseline and associated performance payment in a way that does not reflect actual load reduction. Pre-cooling and snapback are different forms of HVAC load shifting strategies related to reducing load during a prescribed time period. Pre-cooling lowers setpoints in advance of an event to minimize discomfort during the event while snapback represents the additional load required to attain typical setpoints after the event.

   - The ISO could over designate reserves of demand response resources (DRRs) that participate in PA ADR initiatives if their FCM bids are not revised.
- Although the ADR initiative rules specify how co-participation in the PA initiative and ISO-NE FCM should work, the PA initiative rules do not address the ISO's overlap concerns.
- Both ISO and PA staff expressed a willingness to discuss overlap concerns and solutions.

1.6 Evaluation Conclusions, Recommendations, and Considerations

- **Conclusion 1: Requiring participants to sign the PA ADR application after they have signed the agreement with a CSP slows down and complicates the sales process.** Several CSPs noted that asking a customer to sign the additional PA contract hinders the recruitment process. Customers questioned why they needed to sign another agreement when they had already signed an agreement with the CSP. One CSP observed several large prospects refusing to sign the PA's contract after they signed the agreement with the CSP. CSPs also said they are unclear about the purpose of the PA contract for those customers who participate through a CSP, since the customer has already provided the same information and agreed to a contract with the CSP.

   Consideration 1-1: Explore consolidating PA and CSP terms and conditions into a single document or at least presenting the two agreements at the same time during recruiting. National Grid stated that it is standard practice for CSPs to present customers with both the CSP contract and PA application at the same time, but this is not what was reported by CSPs. National Grid staff should check that CSPs are adhering to this practice.

- **Conclusion 2: Feedback from nearly all sources suggests that settlement and payment is a significant administration challenge.** ADR initiative materials (e.g., flyers on National Grid's website, Eversource's Active Demand Reduction application) indicate that payments for the DR summer season will be made by the fall. At least one PA understood that CSPs had communicated the likelihood of a later payment period to participants, but this did not come up in CSP or participant interviews. By December 2019, a minority (less than one-third) of surveyed participants across the three PA initiatives reported that they had received event payments for the 2019 summer season. For Eversource, the settlement calculation and payment processes were difficult due to the interval meters failing to record or transmit information and the decentralized tracking systems. National Grid staff stated that revised procurement requirements created payment delays, and they also reported that a small number (less than 5%) of payments were delayed due to metering and data issues. When asked if they were satisfied with the event payments, the majority of participants (61%) in the Eversource sample and a notable proportion (31%) in the

National Grid sample were uncertain (stated "don't know") since many of them noted that they had not yet received payment. The National Grid effort was in operation with evaluation cycles for two years before this current study, which likely explains why their respondents were more familiar with payment processes. It is also important to note that the survey results did not indicate that the participants were dissatisfied, just that they were not yet satisfied.

Recommendation 2-1: Continue to seek solutions to accelerate the incentive payment process. Eversource is building a centralized program tracking and management system, which will pull customer interval data directly from the CSPs in close to real time. If the system performs as expected, it will shorten the time it takes the PA to complete the settlement calculation and issue payment. National Grid is starting to allow CSPs to access their online day-after data and daily performance summaries. This access should help CSPs more quickly identify faulty meters or reconcile data discrepancies, which affect the payment turnaround time.

Consideration 2-1: Monitor participant satisfaction with event payments by periodically surveying customers after payments are sent out. Consider adjusting initiative materials to more strongly convey the likely payment schedule.

- **Conclusion 3: Fear of facility disruptions, even when minor, is a barrier to participation.** This barrier may be more prominent for the manufacturing facilities than any other customer types. Manufacturing facilities were the majority of Eversource and National Grid participants who experienced slight or temporary disruptions when the events were called and thus reported performing partially or not at all during the DR event(s). CSPs explained that industrial facilities will not always be able to participate due to their inability to decrease production when an event is called.

- **Conclusion 4: All the participants across all three PAs who responded to the surveys stated that their opinion of the PAs was either positively impacted or unaffected by the ADR initiatives.** The initiative is improving the participants' opinions of their PAs. This is despite the fact that, at the time of questioning, many participants had not received their payment for curtailment, which can heavily influence levels of satisfaction with overall engagement.

- **Conclusion 5: The ADR initiative is not consistently branded across the three PAs' marketing materials and initiative administration documentation.** For example, National Grid, Eversource, and Unitil use the "ConnectedSolutions" brand to refer to this initiative in MA, but Unitil does not use the "ConnectedSolutions" brand in NH. Additionally, the Eversource website refers to the initiative as ConnectedSolutions, but the

Eversource DR initiative application is titled the "Active Demand Reduction" application. This can create confusion in the marketplace.

- **Conclusion 6: Customers and CSPs are failing to report shutdown days to the PAs.** Initiative rules include a limited number of announced shutdown days, communicated in advance. The benefit to participants and PAs is that the shutdown day can and should be excluded if it occurs during the baseline period, to avoid an inappropriate reduction in baseline load and claimable load reduction. Disclosing it in advance of what turns out to be an event day is beneficial as well, as it may prevent the penalty of no/low-DR performance and helps PAs and CSPs manage event day expectations. Not knowing shutdown days also interferes with evaluators' effectiveness in using regression-based baselines, even if they occur outside of the baseline periods that affect settlement. Per interviewed PA staff, neither customers nor CSPs reported any shutdown days in the 2019 summer season. Yet one surveyed participant reported their facility was closed prior to the actual DR event, two CSPs had customers report facility shutdown events, and interval meter data showed evidence of shutdowns. CSPs noted three reasons for not reporting shutdown events to the PA: 1) facility shutdown events were unexpected, 2) customers forgot to report it, or 3) shutdown events do not fall within any baseline periods.

  - **Recommendation 6-1:** Remind and educate the CSPs of the shutdown allowance and reporting rule. The PAs could ask for pre-planned shutdown information during the application/enrollment process.
  - **Recommendation 6-2:** Adapt the shutdown rule to account for unexpected facility shutdown events. To exclude a facility shutdown day from a customer's baseline calculation, that customer or their CSP must notify their PA at least seven days in advance of the shutdown. It is difficult to do this when a facility shutdown event is unexpected. Consider allowing customers or CSPs to report the shutdown to the PA 24 hours before an event is called.

- **Conclusion 7: Data quality and analysis requirements are not coordinated or consistent across the PAs.** There is little evidence of coordination across PAs with regards to data sufficiency rules and other load reduction calculation details. Each PA's approach would produce different load reduction estimates if applied to the same group of assets. Within a given state (e.g., Massachusetts), load reduction should be calculated in a consistent fashion across administrators. Inconsistencies of this nature jeopardizes the ability of the PAs to aggregate their estimates at a state level with confidence.

  - **Recommendation 7-1:** Formally standardize all rules related to data quality, baseline calculation methods, and aggregation.

Recommendation 7-2: Establish data quality rules with clear outcomes for poor quality and/or insufficient data. The evaluation team developed several rules as part of this study, which are described in the Data Sufficiency section of this report; these rules may be a useful starting point to develop consistent rules. The data issues encountered in this study were not anticipated. Establishing firmer expectations or providing incentives are two possible means of motivating and ensuring clean and complete data in future initiative cycles.

- **Conclusion 8: The quality of the settlement baseline process can be improved.** A demand response offering is only as good as its settlement baseline process. The PAs should consider changing the initiative to a baseline that is more in line with ISO-NE's baseline. The ISO's baseline has evolved over time to reflect the best baseline methods for attaining an accurate estimate of load reduction. Important aspects for consideration include a symmetric adjustment, no day-ahead notification, and meaningful reporting of shutdown days. While not fully examined, the current initiative baseline has theoretical limited downside risks for assets with highly variable load that do not intend to take load reduction actions but will receive payments when random load variation produces a positive load reduction relative to baseline. A change in notification from day-ahead to day-of would improve the quality of the settlement load reduction estimate by eliminating customers' ability to increase load during a known adjustment hour in anticipation of event hours to follow later that day, with the potentially significant trade-off to the initiatives of more expensive recruitment and/or likelihood of lower MW enrollment, due to customer preference or need for advance notice to participate.

Consideration 8-1: The PAs should consider revising the program design and baseline to be more in line with ISO-NE. There are several options that could move rules closer to ISO-NE rules, including:

- Pay incentives to participants based on the symmetrically adjusted 10 of 10 baselines instead of asymmetric. Incentives may need to increase correspondingly.[4]
- Provide an additional incentive for customers to be dispatched without prior-day notification.
- Increase payment for delivered kW, while adding a penalty or de-rating factor for partial delivery. Delivery below some benchmark (e.g., 80% of enrolled) would be paid on a sliding scale down to some benchmark with no payment below that level.

[4] The symmetric baseline will produce lower estimates of load reduction for assets that shutdown early for events. These assets may require additional payment to choose a symmetric baseline.

- **Conclusion 9: Evaluators found that the 10-of-10 baseline with symmetric adjustment was the most appropriate measure of impact for the 2019 summer season and is the basis for the retrospective realization rate. Evaluated forecasts can be utilized to estimate reductions for future summers.** In each result table, the evaluators provide evaluation symmetric and forecast values that reflect our best estimate of the level of reduction achieved and that might be expected on a forward-looking basis given the use of a symmetrically adjusted baseline with an additional shutdown adjustment.

  Recommendation 9-1: Use the retrospective realization rate to determine past season performance.

  Recommendation 9-2: Use the prospective realization rate to estimate future load reduction.

- **Conclusion 10: Lack of coordination between ISO-NE markets and the PA's ADR initiatives could have unintended consequences.** The lack of coordination and communication between the PA DR and ISO-NE offerings could affect the ISO-NE baseline calculations and could result in the over-estimation of available reserves. Both ISO and PA staff expressed a willingness to discuss overlap concerns and solutions.

  Recommendation 10-1: In the short-term, representatives from ISO-NE, the PAs, and, if feasible, the CSPs should come together at a Demand Resources Working Group (DRWG) meeting and brainstorm mutually beneficial design solutions that would minimize the impact of one entity on the other.

  Consideration 10-1: The PAs could consider incorporating language into their rules that instruct the CSPs to either make themselves unavailable or modify their bids so as not to clear on the ISO platform during PA ADR events.

- **Conclusion 11: There is a natural tension between program design for program implementation and program design for the most comprehensive assessment of reliability.** The ADR initiative design minimizes customer burden while maximizing value to the customer by minimizing the number of events called, offering day-ahead notification, and using the asymmetric baseline. A program designed to produce the most reliable and consistent estimates of load reduction would call more events, forego day-ahead notifications, and use a symmetric baseline for the basis assessing contractual targets. If prioritized, these changes would likely increase customer burden while increasing program costs and/or decreasing the amount of available load reduction. The optimal combination of these program characteristics will flow from a clear recognition of reliability expectations for these programs.

2 Introduction

This section details each PA's initiative design and the evaluation study objectives.

2.1 Initiative Design Summaries

Eversource, National Grid, and Unitil (the Program Administrators, or PAs) administer active demand reduction (ADR) initiatives, commonly called demand response (DR) initiatives, in New England. These initiatives offer incentives to customers that reduce peak demand. This evaluation focuses on the Commercial and Industrial (C&I) curtailment and targeted battery storage initiatives implemented in Massachusetts, New Hampshire and Connecticut in the summer of 2019. Battery storage applies only to Eversource in MA and CT. Eversource MA and National Grid have had one and two years, respectively, of demand demonstrations that preceded 2019. It is important to note that Eversource MA's demonstration projects were technology-specific studies and did not have the same program design as the ADR initiative. The PA ADR initiative goals are as follows:

- Reduce load during the system peak hour to capture system benefits (includes benefits for all rate payers) and participant benefits. The system benefits include reduced costs of avoided capacity, capacity demand reduction induced price effect (DRIPE), transmission, and distribution. The participant benefits include a financial incentive from the PA to reduce load and ICAP charge reduction.
- Reduce carbon emissions on peak demand days. On the days with the highest demand, the marginal generating units are usually inefficient, simple-cycle gas peaker plants. Reduced operation of these plants would result in a beneficial emissions impact.
- Develop load shedding capabilities to assist the utilities in dealing with constrained circuits in the future, if the need arises.

Under the curtailment initiative, PAs offered an incentive of $35/kW-summer for facilities that reduced their demand during peak demand events called by the PA. The PAs selected event hours in part to curtail load during the expected system peak hour.[5]

[5] The ISO-NE levies a charge to retail electricity providers for the cost of having available installed capacity during high load periods. The charge is allocated based on their customers' load during the system summer peak hour. The peak hour cannot be known with certainty until after the summer is over. Eversource and National Grid considered many different sources to forecast the 2019 ICAP hour and monthly peaks. Utility staff sent notifications and completed the settlement process in house. Providers may either charge individual customers according to their demand at that hour or socialize the cost, spreading it to all without regard to peak hour load. Regardless, the ICAP Tag charge must be paid to the ISO. The charge provides an incentive to retailers, curtailment service providers and end users to reduce the ICAP Tag hour load. In addition, limiting peak demand reduces the need for peaking plants, which tend to have higher emissions than other plants.

Under the targeted battery storage initiative, one PA (Eversource) offered an incentive of $100/kW-summer for customers that had battery storage to reduce load during events called by the utility. The PAs contracted with Curtailment Service Providers (CSPs) to recruit customers, notify customers of events, and pay incentives. Customers were only paid to the extent they actually reduced demand, and there were no direct penalties for non-performance. Massachusetts' and Connecticut's winter 2019-20 initiative will be the subject of a separate report once that season is complete. The tables below show the summary of DR initiatives.

Table 2-1. Initiative Design Summaries

Option	Targeted Dispatch	Targeted Dispatch with Batteries
Administrators	Eversource, National Grid, Unitil	Eversource
Incentive (Per Summer)	$35 /kW	$100 /kW
Season	Summer	
Number of Events per Season	2 to 8	
Incentive Lock	None	
Length of Events	3 hours	
Time of Day	2 p.m. to 7 p.m.	
Weekend/Weekday	Weekdays only	
Events on Holidays	None	
Day-Ahead Notification	Yes	
Months	June 1 – September 30	

Figure 2-1 shows the Summer 2019 initiative targets goals by state and PA. Table 2-2, below, summarizes the same goals targets as well as reported demand reduction and participation by PA and state. The table also includes the number of process interviews for later reference.

Figure 2-1. Summer 2019 Demand Response Initiative Goals

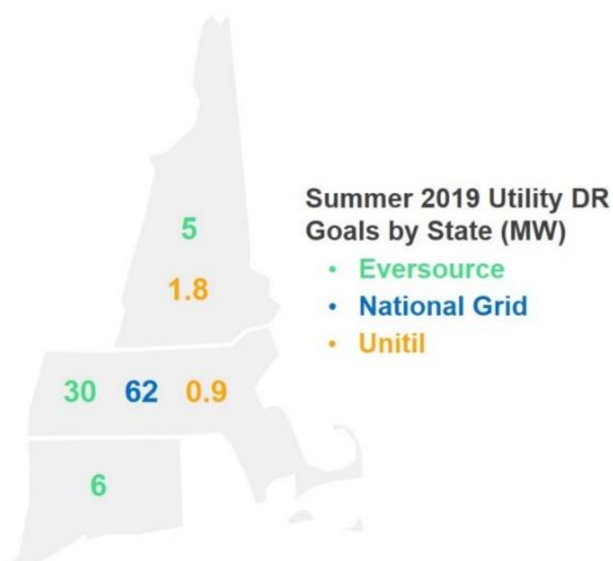


Table 2-2. Summer 2019 Demand Reduction Initiative Targets and Reported Goals

		Demand Reduction (MW)		Number of Curtailment Service Providers[1]		Number of Participants[1]		
Administrator	State	Goal	Reported	Total (N)	Process Interviews (n)[1]	Total Customer Accounts	Total Customers (N)[1]	Process Interviews (n)
Eversource	MA	30.0	23.0	3[a]	3[a]	150	56	17
	CT	6.0	13.0	3	3	96	20	5
	NH	5.0	5.9	3	3	40	9	5
	Subtotal[2]	**41.0**	**42.0**	**3[a]**	**3[a]**	**286**	**74**	**23**
National Grid	**MA**	**62.0**	**74.0**	**4**	**4**	**364**	**147**	**32**
Unitil	MA	0.9	0.9	1	1	3	3	2
	NH	1.8	1.3	1	1	7	4	1
	Subtotal	**2.7**	**2.2**	**1**	**1**	**10**	**7**	**3**
Total		**105.7**	**118.2**	**4**	**4**	**660**	**228**	**61**
State Subtotals								
All	MA	92.9	97.9	7	5	517	206	49
Eversource	CT	6.0	13.0	3	3	96	20	7
Eversource[3] and Unitil	NH	6.8	7.3	3	3	47	13	5
Total		**105.7**	**118.2**	**7**	**5**	**660**	**239**	**61**

[1] CSP and customer subtotals and totals exclude overlapping accounts.

[2] Eversource DR total includes MA battery storage participants.

[3] National Grid is estimated, as one CSP did not provide a customer list.

[a] There were two other vendors with one participant each in Eversource's DR initiative. The evaluation team interviewed one of them.

2.2 Study Objectives

ERS and DNV GL, together the evaluation team, developed and executed both the process and impact evaluations.

The objectives of the process evaluation are to understand the customer experience, barriers to implementation, and PA and vendor success in delivery. The process and impact evaluations will also aim to understand the overlap between the PA ADR initiatives and the ISO-NE market, and provide input on other opportunities for peak demand management.

The primary objective of the impact evaluation is to provide verification of the proper baseline and impacts generated by the DR initiatives in each state after each of the two seasons (summer 2019 and winter 2019-2020). This report convers summer 2019.

The impact evaluation reports the average hourly load reduction across event hours and during the ICAP hour. Aggregate load reduction is reported by state and by PA. Load reduction is based on comparison of measured load against four different alternative/baseline load scenarios and reported separately for each. To the extent that retrospective results differ from expected future results, this is distinguished as well.

A secondary objective is to assess net energy use (e.g., change in kWh) — that is, to determine to what extent DR is being driven more by pure load reduction or by load shifting. Turning lights off is an example of pure load reduction. Pre-cooling, post-event air conditioning bounce-back, and industrial production movement from daytime to evenings are examples of load shifting.

Table 2-3. Evaluation Objectives

Cross-State	
Impact	**Process**
Magnitude of reductions for each solution: What is the demand reduction that each solution is able to provide? How does this compare to planned target and vendor-reported reduction levels?	**Satisfaction:** Are the participants satisfied with their experience? Are they satisfied with the solution? What could be improved?
Magnitude of reductions in aggregate: How much was demand reduced? If there were statewide active demand reduction goals, did the PAs collectively meet them?	**PA satisfaction:** How satisfied are the PAs with their groups of vendors? Were there any challenges or above-and-beyond efforts in design, recruitment, data collection, and data transfer between the parties?
Efficiency: What are the net energy usage implications of the solution, if any?	**Barriers:** Are there technological, economic, or regulatory barriers to deployment of the solutions? Are these barriers affecting number of participants or load reduction per participant? How can these barriers be overcome?
Multiple DR program interaction: How much overlap is there between the PA ADR initiatives and the ISO-NE market? How does it impact the PA ADR initiative performance?	
Potential peak demand management opportunities: What technologies are being used for event-based response? Are there other peak demand opportunities that the utilities should pursue at participating facilities beyond those sought through these initiatives (reducing ISO-NE system peak demand and in turn decreasing capacity needs, costs for all customers, and peaker plant GHG emissions)? What other opportunities do customers have to be compensated for peak demand management in the market?	

3 Methodology and Framework

3.1 Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

3.1.1 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative[6]
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives[7]

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned
- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs

[6] ISO-NE offers their own DR programs through their Price Responsive Demand initiative. Some Eversource, National Grid, and Unitil PA DR initiative participants also enroll in ISO-NE DR programs. . Participating in multiple programs during the same season has ramifications regarding both baseline calculations and resource reliability.

[7] Also note that the CSPs who implement participation in the DR initiatives in multiple states gave perspectives about the initiative as implemented in those states.

- Other opportunities for peak demand management

3.1.2 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the “Unitil” stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group[8], were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

[8] Eversource battery participants are an exception, as the evaluation team received responses from both battery participants.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states[9], to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

[9] The team asked those with sites in multiple states the following question: We understand your organization has participating sites in multiple states (i.e., [LIST STATES]). Does the program experience that you've described above (that is, when you provided answers to our previous questions), reflect your program participation experience across states?

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

3.2 Impact Evaluation Methodology and Framework

This section provides the methodologies used for the C&I Demand Response efforts and the C&I Storage efforts. The demand response section starts with a discussion of the challenges of estimating demand response to put the subsequent discussion of the baseline options in context. Additional discussion of the baselines and their implications occurs in Sections 4.2 and 4.3, the Impact Evaluation Findings and Integrated Impact and Process Evaluation Findings. A data sufficiency section discusses issues related to the evaluation team's recommended statewide data sufficiency and calculation rules for consistency of results across PAs.

3.2.1 C&I Demand Response

The Challenge of Estimating Load Reduction

In the discussion that follows, we review the challenges of estimating load reduction, the baseline methodologies used in this evaluation to estimate load reduction, load characteristics that affect estimation accuracy, and the relationship between program design features and baseline choice and accuracy. Program design and settlement[10] M&V is an iterative process of ongoing re-assessment and refinement. With proper program design and M&V methods, demand response can be a reliable, measurable, and verifiable resource.

A fundamental difference between load reduction and generation as a resource is that it is not possible to directly observe load reduction. Instead, measurement of load reduction necessarily means comparing observed load to a counterfactual load that would have occurred in the absence of dispatch. This counterfactual load or baseline is the estimate of load had the event not occurred. The difference between the baseline and observed load during the event period is the load reduction. The load reduction is positive if the observed load is less than the baseline and negative if the observed load is greater. Because the baseline is an estimate, there is also an

[10] Settlement is the determination of load reduction achieved and payments from the PA to the CSP or participant.

associated error.[11] Random estimation errors are expected to be distributed around zero, with positive errors being indistinguishable from the actual load reduction that is occurring at most sites. Negative errors (where the baseline is higher than actual load) may be visible if the customer has not made any load reduction effort.

Load reduction is a function of many factors:

- **The nature of end use under control.** End uses consigned to reduction can include fixed (e.g., lighting), variable (e.g., HVAC), or either (e.g., process). Each can have its own unique level of achievable reduction, duration of reduction, and limitations in availability.
- **The nature of load reduction engagement**. Means of load control can include automatic, external, control-based, or manual. Each has implications in asset reliability, responsiveness, lag time, and tendency of override.
- **The state of site-level load on event day.** The load shape and consumption levels on an event day relative to other days can be a key determinant of the effectiveness of settlement-type baselines.[12] The weather, probability of shutdown, and behaviors that effect consumption in the pre-event window can be individual or collective factors in load reduction.
- **The state of site-level load on baseline days.** Instances of shutdowns, milder weather, and different production schedules have implications for the development of the load reduction counterfactual.

All of these factors can change day-to-day and event-to-event. In general, despite the dynamic and complex nature of the process, the only information the evaluator has is site-level load. Settlement-type baselines use the site-level load in recent weeks to develop typical load shapes for the site. They also use event-day load to account for the actual conditions at the site on the day of the event. The result is an approach that, with remarkably little variation, is used as a basis of settlement for ISO markets across the country and internationally. Under many conditions, these settlement-type baselines can produce accurate estimates of load reduction, but it is worth remembering that even with those accurate estimates we are unlikely to know

[11] Baseline analyses compare estimated baselines to observed load on days without events. This makes it possible to quantify the error. Across all non-event days, most unadjusted or symmetrically adjusted baselines have errors that are approximately centered around zero, though the errors can be substantial, sometimes even bigger than the expected load reduction. On the subset of extreme days, errors are much less likely to be centered around zero.

[12] Settlement-type baselines are used to determine the load reduction quantity that is used as the basis for performance-based incentives. Generally, settlement-type baselines define baseline load as an average of load during a specified time period occurring prior to an event. Settlement-type baselines are preferred to other impact evaluation methods, such as regression analysis, due to transparency, simplicity, and ability to estimate load reduction within a short time period after an event.

what actions occurred, the process for making those actions occur, and whether, under conditions different than those specific days on which an event occurred, load reduction will occur in a similar fashion. This latter challenge primarily affects the ability to forecast load reduction.

When considering a forecast of what kind of load reduction from an ADR initiative can be expected in the future, we also have to be particularly aware that the evaluator and PA know little about what is occurring at any individual site regarding what end uses may be controlled, frequency and nature of control, and how wider conditions may affect either actual load reduction or the evaluator's ability to estimate that load reduction with a settlement-type baseline. Estimates of load reduction from a single day will reflect the unique conditions of that day only. Estimates of load reduction from multiple days will provide a more informed, generalized basis for a forecasted load reduction. Only Eversource called more than one event in summer 2019.

The limited number of events in summer 2019 raises the question what load reduction estimates are most appropriate to project forward. The load reduction estimates from the current summer's event days reflect the mix of these many factors on those limited number of days (in this case, between 1 and 3 event days, depending on the PA). For most of these factors, there is little choice but to accept this year's conditions to be representative of the future year's condition. For example, the share of customers who successfully provided load reduction is already embedded in the settlement load reduction estimates. While this response rate could be quite different going forward, this year's performance, including the non-performers, is the only empirical evidence of a general response rate available to be projected forward. Similarly, we don't know what weather conditions will be at the future peak, but this year's peak conditions are the most reasonable estimate of next year's peak conditions.[13] Under the current conditions, the one option for adjusting the forecast involves accounting for shutdowns. Rather than assume that the shutdowns that happened to occur on this summer's event day(s) are representative, we can apply probabilistic estimate of shutdown based on the full 2019 summer of load data. More generally, if the initiative is interested demonstrating its ability to provide load reduction in the future, additional event days will provide a sounder basis for those estimates.

Baselines

The evaluation team estimated impacts of the ADR initiative by comparing observed load on event days with five different baselines. Each baseline's calculation methods, advantages, and

[13] DR programs with many events can establish a relationship between load reduction estimates and weather if one exists. This can allow projections across a range of conditions or at some specified design conditions. This kind of extrapolation is not possible with just a few event days.

disadvantages are described below. The ADR initiative uses an asymmetrically adjusted rolling 10-of-10 baseline as the basis for settlement payments from the PAs to CSPs. This is the metric that is listed as 'Reported -Asymmetric' in the impact summary tables.

- **Evaluated-Unadjusted: Unadjusted rolling 10-of-10 baseline.** This settlement-type baseline constructs a pool of the 10 most recent eligible days occurring prior to the day for which load is being estimated. For the ADR initiative, eligible days include non-holiday weekdays that a demand response event did not occur on.[14] The baseline shape is calculated as the rolling average of load in each interval across the 10 most recent eligible days.

  Initiative rules also allow scheduled shutdowns[15] to be excluded from the baseline pool. For a scheduled shutdown to be excluded, the program administrator must be notified at least one week in advance. A maximum of 10 scheduled shutdowns are allowed per season in Massachusetts and Connecticut, and a maximum of 14 is allowed in New Hampshire. As discussed in the findings, no shutdowns were reported prior to events during the summer 2019 season.

  When missing data is present among the 10 most recent baseline days, additional lookback days may be used to construct a pool of 10 baseline days with sufficient data to estimate load. Details of baseline calculations (e.g., use of lookback days, missing data criteria for lookback days, etc.) vary by PA and can found in Appendix C: Settlement Verification. For consistency across results, the evaluation team applied a uniform set of data criteria and related baseline calculation details, discussed later, in the

Enrolled Capacity	Asset-level expected load reduction. Generally estimated by CSP.
Reported: Asymmetric	Load reduction claimed reported by program implementers.
Evaluated: Validation	Evaluation team's attempt to replicate reported asymmetric results. Baseline calculation details (e.g., lookback days) vary by PA.
Evaluated: Unadjusted	Rolling average of load in each interval across the 10 most recent eligible days.
Evaluated: Asymmetric	Shifts the unadjusted baseline upward to meet observed load during pre-event adjustment period. Unadjusted baseline is used in lieu of downward adjustments.
Evaluated: Symmetric	Shifts the unadjusted baseline upward or downward to meet observed load during pre-event adjustment period
Evaluated: Forecast	Modification of evaluated symmetric results by accounting for probability of unreported shutdowns.
Evaluated: Regression	Site-level model of load across the summer. Specification describes load as a function of cooling degree-days, weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit.

[14] Demand response events for the ADR initiative are only called on weekdays; however, the ICAP hour was forecasted on Saturday, July 20, 2019, due to peak load conditions. In response, Eversource called an extraordinary (voluntary) event. For Saturdays, the rolling baseline constructs a pool of the 5 most recent eligible Saturdays.

[15] A scheduled shutdown results in reduced demand. This initiative rule allows for baselines and performance to be unaffected by anomalous shutdown load.

In each PA-specific impact findings section, and at the end of the impact findings section, the evaluators have also calculated performance ratios for multiple combinations of metrics. None of the performance metrics include accounts with insufficient data. The four ratios are:

- Enrollment Ratio: This ratio is the reported asymmetric load reduction to the CSP reported enrolled capacity. This ratio provides insight into what percentage of the reported enrolled capacity was achieved, based on the program baseline and calculation methodology. This ratio is particularly meaningful for planning and sales purposes.
- Asymmetric Ratio: This ratio is the evaluated asymmetric load reduction to the reported asymmetric load reduction. This is an apples to apples comparison of the same baseline methodology between the PAs and evaluators, however, this metric identifies the impact that different calculation rules between the PAs and evaluators has on load reduction.
- Retrospective Realization Rate: This ratio is the evaluated symmetric load reduction to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline is the most appropriate measure of retrospective load reduction for the 2019 summer season. This ratio shows how the choice of baseline adjustment and calculation methodologies impacts the load reduction estimates. The evaluators recommend using this realization rate to calculate the symmetric load reductions at the end of future seasons if there are no evaluations conducted.
- Prospective Realization Rate: This ratio is the evaluated symmetric load reduction with an adjustment for unreported shutdowns to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline accounting for unreported shutdowns is the most appropriate measure of prospective load reduction for future seasons. This ratio provides insight into the magnitude of reductions that could be achieved during future seasons as a function of the validated load reduction estimates. The prospective realization rate should only be used as an ex-ante estimate of future performance for planning purposes and not retrospectively.
    - The product of the enrollment ratio and prospective realization rate is a useful planning metric. It indicates of the percentage of expected future load reduction available from future CSP reports of enrollment (as opposed to PA reports of it).

Data Sufficiency section.

The unadjusted 10-of-10 is an effective baseline for customers with consistent, non-weather-sensitive load, has minimal potential for baseline manipulation, and allows preparatory decreases in load (e.g., shift cancellation) without a negative effect on the estimated load reduction. However, the unadjusted 10-of-10 is inflexible for customers with weather-sensitive load or load variability unrelated to weather and can substantially understate load reduction for customers with weather sensitivity if events are called on extreme hot days. The ADR Initiative calls events to mitigate load during the annual system peak hour, which is correlated with extreme hot days.

- **Evaluated-Symmetric: Rolling 10-of-10 baseline with symmetric additive adjustment.** This settlement-type baseline shifts the unadjusted 10-of-10 to meet observed load during pre-event hours (adjustment period) using a symmetric additive adjustment. For an additive adjustment, the magnitude of the shift for all intervals equals the difference between observed load and the unadjusted baseline during the adjustment period.[16] Symmetric adjustments may shift the unadjusted baseline upward or downward. Generally, however, the resulting baseline may not be less than zero (with the exception of some assets that have behind-the-meter generation).[17]

  The symmetrically adjusted baseline generally improves the accuracy of baseline estimates by accounting for load changes due to weather and variable operations on the event day. Symmetric day-of-event adjustments with prior-day event notification make it possible that the pre-event period used for the adjustment will not be representative of a non-event day. The ADR initiative provides day-ahead event notification; the adjustment period is the hour beginning 2 hours prior to the event period. Examples of event effects during the adjustment period include pre-cooling, shift cancellation, and baseline manipulation.

  The rolling 10-of-10 baseline with a symmetric additive adjustment reduces biases for customers with weather-sensitive load or load variability unrelated to weather. However, this baseline can substantially understate load reduction for customers that decrease load in anticipation or preparation for an event (e.g., shift cancellation), while pre-cooling can inflate the adjusted baseline for weather-sensitive customers. The use of day-ahead event notification increases the potential that load during the adjustment period will be affected by the event, including baseline manipulation. Deliberate ramp-ups may be difficult to distinguish from preparatory load changes. Same-day notification reduces the potential for baseline manipulation and variability due to pre-cooling. ISO-NE uses this baseline with day-of-event notification.

- **Evaluated-Asymmetric: Rolling 10-of-10 baseline with asymmetric additive adjustment.** This settlement-type baseline only applies the additive adjustment if it will increase the

[16] In contrast, a multiplicative adjustment scales the unadjusted baseline by the ratio of observed load to the unadjusted baseline during the adjustment period.

[17] For assets with behind-the-meter generation that often provide net supply to the grid (negative load), a baseline floor of zero can overstate load reduction. Alternatively, the asset's interconnection agreement may be used to establish the baseline floor.

baseline. If the adjustment is negative, the unadjusted baseline is used. The asymmetric adjustment helps in program implementation because it avoids penalizing customers that decrease load in preparation for an event (e.g., shift cancellation). However, it fails to make downward adjustments for customers where load is less than baseline due to event-day conditions unrelated to legitimate actions taken in preparation for the event. By design, this replaces lower or more negative load reduction estimates with higher estimates and creates an upward bias in initiative-level load reduction estimates.

This baseline is used by the PAs as the basis for settlement. The associated load reduction may not be greater than the maximum load during the baseline period. When reporting the results, the PAs treat negative event reduction as a failure to perform and include as zero in reported performance results (**Reported-Asymmetric**). These negative load reduction estimates represent natural variation in random estimation error and zeroing them out introduces an upward-bias in initiative-level load reduction estimates. As discussed here and in the asymmetric adjustment section above, downward adjustments and negative results are, in fact, part of the error process for the estimates and should be included to provide the most accurate estimates of aggregate load. The **Reported-Asymmetric** also include other baseline calculation details that vary by PA (e.g., lookback days) and can found in Appendix C: Settlement Verification. The reported results are provided at the event-level.

The **Evaluated-Validation** set of results is the evaluation team's attempt to replicate the reported performance results using each PA's unique baseline calculations and should be consistent with the initiative-reported results. The evaluated-validation results are calculated at the event level for consistency with the reported-asymmetric results.

Note, assets with highly variable loads uncorrelated with weather are a challenge for any demand response evaluation methodology. For such assets, estimated load reduction and incentives may have limited relationship with actual load reduction actions taken. The upward bias inherent in the asymmetrically adjusted baseline increases the potential for misalignment of incentives and actual load reduction actions taken for these kinds of loads that are participating in the offering. Establishing a formal criterion for predictability of an asset's load is one way to ensure that incentives are aligned with demand reduction actions taken, improve cost-effectiveness, and estimation of impacts.

- **Evaluated-Regression: Regression baseline.** This baseline fits a regression model to an individual customer's load data across the entire season. The regression specification describes load for each hour of the day as a function of cooling degree-days (CDD), weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit. The model is applied to event day conditions without the event day terms in effect to estimate load on that day absent the event. The regression can be an effective baseline for customers with either stable or weather-sensitive load and can control for weather without a day-of-event adjustment. Because the ex-post regression analysis does not have a day-of-event adjustment, it has the potential to facilitate analysis of changes in consumption that occur before and after the event. However, because the regression summarizes all data from the summer, it can systematically underestimate load reduction for customers with unscheduled or unreported shutdowns and is erratic for customers with high load variability unrelated to weather (e.g., operational variability). In the absence of an alternative for the numerous assets that the regression baseline understates load reduction, the regression analysis is not considered the best M&V method of ex-post impact estimation and cannot facilitate ex-ante estimates.

The regression specification features hourly baseload flexible to the month with a consistent hourly cooling trend across the summer. The baseload specification allows the regression to account for some variability in production across the summer though not dramatic shifts in load.

$$L_{idh} = May_h + Jun_h + Jul_h + Aug_h + Sep_h + \beta_h^C C_o + E_{idh} + Wknd_h(1 + \beta_h^C C_o + E_{idh}) + \varepsilon_h$$

Where:

L_{idh}	= Load for a given customer i, day d and hour h
$May_h,\ etc$	= Baseload for hour h in month May, etc
C_o	= Cooling degree days from an optimized degree-day base
β_h^C	= Cooling trends for hour h as a function of CDD
E_{idh}	= Event day load for customer i, event day d, and hour h
$Wknd_h$	= Change in baseload for hour h on weekends and statutory holidays
ε_{it}	= Regression residual

- **Evaluated-Forecast: Rolling 10-of-10 baseline with symmetric additive adjustment and adjustment to account for unreported shutdown days.** A key goal of this evaluation is to produce a best estimate of expected future load reduction. This provides the basis for a prospective realization rate that can be applied to future results absent evaluation. The accuracy of the symmetrically adjusted 10-of-10 baseline, with the limited evidence of pre-cooling or gaming makes it the best choice for the forecast baseline. As discussed above, the estimates from this year already encapsulate our best information regarding aspects of future performance – active response rates, weather, etc. A shutdown adjustment is the one additional adjustment that can be applied to better reflect expected conditions.

Shutdown conditions will vary from site to site. For some sites, we assume there is a low load level that is distinct from typical active load, when it would be unlikely that load reduction would occur. We only identify these days if they were reasonably distinct from the bulk of days. Specifically, using the same hour of the day as used for same-day adjustments (1 to 2 p.m.), we flagged low loads that were clustered below the median load for that hour with separation from the remaining loads equal to at least 20% of the median load. Requiring this substantial break between the shutdown days and typical days allowed us to avoid making purely arbitrary breaks where there was no evidence of differentiation in the data. This approach was effective at identifying the common shutdown days around July 4, for instance.

Savings and Realization Rates

The evaluators computed savings from different perspectives by comparing measured load with each of the above baseline loads.

Table 3-3 summarizes the performance metrics that are available and frequently referenced throughout these findings. Reported asymmetric baseline calculation details may vary by PA. The evaluated validation is the impact evaluation team's attempt to replicate the reported results. Details regarding the reported asymmetric and the evaluated validation can be found in Settlement Verification. All other evaluated results rely on missing data guidelines and baselines calculation details established by the impact evaluation team for consistency across results. Details regarding these can be found immediately below in

Enrolled Capacity	Asset-level expected load reduction. Generally estimated by CSP.
Reported: Asymmetric	Load reduction claimed reported by program implementers.
Evaluated: Validation	Evaluation team's attempt to replicate reported asymmetric results. Baseline calculation details (e.g., lookback days) vary by PA.
Evaluated: Unadjusted	Rolling average of load in each interval across the 10 most recent eligible days.
Evaluated: Asymmetric	Shifts the unadjusted baseline upward to meet observed load during pre-event adjustment period. Unadjusted baseline is used in lieu of downward adjustments.
Evaluated: Symmetric	Shifts the unadjusted baseline upward or downward to meet observed load during pre-event adjustment period
Evaluated: Forecast	Modification of evaluated symmetric results by accounting for probability of unreported shutdowns.
Evaluated: Regression	Site-level model of load across the summer. Specification describes load as a function of cooling degree-days, weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit.

In each PA-specific impact findings section, and at the end of the impact findings section, the evaluators have also calculated performance ratios for multiple combinations of metrics. None of the performance metrics include accounts with insufficient data. The four ratios are:

- Enrollment Ratio: This ratio is the reported asymmetric load reduction to the CSP reported enrolled capacity. This ratio provides insight into what percentage of the reported enrolled capacity was achieved, based on the program baseline and calculation methodology. This ratio is particularly meaningful for planning and sales purposes.

- Asymmetric Ratio: This ratio is the evaluated asymmetric load reduction to the reported asymmetric load reduction. This is an apples to apples comparison of the same baseline methodology between the PAs and evaluators, however, this metric identifies the impact that different calculation rules between the PAs and evaluators has on load reduction.
- Retrospective Realization Rate: This ratio is the evaluated symmetric load reduction to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline is the most appropriate measure of retrospective load reduction for the 2019 summer season. This ratio shows how the choice of baseline adjustment and calculation methodologies impacts the load reduction estimates. The evaluators recommend using this realization rate to calculate the symmetric load reductions at the end of future seasons if there are no evaluations conducted.
- Prospective Realization Rate: This ratio is the evaluated symmetric load reduction with an adjustment for unreported shutdowns to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline accounting for unreported shutdowns is the most appropriate measure of prospective load reduction for future seasons. This ratio provides insight into the magnitude of reductions that could be achieved during future seasons as a function of the validated load reduction estimates. The prospective realization rate should only be used as an ex-ante estimate of future performance for planning purposes and not retrospectively.
    - The product of the enrollment ratio and prospective realization rate is a useful planning metric. It indicates of the percentage of expected future load reduction available from future CSP reports of enrollment (as opposed to PA reports of it).

Data Sufficiency.

Table 3-3. Summary of Performance Metrics

Enrolled Capacity	Asset-level expected load reduction. Generally estimated by CSP.
Reported: Asymmetric	Load reduction claimed reported by program implementers.
Evaluated: Validation	Evaluation team’s attempt to replicate reported asymmetric results. Baseline calculation details (e.g., lookback days) vary by PA.
Evaluated: Unadjusted	Rolling average of load in each interval across the 10 most recent eligible days.
Evaluated: Asymmetric	Shifts the unadjusted baseline upward to meet observed load during pre-event adjustment period. Unadjusted baseline is used in lieu of downward adjustments.
Evaluated: Symmetric	Shifts the unadjusted baseline upward or downward to meet observed load during pre-event adjustment period
Evaluated: Forecast	Modification of evaluated symmetric results by accounting for probability of unreported shutdowns.

Evaluated: Regression	Site-level model of load across the summer. Specification describes load as a function of cooling degree-days, weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit.

In each PA-specific impact findings section, and at the end of the impact findings section, the evaluators have also calculated performance ratios for multiple combinations of metrics. None of the performance metrics include accounts with insufficient data. The four ratios are:

- Enrollment Ratio: This ratio is the reported asymmetric load reduction to the CSP reported enrolled capacity. This ratio provides insight into what percentage of the reported enrolled capacity was achieved, based on the program baseline and calculation methodology. This ratio is particularly meaningful for planning and sales purposes.
- Asymmetric Ratio: This ratio is the evaluated asymmetric load reduction to the reported asymmetric load reduction. This is an apples to apples comparison of the same baseline methodology between the PAs and evaluators, however, this metric identifies the impact that different calculation rules between the PAs and evaluators has on load reduction.
- Retrospective Realization Rate: This ratio is the evaluated symmetric load reduction to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline is the most appropriate measure of retrospective load reduction for the 2019 summer season. This ratio shows how the choice of baseline adjustment and calculation methodologies impacts the load reduction estimates. The evaluators recommend using this realization rate to calculate the symmetric load reductions at the end of future seasons if there are no evaluations conducted.
- Prospective Realization Rate: This ratio is the evaluated symmetric load reduction with an adjustment for unreported shutdowns to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline accounting for unreported shutdowns is the most appropriate measure of prospective load reduction for future seasons. This ratio provides insight into the magnitude of reductions that could be achieved during future seasons as a function of the validated load reduction estimates. The prospective realization rate should only be used as an ex-ante estimate of future performance for planning purposes and not retrospectively.
  - The product of the enrollment ratio and prospective realization rate is a useful planning metric. It indicates of the percentage of expected future load reduction available from future CSP reports of enrollment (as opposed to PA reports of it).

Data Sufficiency

This section offers a high-level introduction to the evaluation team's recommended statewide data sufficiency and calculation rules. These are a uniform set of rules that were established for consistency in evaluated results across PAs, whereas results reported by the PAs vary in terms of data sufficiency and calculation rules. In turn, the methods used by the evaluation team in an attempt to replicate the reported results varies by PA. Relevant details of the data sufficiency and calculation rules used by the PAs in their reported results can be found in the Appendix C: Settlement Verification. Additionally, this section summarizes the quality of the interval meter data received and how it relates to the results presented in the Impact Evaluation Findings.

For rolling baselines, when missing data are present among the 10 most recent baseline days, additional lookback days may be used. If more than 25% of intervals are missing load data during any hour between 1 p.m. and 7 p.m., an additional lookback day is used as a substitute. Additional lookback days are used until each of the 10 baseline days has met this criterion. The maximum lookback for baseline pool days is 42 calendar days.

For the regression analysis, if more than 25% of intervals are missing load data during an hour, the hour is dropped from the hourly site-level regression model.

Ultimately, if either the baseline pool or the event day are missing load data in more than 20% of intervals during the adjustment and event periods, the data is considered insufficient to estimate load reduction.

The evaluation team used enrolled capacity, reported performance data, and interval meter data provided by the PAs for the impact evaluation. In a few instances, the PA provided interval meter data from the vendor CSP. Table 3-4, below, summarizes the extent to which missing load data impacted the results for Eversource in Massachusetts. For each event, the total enrolled capacity, reported asymmetric results, and number of associated accounts are split into two sub-categories: accounts for which the evaluation team has insufficient data, as defined in the paragraph immediately above, to estimate event performance and accounts for which the evaluation team has sufficient data to estimate performance. For example, for the July 19th event, 17 Eversource Massachusetts accounts were missing load data for more than 20% of intervals from 1 p.m. to 7 p.m. in the baseline pool (includes attempt to use lookback days) and/or on the event day. These 17 accounts with insufficient load data to estimate event performance had an enrolled capacity of 2,751 MW and reported asymmetric performance of 633 MW. Alternatively, the evaluation team received sufficient load data to estimate load reduction for 160 Eversource Massachusetts account on July 19th. These accounts had an enrolled capacity of 37,223 MW and a reported asymmetric performance of 18,155 MW. Collectively, the two subsets compose the Eversource Massachusetts ADR initiative.

Note, results presented in the Impact Evaluation Findings do not include results for accounts that the evaluation team cannot estimate load reduction due to insufficient data. The intent is to present apples-to-apples comparisons.

Each of the tables that follow in this section detail data sufficiency for each PA and state. With the exception of Eversource, July 30 was the only called event during the summer of 2019. Due to the varying number of accounts that have sufficient data to estimate load reduction for each event, overall summer results are calculated at the account-level and then aggregated to the initiative-level for Eversource. In other words, each account with sufficient data to estimate load reduction for at least one event is weighted equally in the overall results.

With the exception of National Grid, data sufficiency criteria and baseline calculations used by the evaluation team results in the same number of evaluable account events as those used by the PA. In the case of National Grid, missing data during the most recent 10 eligible baseline days were insufficient to estimate load reduction. National Grid did not use lookback days. In 7 instances, were a lookback period used, a baseline pool with sufficient data could have been constructed. These 7 account events were ultimately not included in the evaluated results in the Impact Evaluation Findings in order to have a consistent basis on which to apply realization rates.

Table 3-4. Eversource - Massachusetts - Data Sufficiency

Event Date	7/19/2019	7/30/2019	8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Enrolled Capacity - Total	39,974	39,974	39,974
Enrolled Capacity - Insufficient	2,751	3,291	6,104
Enrolled Capacity - Sufficient	37,223	36,683	33,870
Reported - Asymmetric - Total	18,788	26,426	24,490
Reported - Asymmetric - Insufficient	633	1,437	2,353
Reported - Asymmetric - Sufficient	18,155	24,990	22,137
Accounts - Total	**177**	**177**	**177**
Accounts - Insufficient	**17**	**26**	**42**
Accounts - Sufficient	**160**	**151**	**135**

Table 3-5. Eversource - New Hampshire - Data Sufficiency

Event Date	7/19/2019	7/30/2019	8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Enrolled Capacity - Total	5,905	5,905	5,905
Enrolled Capacity - Insufficient	-	-	3,225
Enrolled Capacity - Sufficient	5,905	5,905	2,680
Reported - Asymmetric - Total	4,988	5,552	7,114
Reported - Asymmetric - Insufficient	-	-	5,300
Reported - Asymmetric - Sufficient	4,988	5,552	1,814
Accounts - Total	**40**	**40**	**40**
Accounts - Insufficient	**-**	**-**	**21**
Accounts - Sufficient	**40**	**40**	**19**

Table 3-6. Eversource - Connecticut - Data Sufficiency

Event Date	7/19/2019	7/30/2019	8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Enrolled Capacity - Total	12,537	12,537	12,537
Enrolled Capacity - Insufficient	68	18	288
Enrolled Capacity - Sufficient	12,469	12,519	12,249
Reported - Asymmetric - Total	12,575	13,135	13,391
Reported - Asymmetric - Insufficient	-	-	306
Reported - Asymmetric - Sufficient	12,575	13,135	13,085
Accounts - Total	**97**	**97**	**97**
Accounts - Insufficient	**2**	**1**	**2**
Accounts - Sufficient	**95**	**96**	**95**

Table 3-7. National Grid - Massachusetts - Data Sufficiency

Result	Event Average Reduction (kW)
Enrolled Capacity - Total	98,618
Enrolled Capacity - Insufficient	5,484
Enrolled Capacity - Sufficient	93,134
Reported - Asymmetric - Total	74,271
Reported - Asymmetric - Insufficient	2,843
Reported - Asymmetric - Sufficient	71,428
Accounts - Total	**392**
Accounts - Insufficient	**19**
Accounts - Sufficient	**357**

Table 3-8. Unitil - Data Sufficiency

Result	Event Average Reduction (kW)	Event Average Reduction (kW)
State	**MA**	**NH**
Enrolled Capacity - Total	950	1,600
Enrolled Capacity - Insufficient Data	-	-
Enrolled Capacity - Sufficient Data	950	1,600
Reported - Asymmetric - Total	853	1,299
Reported - Asymmetric - Insufficient Data	-	-
Reported - Asymmetric - Sufficient Data	853	1,299
Accounts - Total	**3**	**7**
Accounts - Insufficient Data	**-**	**-**
Accounts - Sufficient Data	**3**	**7**

3.2.2 Targeted Battery Storage

The impact evaluation used battery meter data as the measure for demand reduction. Storage options are listed in Table 3-9. The dispatch schedule called for 3 hours of demand response between 3:00 p.m. and 7:00 p.m. per non-holiday weekday. Performance and payment are based on average demand reduction (kW) during event hours during the season.

Table 3-9. Targeted Battery Storage Dispatch Program Options

Type	Targeted – Summer
Season	June 1 – Sept 30
Max. number of events	8
Event duration (hours)	3

For each storage customer, and in aggregate, the evaluation team quantified the following:

- Demand reduction at the ISO-NE system peak hour
- Average demand reduction across events (all summer non-holiday weekdays for daily dispatch and event dispatches for targeted dispatch)
- Demand reduction for the Eversource weekend event
- Effective reduction of customer monthly peak kW
- Energy reduction by event-hour (including weekend event)
- Total net kWh effect of battery activity at the site over the DR season

If the battery data tracked discharge as positive and charging as negative, then the battery data was added to the interval data to produce a “no-battery” counterfactual. Site-level consumption was observed to decrease during periods when the battery was discharging and increase when it was charging. The evaluation team calculated the total season charge and discharge energy as well as the seasonal efficiency of the battery system.

4 FINDINGS

This section provides evaluation findings. It starts with process evaluation findings, moves to impact evaluation findings and then concludes with a section on integrated process and impact findings.

4.1 Process Evaluation Findings

This chapter documents findings from the process evaluation of the DR Program. The chapter is organized in the following way:

- Section 4.1.1 discusses the cross-cutting and state-level findings.
- Section 4.1.2 discusses the ISO-NE and PA ADR initiative overlap and interplay
- Section 4.1.3 discusses feedback received from Eversource program actors (PA staff, CSPs, and ISO-NE staff) and program participants.
- Section 4.1.4 discusses feedback received from National Grid program actors (PA staff, CSPs, and ISO-NE staff) and program participants.
- Section 4.1.5 discusses feedback received from Unitil program actors (PA staff, CSPs, and ISO-NE staff) and program participants.

4.1.1 Cross-Cutting: All Administrators and States

The PA and CSP interview data as well as DR initiative documentation review revealed two notable cross-cutting findings:

1. Requiring that participants sign a PA application in addition to a separate CSP-customer agreement, both with their own T&Cs, complicates and slows the sales process for all CSPs and administrators, and across states (see additional details in Appendix B, Applications and Contracts Section).
2. Other than lack of winter curtailment options in New Hampshire, the team found no indication that the administration of the DR initiative varied by state.

The team also examined participant survey responses to assess whether any aspects of the initiative might be different or are perceived differently in different jurisdictions. The team found no indication that customer experience varied by state (see additional details in Appendix A, State-Level Detailed Results Section). However, interpret this finding with caution since DR initiative participant populations in Connecticut (n=20) and New Hampshire (n=13) were small, and thus very few Connecticut (n=7) and New Hampshire (n=5) initiative participants responded to our survey. These small samples impeded the team's ability to more robustly compare customer survey responses by state.

4.1.2 ISO-NE and PA ADR Initiative Overlap

The first part of this section provides a high-level description of the ways in which resources can bid into the ISO-NE forward capacity market (FCM) and the responsibilities they take on when doing so. The evaluation team has included this section to further describe and clarify the context in which the overlap of the PA ADR initiative and the FCM takes place. Later subsections explore ISO-NE staff[18], CSP, and PA perspectives and concerns about this overlap.

ISO-NE Forward Capacity Market Description

Customers who participate in the ADR initiative (either through load reductions or with battery storage systems) are also able to bid into and receive compensation from participating in the ISO-NE forward capacity market (FCM). The ISO refers to individual customers as Demand Response Assets (DRAs). DRAs cannot directly participate in the FCM, however – DRAs must be "mapped to," or grouped together to form, a Demand Response Resource (DRRs), which *are* able to directly participate in the FCM. Most often[19], CSPs aggregate DRAs into DRRs and bid into the FCM.

An active DRR bids load reduction (in MW) into the capacity market at a chosen price. If the market price goes above a certain threshold (bid clears), the DRR will be awarded a capacity supply obligation (CSO) for the MW reduction bid amount. By selling capacity into the FCM and receiving a CSO, a DRR has assumed responsibility for a share of system requirements and is expected to dispatch to its full CSO capability during capacity scarcity conditions (CSCs). CSCs[20] occur when the electricity supply is constrained and reserve capacity (i.e. backup electricity supply) is scarce. The ISO compensates CSOs for assuming this responsibility through monthly incentive payments, called Monthly Capacity Payments. Monthly Capacity Payments comprise two payment streams: a base payment and a performance payment.

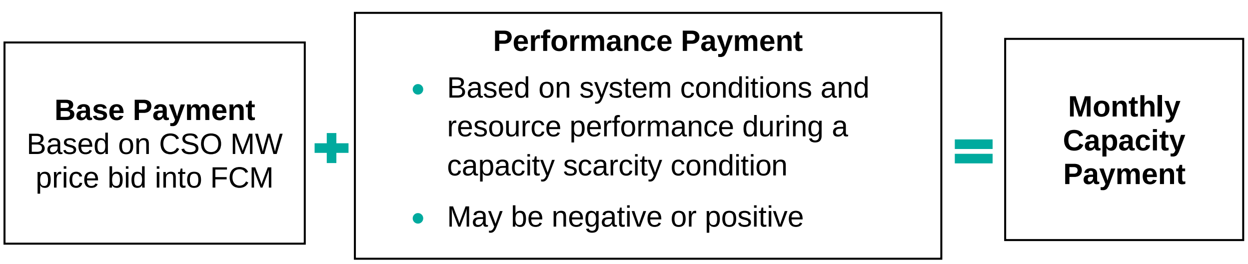
[18] The evaluators interviewed one member of the ISO's Demand Resource Strategy, Market Operations and Asset Registration and Auditing Teams (three staff in total) in January of 2019, and conducted a second interview with a member of the ISO's Market Operations Team in January of 2020.

[19] Instances where one customer or DRA forms its own DRR and bids into the FCM independently are rare.

[20] Definition of capacity scarcity condition from ISO NE Market Rule 1: A Capacity Scarcity Condition shall exist in a Capacity Zone for any five-minute interval in which the Real-Time Reserve Clearing Price for that entire Capacity Zone is set based on the Reserve Constraint Penalty Factor pricing for: (i) the Minimum Total Reserve Requirement; (ii) the Ten-Minute Reserve Requirement; or (iii) the Zonal Reserve Requirement, each as described in Section III.2.7A(c); provided, however, that a Capacity Scarcity Condition shall not exist if the Reserve Constraint Penalty Factor pricing results only because of resource ramping limitations that are not binding on the energy dispatch.

- The **base payment** is the bid clearing price per MW of the CSO and is paid monthly whether or not a scarcity condition occurs. (It is possible that a DRR could receive a monthly payment without ever being required to reduce load.)
- The **performance payment** is based on the load reduction a DRR provides relative to their obligated share of system requirements (i.e., their CSO MW) during scarcity conditions. ISO-NE uses its standard adjusted 10 of 10 baseline to determine performance during a scarcity condition. If a DRR over-performs its CSO, the ISO provides an incentive. Under-performance results in a penalty charge. A few notes:
  - The ISO refers to this as pay-for-performance. It is different from the PA ADR initiative pay-for-performance in that if a DRR delivers less than their obligated share of system requirements during a scarcity event, they will incur a penalty payment from the ISO.[21]
  - The Monthly Capacity Payment includes a performance payment element only if a scarcity condition occurs within any given month; if a scarcity condition does not occur, the Monthly Capacity Payment only includes the base payment.

Figure 4-1. CSO Monthly Capacity Payment Components



A DRR with a CSO must participate in the wholesale energy and reserve markets:

- **Wholesale energy market.** The wholesale energy market is a system for purchasing and selling electric energy using supply and demand to set the price of electricity.
- **Reserve market.** The reserve market is the backup capacity available for dispatch in case of unexpected generation or supply shortages.[22]

Put simply for the purposes of this report, this means that a DRR must bid into the energy market as if it were a generation resource. The ISO takes note of this bid in their tracking

[21] In this sense, "performance payment" is a misnomer. It could very well be a performance *penalty*, or, in cases where DRRs perform perfectly/as obligated, they are neither credited nor charged.

[22] The ISO defines the reserve market as "capacity available for dispatch during system contingencies, which are unplanned disconnections of power system elements, such as transmission facilities or generators, from the electricity grid." https://www.iso-ne.com/markets-operations/settlements/understand-bill/item-descriptions/reserve-market

system, noting the bid price and MW offer. Using these factors, among others, the ISO will determine whether it is most economical to keep the resource available in the energy market (to be utilized as a generation resource, if necessary) or to designate it as reserve capacity (to flag as a resource ready to be dispatched in the case of an unexpected supply deficiency).

While bidding into the energy and reserve markets present an additional opportunity for DRRs to yield financial rewards, in most cases, the assets that make up these DRRs (i.e. customers) are not interested in regular load curtailment; they prefer to have minimal interruptions and dispatch infrequently. A possible strategy for DRRs participating in the ISO-NE involves bidding into the energy market at levels that make it unlikely the resource will ever clear.

Potential Concerns and Magnitude of Overlap

Based on interviews of CSPs and customer survey results, between 40% and 60% of PA ADR initiative participants also participate in the ISO-NE FCM, indicating that the potential for overlap is substantial. ISO-NE staff reported two main concerns regarding the interplay between PA DR participation and bidding into the ISO markets:

Participation in PA DR events could result in eroding the ISO-NE baseline for performance, or vice versa.

From the perspective of the PAs, this concern depends on the technology and dispatch strategies employed. The two technologies that were utilized this summer and their treatment in the ISO-NE market is as follows:

- Curtailment resources - Curtailment resources would likely participate in the capacity market as an active demand resource.
- Batteries - Batteries would participate in ISO-NE markets as generation resources. They would provide positive and negative performance data for all hours, avoiding baseline-related issues. Depending on the strategy for participating in the ISO-NE markets, there is the possibility for conflicts with the PA ADR initiative.

The ISO uses a similar 10-of-10 model to calculate the baseline of a customer that has bid into the wholesale markets. This method calculates the baseline using the customer's last 10 weekday, non-DR non-holidays. If the customer responds to a PA-called DR event on a day that falls within those ten days, the customer's ISO-NE baseline will be affected.[23]

[23] Erosion of the ISO-NE baseline caused by the load effects of an ADR event could misleadingly increase or decrease ISO-NE's estimates of load reduction for that asset. If the ADR event load reduction affects only hours during the ISO-NE activation period, then ISO-NE's estimate load reduction would be decreased. If the ADR event load reduction affects the hour of the ISO-NE adjustment period then ISO-NE's estimate of load reduction would be increased. Either way, ISO-NE's ability to accurate assess the available resource is undermined.

ISO-NE staff is equally concerned about the potential impact of PA ADR participation on the day-of baseline adjustment. The ISO uses the 15-minute period prior to a CSC (i.e. the three 5-minute intervals that end most closely to the dispatch signal) to adjust a customer's baseline. If a CSC occurs while DRRs are responding to a PA event, the ISO's 15-minute baseline adjustment would effectively remove any performance the ISO could account for.

The two CSPs with a majority of participants in the 2019 summer season stated that the same-day adjustment to the ISO-NE baseline due to curtailment for an overlapping PA ADR event was a major pain point. One CSP remarked, "baseline erosion does have a slight impact on a customer's performance, but the primary metric that determines performance during an ISO event is the baseline adjustment, which happens day-of…. Coincident events are not the problem – it's *overlapping events* that are an issue, particularly when a PA ADR event occurs first and curtailment is already taking place." In these scenarios, customer participation in the ADR initiative would likely result in a penalty from the ISO. Note that this risk is stated within the PA ADR initiative rules, and CSPs choose to participate with that understanding. One CSP reported that they "insulate" their customers from this risk; in scenarios where customers are curtailing when an ISO event is called, the CSP calculates customer baselines as if they were not already dispatching, pay them accordingly, and take on the non-performance penalty imposed by the ISO. This has helped many of their larger customers consent to participating in both the ADR initiative and the ISO markets.

Scenarios in which events overlap or are called coincidently (i.e. begin at the same time) are expected to be rare, as scarcity conditions occur because of a supply constraint (at the transmission level) while PA DR events are typically called to mitigate load during the system peak hour. Additionally, there were no scarcity conditions that occurred during the summer 2019 season and thus no opportunities for event overlap.[24] While interviewed CSPs and ISO-NE staff did acknowledge that PA DR event and CSC overlap is unlikely, they still expressed concern at the prospect of an overlap having a negative impact on baselines and day-of adjustments. One CSP has suggested repeatedly to the PAs that they coordinate in petitioning or opening a dialogue with the ISO to get the baseline adjustment rule revised.[25]

ISO-NE staff explained that the ISO-NE tariff does include provisions that are similar to the PA ADR initiative's shutdown day allowance; the ISO will remove days on which forced curtailments (i.e. unexpected outages) or scheduled curtailments (i.e., planned shutdowns of facilities) occurred from a customer's baseline. There have been discussions at the ISO about

[24] Since the summer of 2017, the first season of National Grid's DR program, there has only been one CSC, which occurred on September 3, 2018.

[25] Per Eversource and National Grid staff, the PAs have reached out to the ISO repeatedly to discuss coordination opportunities between the two offerings. This is further discussed below.

including provisions for days on which customers respond to local (PA) demand response events, and ISO staff posited that a similar provision could also be made for the day-of baseline adjustment. Such a change to the language in the tariff might alleviate the issue of a corrupt baseline and baseline-adjustment, but it would not address another major ISO concern as discussed below.

Potential for the over-estimation of available reserves. It is crucial that the ISO maintains a certain amount of reserves, not only for the sake of reliability, but also because the North American Electric Reliability Commission (NERC) requires it. As such, the ISO needs an accurate record of available reserves in the marketplace at all times. ISO staff explained that the ISO accounts for reserves through the bidding process and designates whatever MW a DRR has bid into their platform as available reserves.

If a DRR dispatches some of its assets in response to a PA DR event, the DRR no longer has its full dispatch capability, or the capability it had originally bid into the market. In these cases, the ISO both requires and expects that the market participant will revise their physical bid (i.e. kW dispatch capability) in the ISO's platform so they can maintain an accurate account of reserves. This is the only way for the ISO to be made aware that the resource is no longer available or is available at a reduced capacity. If a participant does not reduce its bid, the ISO is unknowingly designating more resources/kW as reserves than are available. This creates system-wide reliability risks and puts the ISO in violation of NERC regulations, subjecting them to sanctions.

ISO staff stated that the failure of DRRs/market participants to manage their bids when their availability changes is a major concern, especially if they dispatch in response to frequent PA DR event calls. Although the evaluation team did not ask CSPs directly about their management of participant FCM bids when a PA event has been called, during an interview for a separate evaluation effort, one CSP engaged in the DR initiative mentioned that they do not reduce their customer bids into the FCM when their availability changes.

ADR Initiative Rules Concerning the ISO-NE Market

Prior to the roll out of the ADR initiative, in March of 2016, the PAs met with members of the ISO's marketing, operations and policy teams as well as DOER and other stakeholders to discuss the proposed design of the ADR initiative and how it might interact with the ISO markets. Per Eversource and National Grid implementation staff, the ISO did not seem concerned with program overlap or interplay and asked that the PAs return for a discussion once the initiative has reached 400 MW in enrolled capacity. National Grid implementation staff noted that in designing the initiative, the PAs considered potential concerns the ISO might have moving and implemented certain rules to reduce the risk to a customer's PA ADR performance calculations posed by participating in ISO-NE scarcity events. The rules are as follows:

1. "If the ISO-NE calls an OP4 event[26] during the baseline period of a Connected Solutions event, this day will not be counted in the baseline."
2. "Although rare, it is possible that both Connected Solutions and ISO-NE will call on a customer to curtail on the same day. This will not affect how the customer performance is calculated in the Connected Solutions program."

PA staff stated that these rules were put in place for the following reasons:

- Participation in the ISO-NE market should not adversely affect the customer's baseline in the PA ADR initiative.
- Conflict with ISO-NE's events should be avoided since they are more critical than the PA ADR events.
- The baseline calculations are meant to use the last ten "similar" non-holiday weekdays, or days where operations were scheduled and proceeded as normal[27]. Days on which a customer participates in an ISO-NE event are not "similar" days, so it makes sense to exclude those days from the baseline.

In discussions with ISO staff, the evaluation team reviewed the ADR initiative rules. When asked if these rules addressed any of the ISO's concerns, ISO staff explained that they did not. While the first creates a work-around for potential ADR initiative baseline erosion and the second mitigates the risk of coincident events, neither address the day-of baseline adjustment, reserve designation issues, or ISO baseline contamination. CSPs echoed this sentiment.[28]

Additional Interplay Issues

ISO staff also mentioned several flaws inherent in the way capacity costs are currently allocated to customers that can be exacerbated by PA ADR programs.

The ISO allocates capacity costs amongst customers based on each customer's load during the system's annual single peak hour of the prior commitment period (this load is called a customer's installed capacity (ICAP) tag). If a customer was dispatched for a PA DR event (or due to a CSC) during that annual system peak hour, the customer would not only get paid for their performance but would also only be required to pay capacity costs on their reduced system peak hour load. ISO staff contended that this reallocation of capacity costs reflects the ability of customers to provide DR services rather than reflecting their overall contribution to

[26] OP4 events include CSCs.

[27] Note, shutdown days are not considered "similar," but if they go unreported, there is no way for a PA to know the day should be removed from the baseline.

[28] As mentioned earlier in this section, the PA rules fully acknowledge that participating in the PA ADR initiatives will pose some risks to the customer's ISO-NE performance. CSPs choose to enroll their customers in both offerings with this understanding.

capacity requirements. By reducing their load during the system peak hour, ISO staff contended, customers are not necessarily reducing costs, but instead are shifting their costs to others; they still require the capacity they would have used during the system peak hour during many other of the top hours during the year, but because they were able to reduce during the system peak hour, they are not paying for the capacity they need. Here, the PAs pointed out that any reduction in customer demand during the system peak hour indeed does bring a reduction in the actual installed capacity requirement (ICR), which in turn should bring costs down for all customers. The PAs also acknowledged, however, that customers that participate in ADR would get more benefits than those who do not, in the form of event payments and the potential for a reduced ICAP charge.

ISO staff acknowledged that this dynamic is enabled by how ISO-NE has designed capacity cost allocation, and there have been conversations to change this methodology (e.g., allocate costs based on the top 10 hours, rather than the one peak hour). With the cost allocation as is, ISO-NE staff stated that if PA ADR initiatives are intentionally targeting the system peak hour as a period to call a DR event, they are facilitating customer benefits that aren't reflective of actual conditions/capacity needs throughout the year.

As discussed above, all three PAs indicated that reducing system peak hour load was a major objective of the ADR initiatives, and that they expected system-level capacity cost reduction as a result. This would benefit all customers. Their CSPs are targeting those hours. Reducing individual participating customers' ICAP tag capacity charges (and shifting some of the lower overall costs to others) is not an objective. Nonetheless, some implementers are marketing this cost reduction to prospective participants in addition to their event payment benefit.

ISO-NE – Potential Solutions to Allay Overlap Concerns

When asked what should be done to address ISO concerns, ISO staff suggested that, in the immediate term, stakeholders should come together to identify easy-to-implement strategies that will allow for better coordination between the ISO markets and PA ADR initiative. In the long term, the ISO will be shifting how DRRs can bid into the energy, reserve and forward capacity markets and will also be increasing penalties for non-performance.

National Grid staff explained that there are established channels whereby the ISO can formally raise any concerns they have with the ADR initiative. One such channel[29] is the Demand Resources Working Group (DRWG), which consists of ISO, PA and CSP representatives and meets on a monthly basis. According to National Grid and Eversource, the ISO has not

[29] The EEAC serves a second forum where the ISO (or other stakeholders) could bring forth concerns about the interplay of DR offerings in MA.

previously responded to repeated outreach attempts or brought forth concerns with the ADR initiative through the DRWG or other channels.

It may be that PA C/I ADR has historically been too small to be of concern to ISO, but recent initiative growth means the prospect of reaching the aforementioned 400 MW threshold is real and thus becoming a matter of interest for the ISO and a good time for re-engagement.

The evaluation team suggests that the utilities incorporate language into program rules that states that CSPs are expected to make DRRs unavailable in the ISO platform for the following day if they will be dispatching for a PA ADR event.[30]

Summary of Findings

- ISO-NE staff concerns regarding PA ADR initiative overlap are:
  - Participation in PA ADR events could result in eroding the ISO-NE baseline calculation and same-day adjustment for performance, or vice versa.
  - The ISO could over designate reserves if DRRs participate in the PA ADR initiative do not revise their FCM bids.
- Although ADR initiative rules specify how co-participation in the initiative and ISO-NE FCM should work, the initiative rules do not address the ISO's overlap concerns.
- Both ISO and PA staff expressed a willingness to discuss overlap concerns and solutions.

4.1.3 Feedback on the Eversource DR Initiative Delivery

This section documents initiative delivery challenges reported by Eversource staff and CSPs, as well as customer feedback.

Marketing and Enrollment

The Eversource DR initiative promotional efforts could use improvement. One CSP mentioned that when pitching the PA initiative to a prospective customer, they referred the customer to the PA website to learn more about the program. This CSP reported difficulty navigating the website to find information about the program. The evaluation team also observed that the program name differs on the marketing material and customer application. Eversource staff reported that the program was branded as "ConnectedSolutions" in all states, which is also

[30] By managing their bid (i.e. partly or completely reducing their bid according to their availability), market participants not only fulfill their responsibility to the ISO, but are also effectively removing their bid from the FCM. This would make it impossible for CSPs to bid into the FCM (through the day-ahead market) at a lower price so as to clear in the energy market during the PA event period. This scenario could have occurred during summer 2019 and neither PAs nor evaluators would be able to confirm it with current available data sharing and analysis capabilities. If it were the case, ISO-NE's energy market would have been distorted by actions of the CSPs.

stated on the website. However, on the application, the program is called "Active Demand Reduction," and at the bottom of the application page it is stated: "To submit your application or to learn more about the Eversource Active Demand Reduction Program, please email connectedsolutions@eversource.com."

Appendix B, Eversource Marketing and Enrollment Section, describes how the program is marketed.

Initiative Administration After Enrollment and Before Payment

Eversource and CSP staff experienced challenges with contracting, application submissions, data tracking, monitoring, as well as management processes (see detailed findings in Appendix B). To alleviate these challenges, Eversource implementation staff is making the application process electronic and designing infrastructure to enable reporting, monitoring, and dispatching of event communications within one Distributed Energy Resources Management System (DERMS). This DERMS platform will also facilitate initiative tracking and customer segmentation. It is expected that many of the challenges experienced in 2019 will likely be alleviated when the DERMS platform comes online.

Payments

Eversource staff reported that settlement and payment were the most challenging of all initiative administration processes during the 2019 summer season, explaining that the implementation team's decentralized management systems made the settlement calculation and payment processes difficult. Basic accounting and customer labeling issues slowed the payment process. It was particularly difficult to calculate payments for customers who implemented multiple DR strategies using the same asset (e.g., a facility that uses a battery to participate in both targeted and daily dispatch).[31] Eversource staff also noted that some customers were not able to participate and receive payments because their CSP-installed interval meters did not record or transmit information.[32]

Eversource's Active Demand Reduction application[33] T&Cs state that payments for the summer season will be made in the fall. Very few (17%) participants reported they had received payments when the evaluation team surveyed them in November and December of 2019. One CSP who was interviewed in January of 2020 also noted not yet having received payments from

---

[31] In the future, customers will not be allowed to participate in more than one program option with one asset, as this causes too many accounting issues. This does not prevent customers from using a single asset to both participate in a DR initiative and manage monthly utility-billed peak demand.

[32] All Eversource customers have utility meters; however, customers are still required to install CSP meters capable of measuring usage at the 5-min interval and sending data to the PA in real time (which some utility meters are not capable of).

[33] https://www.eversource.com/content/docs/default-source/save-money-energy/demand-reduction-application.pdf?sfvrsn=7bb2ca62_6

Eversource. It is unclear how this affects customer satisfaction since a substantial number of surveyed participants stated “don't know” when asked about payment satisfaction (see Eversource Participant Satisfaction Section below).

The Eversource staff expects that the centralized DERMS platform will improve payment processing for the next summer season. Eversource uses CSP data to make baseline and settlement calculations and requires that CSPs install five-minute interval meters at all customer sites. (These meters are in addition to utility meters.) Eversource is planning to pull customer interval data directly from the CSPs through their new centralized platform. This will enable a near-constant flow of usage information. In close to real-time, CSPs will be able to identify meter malfunctions and correct data transmission issues. The Eversource staff will aim to work out any data discrepancies with the CSPs prior to beginning baseline and settlement calculations.[34]

One CSP noted that the meter installation requirement makes it cost-prohibitive for smaller customers who need to pay for the meters to participate in DR. Although Eversource customers can apply for an incentive of up to $1,500 per meter, those funds, on average, typically cover only 40% of the CSP installation costs. This CSP suggested either relaxing the requirement, increasing the incentive funds available, or using utility interval meter data to calculate baseline and settlements, where applicable.

CSP Performance

Despite having faced several administration challenges, Eversource staff expressed satisfaction with the 2019 summer season results. Staff noted that CSPs performed nicely. Four[35] events were called and 6 of the top 10 system demand hours occurred during called events, including both July and August peaks and the ICAP hour (5 p.m. to 6 p.m. on Tuesday, July 30th).

Staff reported only two event delivery challenges:

- Eversource staff decided to call an event on a Saturday, even though the initiative as currently designed does not account for weekend events. When they called this event, they explained it would be “voluntary,” which meant that if a customer's performance during the event was lower-than-average or if a customer was unable to perform, their baseline would not be impacted. However, the PA would take into account customer performance that was higher-than-average when calculating baselines to ensure that overperformance was reflected in their event payments. One CSP reported not

[34] Eversource also contracted with a vendor to assess performance for customers with interval data. Initially, the vendor's results varied drastically from CSP calculations. It was found that they had erroneously accounted for an event on a day when no event was called. When the vendor corrected the error, the vendor's final calculations varied only slightly from those of the CSPs.

[35] Includes the voluntary weekend event.

understanding how the event would impact their customers and what the benefit would be, prompting them to ask the PA to clarify multiple times. This CSP remarked about this confusion and multiple follow-up discussions by saying, "that's not the sort of communication you want to have 24 hours before an event."

- The second event delivery challenge noted by Eversource staff related to the event called on July 30th, which ended up being when the ICAP hour occurred. Since it was the first event of the summer that all three PAs (Eversource, National Grid, and Unitil) planned to call, they aimed to call the event together, and for the same event window. Eversource staff noted there had been a concerted effort to design the initiative together; hence, the PAs believed there would be value in making dispatch homogenous across New England. Eversource originally intended to call the event from 4 p.m. to 7 p.m. but ended up calling the event one hour earlier (from 3 p.m. to 6 p.m.) along with the other PAs. This coordination required more real-time system monitoring and decision-making than the Eversource staff thought was practical, and they also noted risking missing the ICAP hour. They also did not perceive that the coordinated event was beneficial for the initiative or the distribution system. For the rest of the season, Eversource called events independent of the other two PAs, and they were the only PA to call more than one event over the summer. Eversource staff plans to operate independently for the upcoming DR seasons as well.

Participant Satisfaction

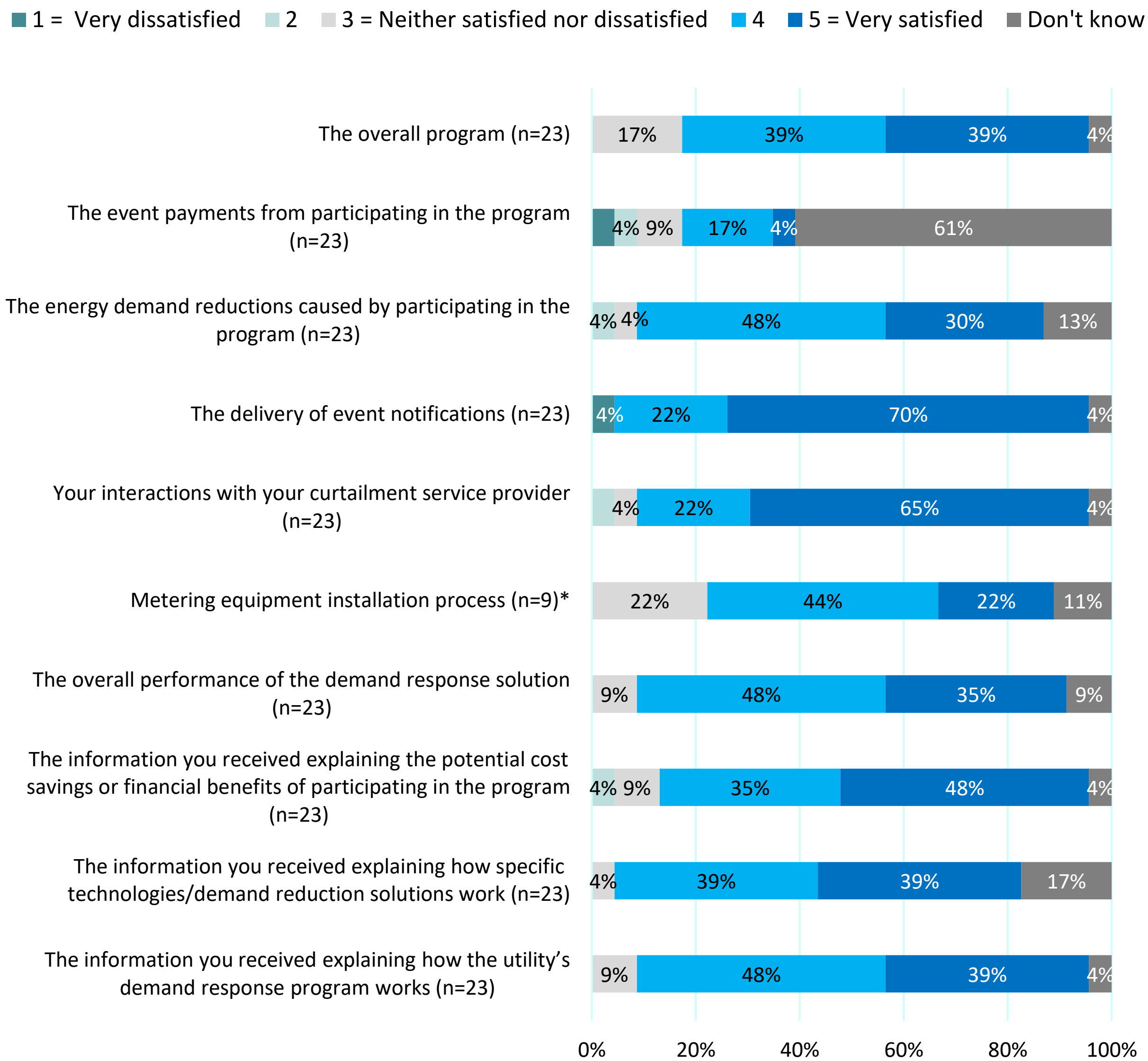
Twenty-three Eversource participants rated their satisfaction with various aspects of the program on a scale of 1 to 5, with 1 being "very dissatisfied," 3 being "neither satisfied nor dissatisfied," and 5 being "very satisfied."

Mostly notably, participants were highly satisfied with the delivery of event notifications (70% gave a score of 5 out of 5) and their interactions with their CSPs (65% of customer gave a score of 5 out of 5). Participants expressed overall satisfaction with nearly all of the remaining program elements except for one: the event payments (Figure 4-2). A large number of participants (70%) did not know how to rate the event payments, saying "don't know" or providing a rating of 3 (i.e., "neither satisfied nor dissatisfied"). This is either because respondents were not involved in or aware of the payment process (i.e. others were responsible for their organizations' financial activity), or because it payment had not yet been received at the time of the interview. Four respondents in the sample (17%) reported they had received event payments. Among them, two rated their satisfaction with their payments as a 4 out of 5 and two gave a "don't know" response.

Fifty-two percent (12 of 23) reported not yet receiving event payments, and 30% (7 of 23) noted not knowing whether they had received payments. Among the 12 respondents who had not received payments, eight could not provide a satisfaction rating, saying "don't know"; two gave a rating of 1 or 2, expressing dissatisfaction with this program element; one stated "neither satisfied or dissatisfied;" and one gave a rating of 4, expressing satisfaction. Two dissatisfied participants explained that they were dissatisfied because they had not yet received payments and one of them heard that others were receiving payments from Eversource, which concerned them. Note that each CSP has a different payment schedule, which means there is no consistent payment schedule.

It is unclear at this time whether participants are generally satisfied with the event payments. Collectively, these findings suggest that participant payment satisfaction needs to be monitored.

Figure 4-2. Participant Satisfaction Ratings



* Only a subset of participants reported installing metering equipment. Only those participants were asked this question, which is why the sample is much smaller for this element.

> The program positively impacted 43% of participant opinions of Eversource and negatively impacted 0%.

Participants were asked whether their program experience affected their opinion of their PA. About 56% (13 of 23) reported their opinion of their PA remained the same after participating. The remaining participants (43%, or 10 of 23) reported their program experience positively impacted their opinion of their PA.

Facility Disruptions - Curtailment Participants

> The majority reported experiencing disruptions when participating in an event. For some, these disruptions interfered with their participation. For others, these disruptions had no effect on their participation.

Twenty-two[36] curtailment participants (i.e., targeted dispatch participants) provided feedback on facility disruptions when participating in a program event. About 55% (12 of 22) reported experiencing slight or temporary disruptions to their company's core operations as a result of reducing demand this summer. Three reported complaints from building occupants. About 36% (8 of 22) reported having no disruptions at all. Of those who reported slight disruptions, three said that those disruptions prevented them from curtailing load during one or more called DR events this summer. These three participants (two manufacturing plants and a research center) achieve demand reduction through curtailing production activity or using HVAC equipment, upon which critical systems rely to continue regular operations. These specific sites could not afford even the slightest disruption in core operations caused by curtailment during DR events this summer, as production and research objectives cannot be compromised. All CSPs acknowledged that industrial facilities might not always be able to participate for these reasons as well. Another participant, a public school, participated only partially for all DR events; they stated that because school was in session, comfort and safety of facility occupants was the priority, which meant not all curtailment activities could be completed.

Grocery and retail participants also explained their difficulties in participating in initiative events. One of them remarked: "It's never convenient to shut off half your lights in a supermarket; customers have to see the price of what they're paying for." This respondent also mentioned safety concerns, that is, butchers being able to see what they are doing as they work with sharp tools. The other two respondents mentioned heat or humidity being an issue. One wanted events to be shorter so that the stores would not feel the temperature difference. The other wanted more opportunities to engage with the initiative on days that are not as humid (i.e., more events called). One CSP mentioned that grocers that are regulated to keep their stores at a certain temperature might not be able to participate in an event.

Facility Disruptions – Battery Storage Participants

Eversource had two initiative participants that had a battery system. One of them did targeted dispatch (with both a battery and through curtailment) and the other did targeted and daily dispatch[37] with the same battery. Only the latter participant reported slight or temporary

---

[36] Twenty-three participants responded to the survey, but one of the battery storage participants did not engage in curtailment actions and thus is excluded from this analysis.

[37] Eversource offered battery daily dispatch as part of a separate demand response demonstration program. The demonstration is being evaluated separately from Eversource's full-scale ADR initiative.

disruptions to their operations because of their DR initiative participation activity. Their site has a cogeneration system that frequently went offline unexpectedly when the storage system dispatch lowered load past a certain threshold. Often, this diverted maintenance personnel from routine tasks to deal with the cogeneration system outage. This issue did not prevent this site from participating in any DR events, but the customer indicated that it limited the amount of kW reduction they could deliver.

DR Actions Taken to Curtail Load

More participants engaged in manual than automatic actions to curtail load.

Under the plan with their CSP, 50% (11 of 22)[38] of surveyed curtailment participants reported being responsible for making manual adjustments to curtail load. Fewer participants (7 of 22) reported their CSP planned to make automatic adjustments. Among those, only two reported also making some manual adjustments to curtail load.

Most cited curtailment actions related to cooling (see Table B-1 in Appendix B).

4.1.4 Feedback on the National Grid DR Initiative Delivery

This section documents initiative delivery challenges reported by National Grid staff and CSPs, as well as customer feedback.

Marketing and Enrollment

National Grid staff mentioned that recruitment for the 2019 summer season was more difficult for the PA than it had been over the first two years of the DR initiative. Since a significant percentage of the PA's largest commercial and industrial customers had already signed up, National Grid staff noted having to increase their sales efforts in 2019 to achieve the same amount of MW reduction that had been reached in prior years with less resources spent on recruiting. They reached out to more customers, increased frequency of email marketing, improved the initiative website and informational materials, and worked closely with CSPs and AEs to recruit additional customers into the initiative. The PA anticipates this need for increased promotion to continue.

Appendix B, National Grid Marketing and Enrollment Section, describes additional details on how the ADR initiative is marketed.

Payments

The payment processing remains challenging. DR initiative marketing materials state that incentives will be paid out in October[39]. When the evaluation team interviewed National Grid

[38] Excluded one respondent who did a targeted, daily dispatch with their on-site energy storage.
[39] https://www.nationalgridus.com/media/pdfs/bus-ways-to-save/connectedsolutions-matargeteddispatchflyer.pdf

staff in December of 2019, there were still a few payments that had not yet been made. During a follow-up conversation, National Grid staff explained that the delay in payments was largely because National Grid's procurement protocols had been revised. This meant that National Grid staff had to have CSPs re-sign contracts, NDAs and ISAs in order to process DR initiative payments.

Staff also explained that each summer season, a small percentage (between one and five percent) of customers experience metering or data issues. With the increase in initiative enrollment over the past three summers, it has become more time-consuming for staff to resolve these data issues.

Two CSPs expressed dissatisfaction with National Grid's data sharing and the payment process. It was not until after the 2019 summer season ended that National Grid provided customer performance data to the CSPs, who were unable to reconcile any data issues until they had that data. CSPs reported wanting to obtain the utility's meter data sooner (i.e. within a few days following an event) rather than at the end of the season, in order to more quickly identify faulty meters and reconcile data discrepancies. This would also allow CSPs to identify customers that performed sub-optimally and help improve curtailment strategies for the next DR event. One CSP reported that they were discussing with National Grid the ability to access the PA's online energy portal, which the CSP understands provides day-after data and daily performance summaries. National Grid staff noted that all CSPs have access to their online energy portal, but that the CSPs need to get customer permission for National Grid to share customer data with them.

One challenge that access to the PA's online energy portal will not solve, according to the CSP looking to access customer data through National Grid's online energy portal, is the inability of the utility's meters to record customers' export of power. This CSP has several large customers that export power, and National Grid's meters have either failed to record or have incorrectly accounted for their power export. For example, a customer's power generation can bring a load from 200 kW to 0 kW, and then export 100 kW back to the grid. Instead of registering a 300-kW reduction, the PA has subtracted the initial and ending demand during the event (200 kW to 100 kW), showing only a 100 kW reduction. For some of these cases, the CSP had to involve the customer and obtain their generation data to reconcile the data discrepancy. These particular cases had a long settlement and payment process for the 2019 summer season.

There is also an indication that customers are starting to express frustration with the payment turnaround time. By mid-December, one CSP reported receiving numerous calls from frustrated participants, asking when they would receive their payments.

Participant Satisfaction

Thirty-two National Grid participants rated their satisfaction with various aspects of the initiative on a scale of 1 to 5, with 1 being "very dissatisfied," 3 being "neither satisfied or dissatisfied," and 5 being "very satisfied."

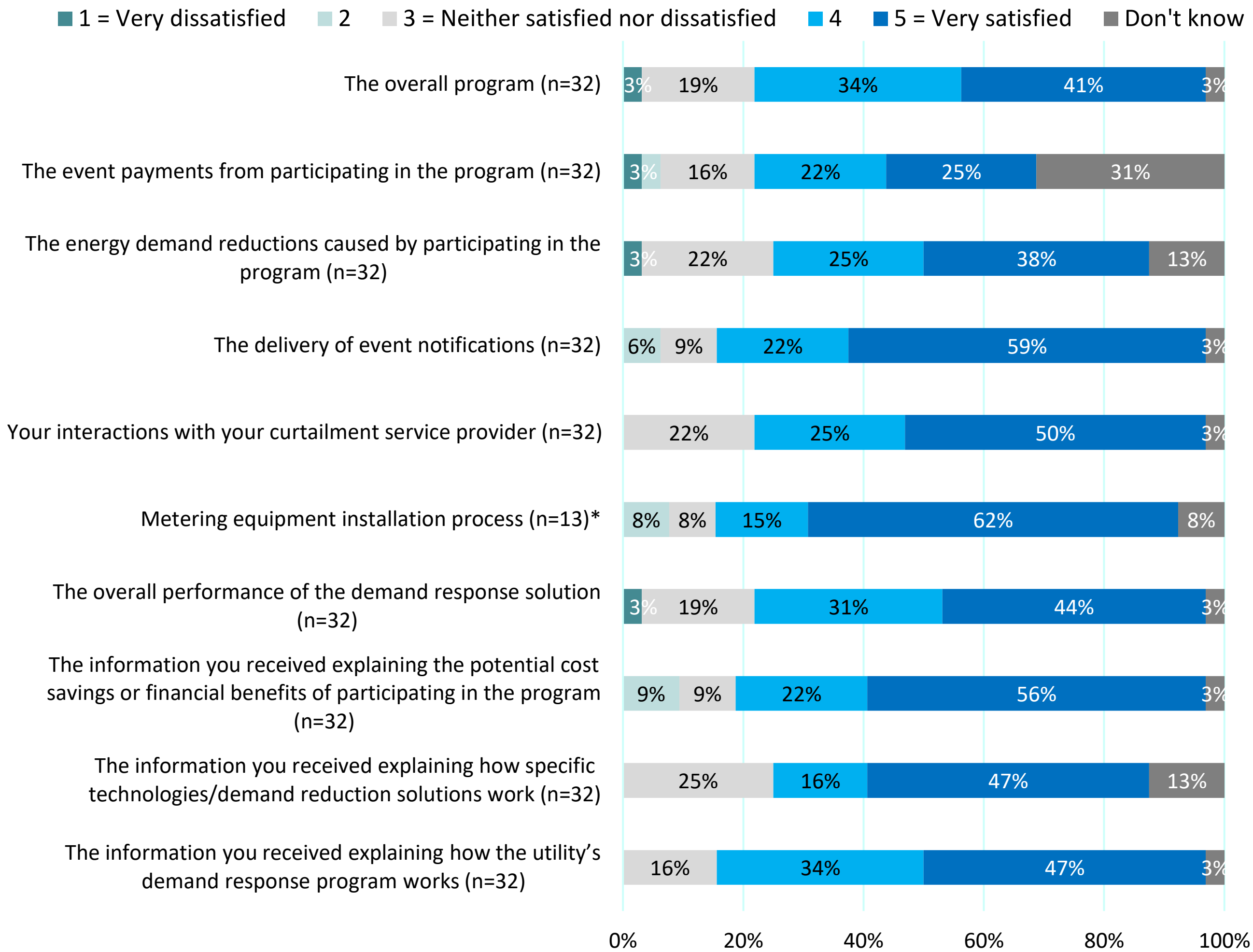
National Grid participants indicated moderate to high satisfaction for many, but not all, aspects of the initiative. Over 75% gave ratings of 4 or 5 for their interactions with their CSPs, the overall performance of their demand response solution, the delivery of event notifications, the metering installation (if applicable), the information with how the initiative works and its potential financial benefits, and the initiative in general (Figure 4-3). A smaller proportion (60% to 65%) were satisfied with the energy demand reductions achieved through the initiative and information on how technologies or demand reduction solutions work (Figure 4-3).

Customer satisfaction with event payment is unclear. While 47% of participants were satisfied with the event payments, the same proportion of participants did not provide a definitive rating of the event payments (31% "don't know," and 16% "neither satisfied nor dissatisfied"). This was to be expected, since only a minority of respondents (34%, or 11) in the sample reported they had received event payments. Of those 11 respondents, eight rated event payments as a 4 or 5 (high satisfaction), one gave a rating of 3 (neither satisfied nor dissatisfied), and the rest indicated "don't know." None were dissatisfied.

Forty-four percent (14 of 32) reported not yet receiving event payments, 19% (6 of 32) noted not knowing whether they had received payments, and one respondent did not answer this question. Among the 14 who had not yet received payments, seven could not provide a satisfaction rating, saying "don't know;" three stated "neither satisfied or dissatisfied;" three gave a rating of 4 or 5, expressing satisfaction; and one gave a rating of 1 or 2, expressing dissatisfaction with this initiative element. One participant expressing dissatisfaction explained they were dissatisfied because they did not have a clear understanding of how payment calculations and protocols worked in general.

Due to a notable number of "don't know" responses, it is unclear at this time whether participants are satisfied with the payments. Collectively, these findings suggest that participant payment satisfaction needs to be monitored.

Figure 4-3. Participant Satisfaction Ratings



***** Only a subset of participants reported installing metering equipment. Only those participants were asked this question, which is why the sample is much smaller for this element.

Participants were asked whether their initiative experience affected their opinion of their PA. About 63% (20 of 32) reported their opinion remained the same after participating. The rest (38%, or 12 of 32) reported their initiative experience positively impacted their opinion of their PA.

Facility Disruptions and Opting Out

A notable number of participants reported experiencing slight or temporary disruptions when participating in an event, and some experienced complaints from facility occupants (Table 4-1).

Table 4-1. Participation Challenges (Multiple responses allowed)

Challenges	
	n=32
Slight or temporary disruptions to your company's core operations	15 (47%)
Complaints from facility occupants	6 (19%)
Major disruptions	1 (3%)
Other	2 (6%)
Don't know	2 (6%)
No disruptions	12 (38%)

Eight respondents, all representing manufacturing facilities, reported performing partially or not at all during the DR event(s)[40] called this summer. These respondents account for 2,884 kW of enrolled reductions. Four explicitly cited customer demand or the inability to decrease production as the reason for their lack of participation. All CSPs acknowledged that industrial facilities might not always be able to participate for these reasons as well. The impact evaluation found that four of these customers underperformed, three overperformed, and one performed as expected.

DR Actions Done to Curtail Load

More participants engaged in manual than automatic actions to curtail load.

Under the plan with their CSP, 69% (22 of 32) of surveyed curtailment participants reported being responsible for making manual adjustments to curtail load. Less (7 of 32) reported their CSP planned to make automatic adjustments. Among all those who opted for automatic adjustments (n=7), only three reported also relying on manual adjustments to curtail load.

Most cited curtailment actions related to cooling (see Table B-2 in Appendix B).

4.1.5 Feedback on the Unitil Initiative Delivery

This section documents initiative delivery processes reported by Unitil staff and CSPs, and customer feedback.

Marketing and Enrollment

To meet the 2020 season demand reduction targets, Unitil staff are considering making several changes to how they promote their DR initiative. In 2019, participants were mainly recruited by one CSP who enrolled a few of their eligible existing participants into Unitil's DR initiative. Unitil staff explained that the CSP did most of the outreach and worked with Unitil's AEs as needed but that ultimately, the PA's AEs did not play a large role in recruitment. Staff had also felt it was not a priority to develop initiative content for the PA website in time for the summer

[40] Some participants recalled having been notified of more than one DR event, even though National Grid only called one event.

season, as it was more likely for large customers to contact their AEs directly to get initiative information than to reference the PA's website. Information about Unitil's DR initiative offering is not easily found elsewhere on the web. Unitil is listed with the other PAs on the initiative application posted on the MassSave website but is not mentioned alongside Eversource and National Grid on MassSave's landing page for the initiative[41]. There is no information about the DR initiative on the NHSaves website.

In 2020, Unitil staff plans to more actively promote the pilot by:

- **Developing pilot content on their websites.** The PA will publish pilot content on the Unitil and NHSaves websites and will make any needed updates to the MassSave website by April 2020.
- **Working with an additional CSP.** In New Hampshire, Unitil is exploring contracting with another CSP. The Unitil staff plan to first solicit feedback from AEs, who are the most familiar with Unitil customers, on the implications of contracting with another CSP and the best way to promote the pilot. Staff were concerned that customers might be fatigued or frustrated if they are contacted by multiple CSPs.

The one CSP working with Unitil reported that it would help them to discuss recruitment with Unitil's AEs. They had reached out multiple times to a few customers without success and stated that meeting with the AEs early on in the sales season to discuss how to reach large customers would be a boon to their recruitment efforts. This process is starting to take shape.

Payments

There is no indication that Unitil staff experienced challenges with the payment process.

The baseline and settlement calculations were relatively straightforward. Many of the customers who were enrolled in Unitil's DR pilot also participated in the ISO-NE Forward Capacity Market with the same CSP. Since they participated in the ISO-NE market first, they had CSP or their own meters installed. Every customer also has a Unitil interval meter already in place.

Participant Satisfaction

This and subsequent sections describe customer responses to a pilot participant survey. These survey responses should be interpreted with caution, due to small sample size. Three of seven customers who participated in the Unitil DR initiative in the 2019 summer season responded to our survey.

[41] https://www.masssave.com/en/saving/business-rebates/demand-response-and-storage

Three Unitil participants rated their satisfaction with various aspects of the pilot on a scale of 1 to 5, with 1 being "very dissatisfied," 3 being "neither satisfied nor dissatisfied," and 5 being "very satisfied."

These participants were highly satisfied with the initiative. All three indicated high satisfaction (rating of 4 or 5) with the following elements of the initiative:

- Information on the potential financial benefits of participating
- The information on how technologies or demand reduction solutions work
- The performance of their demand response solution
- The energy demand reductions achieved through the pilot
- The delivery of event notifications
- The metering installation (if applicable)
- The payment system
- The pilot in general

Two of the three participants were also highly satisfied with the interactions with their CSP and the information on how the initiative works. One respondent who did not indicate high satisfaction with the information on how the initiative works rated this element as "neither satisfied nor dissatisfied."

Note that although all three respondents gave ratings of 4 out of 5 for event payments, , two noted not yet having received event payments, and the remaining respondent was unsure of whether they had received payments.

Participants were also asked whether their pilot experience affected their opinion of their PA. Among two who gave a response, one reported their opinion of their PA remained the same and the other reported their pilot experience positively impacted their opinion.

Facility Disruptions and Opting Out

Only one participant reported experiencing slight or temporary disruptions when participating in an event. This participant also explained that the time of the event was later in the day, and thus they could not respond.

DR Actions Done to Curtail Load

Under the plan with their CSP, only one of three surveyed participants reported being responsible for making manual adjustments to curtail load. One respondent was unsure if they had opted to perform manual or automatic adjustments, and the third respondent gave an unclear response.

The one participant responsible for making manual adjustments noted conducting the following actions to curtail load:

- Turning off unnecessary electric domestic hot water heaters
- Turning off inessential split system units

Participant Suggestions for DR Initiative Improvement

Participants gave no suggestions on how to improve the initiative.

4.2 Impact Evaluation Findings

This section provides impact evaluation findings described in the C&I Interruptible Findings of the Executive Summary by PA, state, and event.[42] The called events during summer 2019 were:

- 7/19 from 4:00 to 7:00 p.m. (Eversource only)
- 7/30 from 3:00 to 6:00 p.m. (Eversource, National Grid, and Unitil)
- 8/19 from 4:00 to 7:00 p.m. (Eversource only)

The ISO-NE system peak hour (ICAP hour) occurred on July 30, from 5:00 to 6:00 p.m. ICAP hour results are the average load reduction during the ISO-NE ICAP hour. Event day results are calculated as the average hourly load reduction. Overall results are the average load reduction across events.

4.2.1 Guide to Impact Evaluation Results

Table 4-2 replicates Table 3-3 from the Methodology and Framework section. It summarizes the summary of performance metrics that are presented and discussed throughout these findings. The load reduction is the difference between these baselines and the actual observed load during the event hours. Note, in order to make results comparable, all results included in the findings are for accounts that the evaluation team had sufficient data to estimate load reduction. Details of the evaluation team's data sufficiency criteria, missing data, and applicable baseline calculations can be found in the

[42] Demand response events for the ADR initiative are only called on weekdays; however, the ICAP hour was forecasted on Saturday, July 20, 2019, due to peak load conditions. In response, Eversource called an extraordinary (voluntary) event. Results for this extraordinary event are reported in Appendix D: Eversource Extraordinary Weekend Event.

Enrolled Capacity	Asset-level expected load reduction. Generally estimated by CSP.
Reported: Asymmetric	Load reduction claimed reported by program implementers.
Evaluated: Validation	Evaluation team's attempt to replicate reported asymmetric results. Baseline calculation details (e.g., lookback days) vary by PA.
Evaluated: Unadjusted	Rolling average of load in each interval across the 10 most recent eligible days.
Evaluated: Asymmetric	Shifts the unadjusted baseline upward to meet observed load during pre-event adjustment period. Unadjusted baseline is used in lieu of downward adjustments.
Evaluated: Symmetric	Shifts the unadjusted baseline upward or downward to meet observed load during pre-event adjustment period
Evaluated: Forecast	Modification of evaluated symmetric results by accounting for probability of unreported shutdowns.
Evaluated: Regression	Site-level model of load across the summer. Specification describes load as a function of cooling degree-days, weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit.

In each PA-specific impact findings section, and at the end of the impact findings section, the evaluators have also calculated performance ratios for multiple combinations of metrics. None of the performance metrics include accounts with insufficient data. The four ratios are:

- Enrollment Ratio: This ratio is the reported asymmetric load reduction to the CSP reported enrolled capacity. This ratio provides insight into what percentage of the reported enrolled capacity was achieved, based on the program baseline and calculation methodology. This ratio is particularly meaningful for planning and sales purposes.
- Asymmetric Ratio: This ratio is the evaluated asymmetric load reduction to the reported asymmetric load reduction. This is an apples to apples comparison of the same baseline methodology between the PAs and evaluators, however, this metric identifies the impact that different calculation rules between the PAs and evaluators has on load reduction.
- Retrospective Realization Rate: This ratio is the evaluated symmetric load reduction to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline is the most appropriate measure of retrospective load reduction for the 2019 summer season. This ratio shows how the choice of baseline adjustment and calculation methodologies impacts the load reduction estimates. The evaluators recommend using this realization rate to calculate the symmetric load reductions at the end of future seasons if there are no evaluations conducted.
- Prospective Realization Rate: This ratio is the evaluated symmetric load reduction with an adjustment for unreported shutdowns to the reported asymmetric load reduction. The evaluators determined that the symmetrically adjusted baseline accounting for unreported shutdowns is the most appropriate measure of prospective load reduction for future seasons. This ratio provides insight into the magnitude of reductions that could be achieved during future seasons as a function of the validated load reduction estimates.

The prospective realization rate should only be used as an ex-ante estimate of future performance for planning purposes and not retrospectively.

- The product of the enrollment ratio and prospective realization rate is a useful planning metric. It indicates of the percentage of expected future load reduction available from future CSP reports of enrollment (as opposed to PA reports of it).

4.3 Data Sufficiency subsection of the Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

4.3.1 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned
- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs
- Other opportunities for peak demand management

4.3.2 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology.

Table 4-2. Summary of Performance Metrics

Enrolled Capacity	Asset-level expected load reduction. Generally estimated by CSP.
Reported-Asymmetric	Load reduction claimed reported by initiative implementers.
Evaluated-Validation	Evaluation team's attempt to replicate reported asymmetric results. Baseline calculation details (e.g., lookback days) vary by PA.
Evaluated-Unadjusted	Rolling average of load in each interval across the 10 most recent eligible days.
Evaluated-Asymmetric	Shifts the unadjusted baseline upward to meet observed load during pre-event adjustment period. Unadjusted baseline is used in lieu of downward adjustments.
Evaluated-Symmetric	Shifts the unadjusted baseline upward or downward to meet observed load during pre-event adjustment period
Evaluated-Forecast	Modification of evaluated symmetric results by accounting for probability of unreported shutdowns.
Evaluated-Regression	Site-level model of load across the summer. Specification describes load as a function of cooling degree-days, weekends and holidays, calendar month, and event day terms. The cooling degree-day base is determined by regression best fit.

The tables presented in the following sections describe load reduction estimates across the PAs, states, or event days. The tables track a progression of results from enrolled capacity to reported asymmetric load reduction to evaluated results.

Enrolled capacity is the nominated or expected load reduction, generally provided by the vendor CSP. The reported asymmetric estimates are based on the Initiative's chosen settlement baseline and provided by the PAs. Differences between the PA-reported asymmetric and the evaluated validation results reflect the evaluation team's inability to perfectly replicate some aspect of the PA's calculation process and, in some instances, may reflect different underlying data used in analysis.[43] Note the reported asymmetric results are provided at the event level. Evaluated validation results are calculated and provided at the event-level for consistency with the reported asymmetric results. The process from Reported-Asymmetric to Evaluated-Validation varies by PA and is explained in detail in Appendix C: Settlement Verification.

The Evaluated-Asymmetric baseline results use the initiative's asymmetric baseline framework but apply the evaluation team's recommended statewide data sufficiency and calculation rules for consistency across results. The difference between these Evaluated-Asymmetric results and the validated results indicates the necessity of consistent reporting algorithms across the PAs. Generally, differences may be caused by data sufficiency requirements, use of baseline look-back days, and reporting of negative load reduction as zero.

Each of the last four rows of the tables presented are evaluated results following the evaluation team's consistent set of data sufficiency and calculation rules. The unadjusted, symmetric, and asymmetric baselines all use the same 10-of-10 baseline shape. The unadjusted baseline shape is calculated as the rolling average of load in each interval across a 10-day baseline pool. The symmetrically adjusted baseline shifts the unadjusted baseline to meet observed load during pre-event hours (adjustment period). The asymmetric baseline only shifts the unadjusted baseline upward. If the adjustment is negative, the unadjusted baseline is used. As such, the asymmetrically adjusted baseline is a combination of the unadjusted and symmetrically adjusted baselines.

The symmetrically adjusted baseline reduces biases for sites with variable load due to weather or other production factors. The symmetrically adjusted baseline methodology is the most commonly used baseline approach and is used by ISO-NE. All three baselines are provided primarily to demonstrate the extent to which the combined, asymmetrically adjusted baseline artificially increases the estimate of load reduction over either of the unadjusted or symmetrically adjusted baselines. **The evaluation team recommends using a symmetrically adjusted baseline as the most appropriate measure of event period load reduction for the 2019 summer season.** The symmetrically adjusted baseline is also the basis of the evaluation forecasted results. As discussed in Section 3.2.1 Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

[43] Reported-Asymmetric results may reflect interval meter data from the CSP, while the evaluation team generally uses interval meter data provided by the PAs, though there are exceptions.

4.3.3 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned
- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs
- Other opportunities for peak demand management

4.3.4 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology, **the forecast estimate of load reduction is based on the symmetrically adjusted baseline results with an additional adjustment for propensity to shut down.** There are many other factors that the evaluator might be able to adjust for, given more information about the site and the load reduction occurring there. Understanding more about the load reduction (e.g., fixed or variable, nature of control, etc.) would make it possible to improve forecasts of future load reduction.

4.3.5 C&I Interruptible

4.2.2.1 Eversource

Table 4- provides the load reduction estimates for Eversource across three states. The event average results are averages across the three events Eversource called. Due to the varying number of accounts that have sufficient data to estimate load reduction for each event, these results are calculated at the account-level and then aggregated to the initiative-level. In other words, each account with sufficient data to estimate load reduction for at least one event is weighted equally in the event average. The bottom row shows the total number of assets that the evaluation team has sufficient data to estimate load reduction for at least one event. The reported results were only provided on an average event basis. The evaluated-validation results are calculated and provided at the event level for consistency with the reported results.

Table 4-3. Eversource Impact Summary

Result	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
	MA		NH		CT	
Enrolled Capacity	37,248	36,683	5,905	5,905	12,519	12,519
Reported-Asymmetric	22,261	N/A	5,156	N/A	13,085	N/A
Evaluated-Validation	21,528	N/A	5,786	N/A	12,558	N/A
Evaluated-Unadjusted	6,561	7,371	3,202	2,689	7,951	7,275
Evaluated-Asymmetric	19,912	21,760	5,661	5,981	12,158	11,564
Evaluated-Symmetric	17,432	20,558	5,147	5,947	11,462	10,921
Evaluated-Forecast	16,523	19,380	4,953	5,674	11,330	10,779
Evaluated-Regression	12,749	15,385	4,446	3,823	9,594	9,824
Accounts	**161**	**151**	**40**	**40**	**96**	**96**

The enrolled capacity is the nominated or expected load reduction, generally provided by the vendor CSPs at the asset level.

The Reported-Asymmetric results are the load reduction estimates provided by the PA.

The Evaluated-Validation results represent the evaluation team's attempt to replicate the Reported-Asymmetric results. The inability to fully validate Eversource's Reported-Asymmetric load reductions has multiple causes. Eversource uses a distributed energy resources management system (DERMS) as the primary source for load reduction estimates. Vendor CSP estimates are used as a secondary source when the DERMS vendor is unable to estimate load reduction. In these instances, the underlying interval meter data used to estimate load reduction may be different. Other differences between the Reported-Asymmetric and the validated results represent challenges in replicating the baseline. For example, there may be differences in how the meter-level interval data was aggregated to the account-level. More details on these discrepancies are provided in Appendix B: Settlement Verification.

4.4 The Evaluated-Asymmetric results are lower than the reported results, reflecting the evaluation team's application of recommended statewide data sufficiency and calculation rules. For Eversource, the difference between the Evaluated-Validation and Evaluated-Asymmetric results is primarily driven by the inclusion of negative load reduction estimates in the evaluated results, rather than setting them to zero. Reported-Asymmetric results are provided at the asset level for each event. In instances that an asset is estimated to have negative load reduction, the result is reported as zero, as opposed to the actual value estimated. When the event results are aggregated to the initiative level, the cumulative effect can be significant. As discussed in Section 3.2.1 Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

4.4.1 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned
- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs

- Other opportunities for peak demand management

4.4.2 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology (Baselines), negative load reduction estimates are a natural part of baseline estimation. Baseline estimation error is always present but is only visible when negative load reduction is estimated, which is only likely to occur for an asset that did not take any load reduction action. These negatives need to be included to avoid another form of upward bias on the estimates.

The Evaluated-Symmetric load reduction estimates used the closest baselines to the approach used by ISO-NE and other ISOs. As discussed in Section 3.2.1, it is an accurate estimate of load reduction to the extent that load in the pre-event period is representative of a non-event day. Because it is adjusted to actual load just prior to the event period, it is the least variable estimate of load reduction when compared to other baselines across multiple days. However, if the day-ahead notification causes load to be unrepresentative of a non-event day,[44] then the estimate of load reduction would be biased either up or down. The Evaluated-Symmetric final estimates are lower than the Evaluated-Asymmetric results because of the inherent upward bias in the

[44] At the aggregate level there may be a mix of early shutdowns and pre-cooling. The net effect of these negative and positive effects on load would represent the bias.

asymmetrical baseline.[45] Using the symmetric adjustment further reduces the event average load reduction.

The evaluated-forecast results further adjust the symmetric results to address the likelihood of shutdown. For a forward-looking estimate of expected load reduction, given the empirical evidence available from this summer, this is the most informed estimate possible. In the future, with additional information about participating assets and more empirically based load reduction estimates, improved forecast estimates may be possible.

The Evaluated-Regression results provide load reduction estimates based on the model specification. The regression baseline appears to do a poor job of estimating event day load. There may be non-weather-related variability in the load data that the regression is unable to characterize. In addition, if extreme-day weather has different load effects than those expressed in the data on less extreme warm days (non-linearities), then the regression, though characterizing much of the weather-correlated load well, will not be able to characterize all weather-correlated effects on the event day.

Figures 4-4 through 4-6 provide visual representations of the July 30 ICAP event day for Eversource by state. Each of these figures show actual load (orange line), the unadjusted 10-of-10 baseline (dark blue line), the symmetrically adjusted baseline (light blue line), the asymmetrically adjusted baseline (cyan line), and the regression baseline (green line). The adjustment window and event period are shaded in gray. The ICAP hour occurred on July 30 from 5 to 6 p.m. and is identified with darker shading.

The plots include all the evaluated baselines with the exception of the forecast baseline.[46] These initiative-level plots are discussed more extensively in the Integrated Impact and Process Evaluation Findings Section 4.5.2 where these plots are replicated and compared to plots of customers with and without weather correlation.

In the three following plots, the key takeaways are:

- The asymmetric baseline is consistently higher than actual load and all other baselines. This is a result of its inherent upward bias.
- The symmetric baseline is identical to actual load during the adjustment period but diverges in hours further away from the adjustment hour. This is likely due to the

[45] To the extent the asymmetric baseline used the symmetric baseline (i.e., for upward shifts), it will include whatever bias is present due to pre-event load changes related to the event. There is additional upward bias when the asymmetric baseline forgoes a downward shift in favor of the unadjusted baseline.
[46] Because the forecast baseline is a de-rated load-reduction estimate based on the symmetrically adjusted baseline, it does not make sense to represent it as a baseline.

generally flatter shape of the 10 of 10 baseline based on milder days but could also include bias due to pre-event load altering (this is discussed further in Section 4.5.2).

- The regression baseline consistently underestimates actual load throughout the day. While it underestimates load by under 5%, it reduces the load reduction estimates by 25% or more.

Figure 4-4. Eversource Massachusetts – July 30 Event

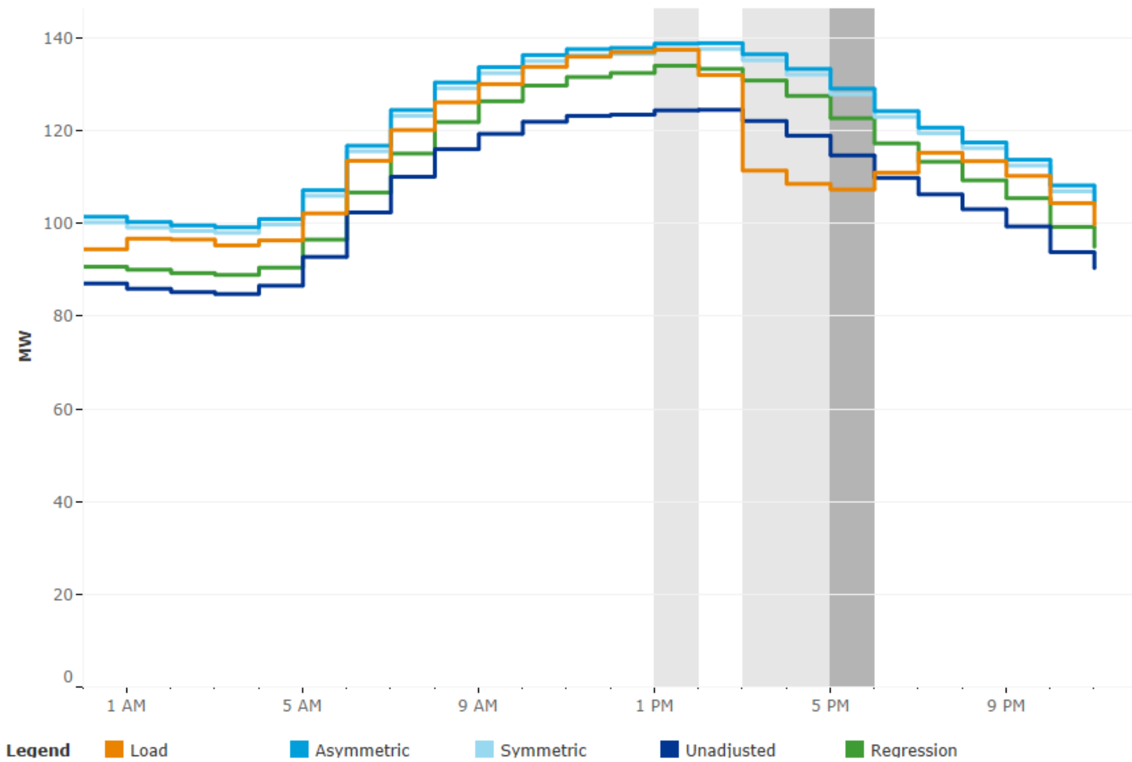


Figure 4-5. Eversource New Hampshire – July 30 Event

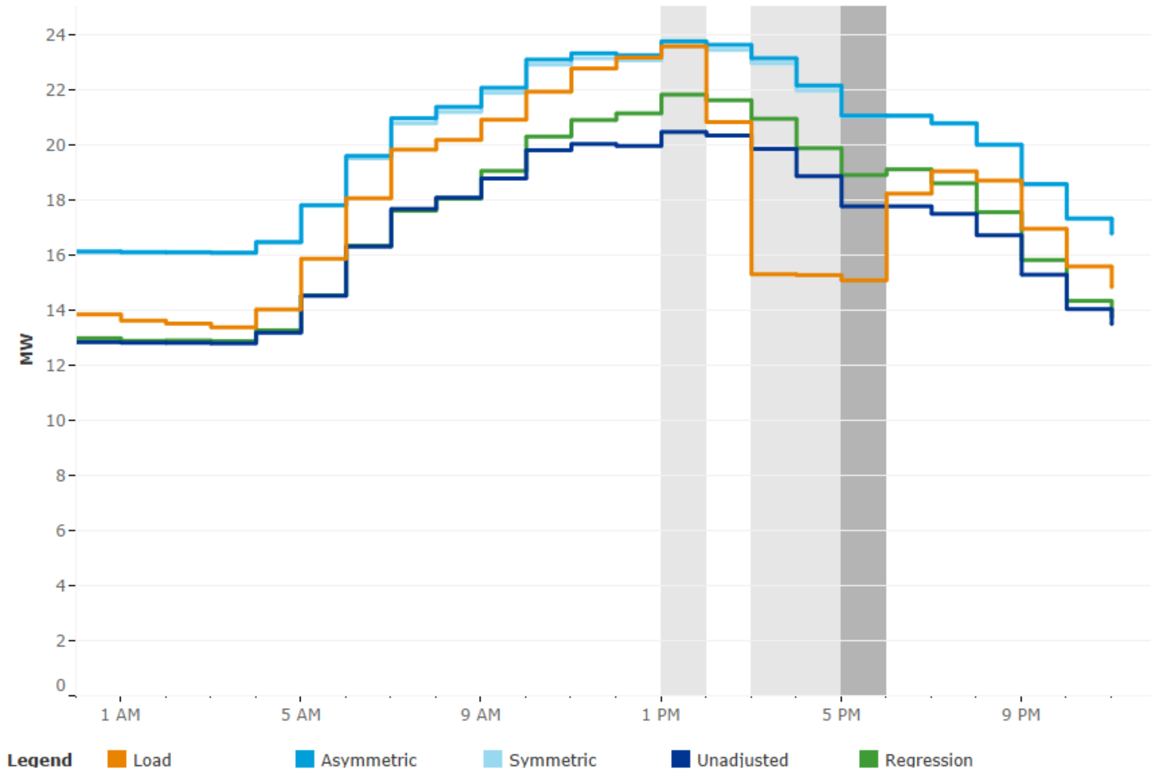
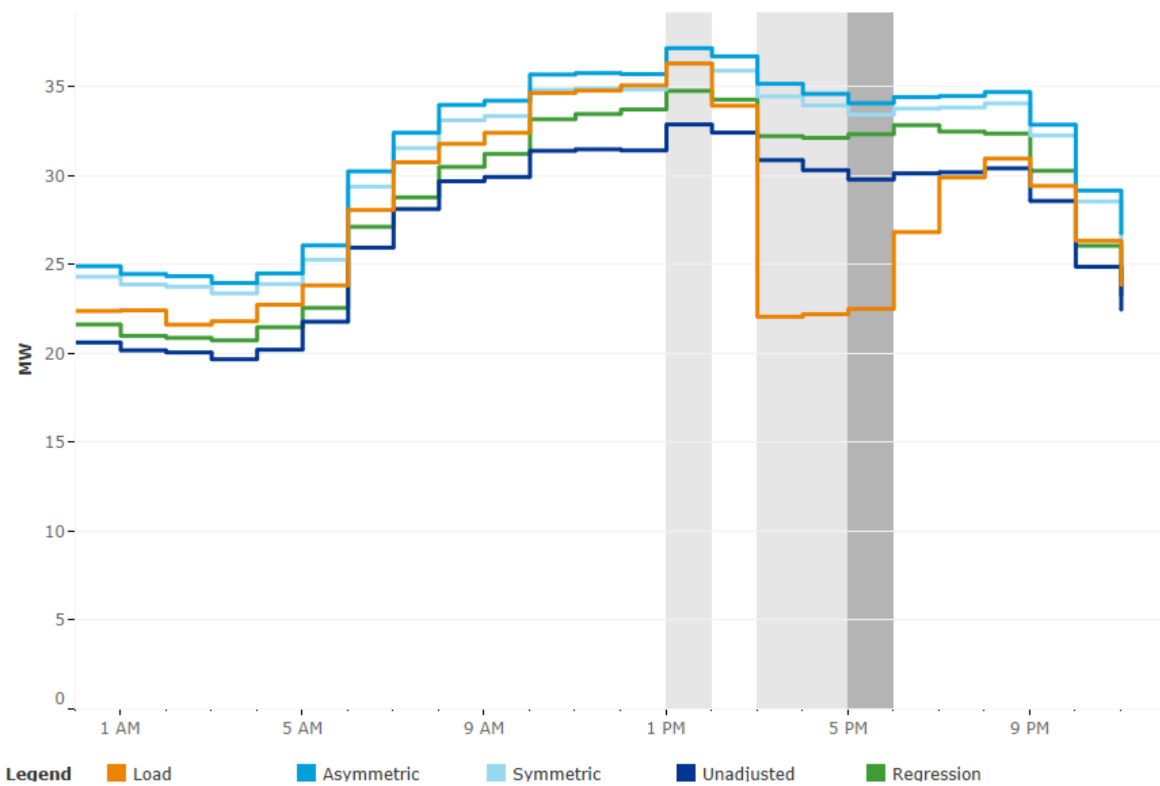


Figure 4-6. Eversource Connecticut – July 30 Event



The evaluation team considers the Evaluated-Symmetric load reduction estimates the most appropriate measure of ex-post load reduction estimates of the 2019 event days. Aside from the potential for potential biases related to adjustment period load movement (up or down), this baseline provides the most accurate and least variable estimate of load reduction. With limited to no evidence of gaming or pre-cooling, the concern of over-estimating load reduction is small. The evaluation team recognizes that this approach is likely to under report load reduction for assets that reduce load to prepare for the event prior to the adjustment period; however, this is a much smaller magnitude concern than the bias associated with only applying positive adjustments, per the asymmetrically adjusted baseline, thereby truncating the natural variability of asset loads. That is, there is limited evidence of clear early shutdown activity on these event days and instead, the unadjusted baseline is most frequently the chosen baseline when natural variability has forced load levels below the unadjusted baseline levels with no evidence of dispatch activity. This truncation replaces reasonable, small to slightly negative load reduction estimates with larger, positive estimates of load reduction where there is no evidence of dispatch activity. The reason ISOs use a symmetrically adjusted baseline is because of the importance of maintaining an unbiased estimate of load reduction.

The evaluation team considers the evaluated-forecast load reduction estimates the best forward-looking load reduction estimates based on the evidence from the 2019 event days. The evaluation has moved away from using the regression approach as the primary evaluated result because it consistently underestimates load across the event days. This produces an estimate of load reduction that is most likely to be downwardly biased due to the inability of the regression to appropriately characterize customer load. Furthermore, inability to estimate load has a magnified effect on estimates of load reduction.

Table 4-4 provides performance ratios for Eversource based on the results in Table 4-3. These performance ratios are described in the Savings and Realization Rates section of the Impact Evaluation Methodology and Framework in this report. The enrollment ratio for Eversource MA was lower than the others because a small number of large customers underperformed significantly.

Table 4-4. Summary of Performance Ratios for Eversource

PA and State	Enrollment Ratio (Reported Asymmetric / Enrolled Capacity)	Asymmetric Ratio (Evaluated Asymmetric / Reported Asymmetric)	Retrospective Realization Rate (Evaluated Symmetric / Reported Asymmetric)	Prospective Realization Rate (Evaluated Forecast / Reported Asymmetric)
Eversource MA	59.8%	89.4%	78.3%	74.2%
Eversource NH	87.3%	109.8%	99.8%	96.1%
Eversource CT	104.5%	92.9%	87.6%	86.6%

The following set of tables (Tables 4-5, 4-6, and 4-7) present event day results at the state-level for Eversource. The event-level results demonstrate the high level of variability across event days. In Massachusetts, the evaluated symmetric load reduction estimates were almost double on the ICAP day (July 30), compared to a week and a half earlier on July 19. The other states' event results are less extreme, with the same comparison of days in Connecticut showing only a 7% increase. These differences are caused by some combination of different weather across the states, a different mix of weather sensitivity in the asset populations of each state, and different non-weather-correlated variability. Missing data among New Hampshire participants resulted in the inclusion of only 19 of 40 customers in the analysis of performance on August 19. Note that the average event load reduction on the ICAP day was greater than load reduction during the ICAP hour for each state. From the graphs of the ICAP day, the baselines show that load generally begins to decrease by 2 p.m., which likely reduces the amount of possible load reduction. Note, observed load is relatively flat during the event period. As the baselines decrease relative to observed load during the event period, less load reduction is estimated later in the event than earlier in the event.

Table 4-5. Eversource Events – Massachusetts

Event Date	7/19/2019	7/30/2019		8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	ICAP Hour Reduction (kW)	Average Hourly Reduction (kW)
Enrolled Capacity	37,223	36,683	36,683	33,870
Reported-Asymmetric	18,155	24,990	N/A	22,137
Evaluated-Validation	19,500	24,346	N/A	18,847
Evaluated-Asymmetric	15,557	23,877	21,760	18,656
Evaluated-Unadjusted	4,048	9,489	7,371	6,267
Evaluated-Symmetric	11,705	22,652	20,558	16,385
Evaluated-Forecast	11,096	21,290	19,380	15,713
Evaluated-Regression	5,313	17,929	15,385	14,586
Accounts	**160**	**151**	**151**	**135**

Table 4-6. Eversource Events – New Hampshire

Event Date	7/19/2019	7/30/2019		8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	ICAP Hour Reduction (kW)	Average Hourly Reduction (kW)
Enrolled Capacity	5,905	5,905	5,905	2,680

Reported-Asymmetric	4,988	5,552	N/A	1,814
Evaluated-Validation	4,796	7,012	N/A	1,679
Evaluated-Asymmetric	4,619	6,902	5,981	1,659
Evaluated-Unadjusted	2,885	3,610	2,689	686
Evaluated-Symmetric	4,017	6,769	5,947	1,013
Evaluated-Forecast	3,962	6,427	5,674	970
Evaluated-Regression	4,273	4,692	3,823	1,033
Accounts	**40**	**40**	**40**	**19**

Table 4-7. Eversource Events – Connecticut

Event Date	7/19/2019	7/30/2019		8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	ICAP Hour Reduction (kW)	Average Hourly Reduction (kW)
Enrolled Capacity	12,469	12,519	12,519	12,249
Reported-Asymmetric	12,575	13,135	N/A	13,085
Evaluated-Validation	12,331	12,663	N/A	12,271
Evaluated-Asymmetric	11,898	12,347	11,564	11,822
Evaluated-Unadjusted	8,434	8,058	7,275	7,133
Evaluated-Symmetric	10,901	11,682	10,921	11,393
Evaluated-Forecast	10,764	11,495	10,779	11,322
Evaluated-Regression	9,118	9,963	9,824	9,383
Accounts	**95**	**96**	**96**	**95**

Prior Year Results

Prior year results illustrate how expected load reduction may vary due to the timing of events across baselines. Figures 4-7 and 4-8 show event results for Eversource's 2018 demand demonstration in Massachusetts. Load reduction estimates for the unadjusted, symmetrically adjusted, and regression baselines are shown for each event. Note, the Eversource demand demonstration used the symmetrically adjusted baseline for settlement, in contrast to the use of the asymmetrically adjusted baseline by the ADR initiative. As such, results for the asymmetrically adjusted baseline are not shown here. Importantly, while load reduction estimates vary significantly across events for each baseline, load reduction for the symmetrically adjusted baseline is always greater than regression baseline, with the exception of the July 2, July 3, and July 5 events that occurred during the week of the July 4 holiday. For these events, the regression baseline is greater than the symmetrically adjusted baseline because it is impacted less by the shutdowns that occur around July 4.

Figure 4-7. Eversource Massachusetts – 2018 Demand Demonstration Results

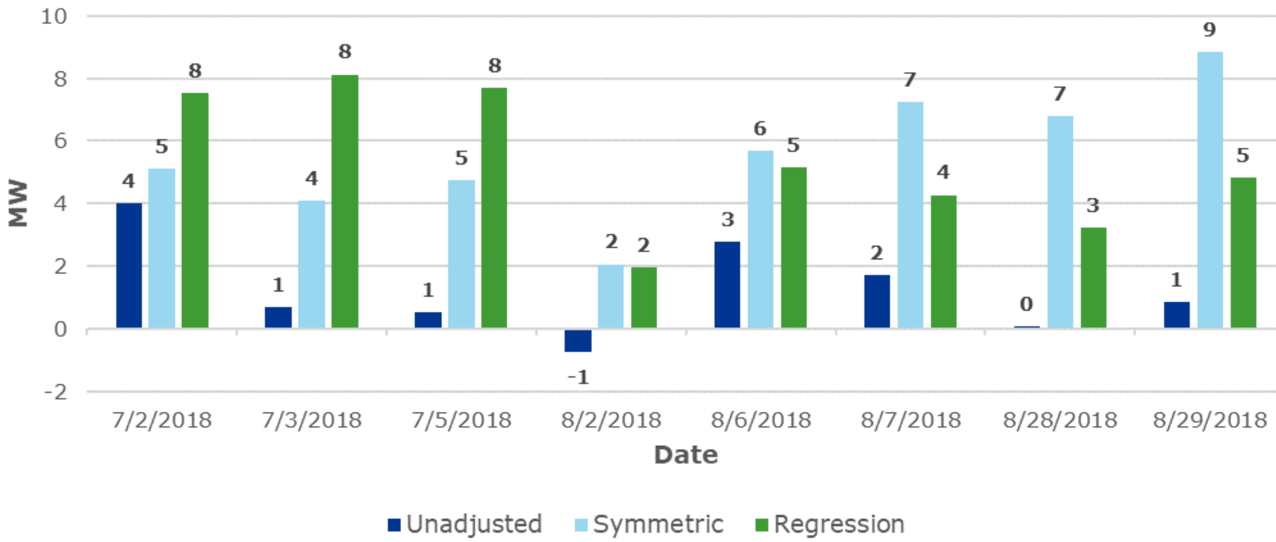


Figure 4-8. Eversource Massachusetts – 2018 Demand Demonstration Results

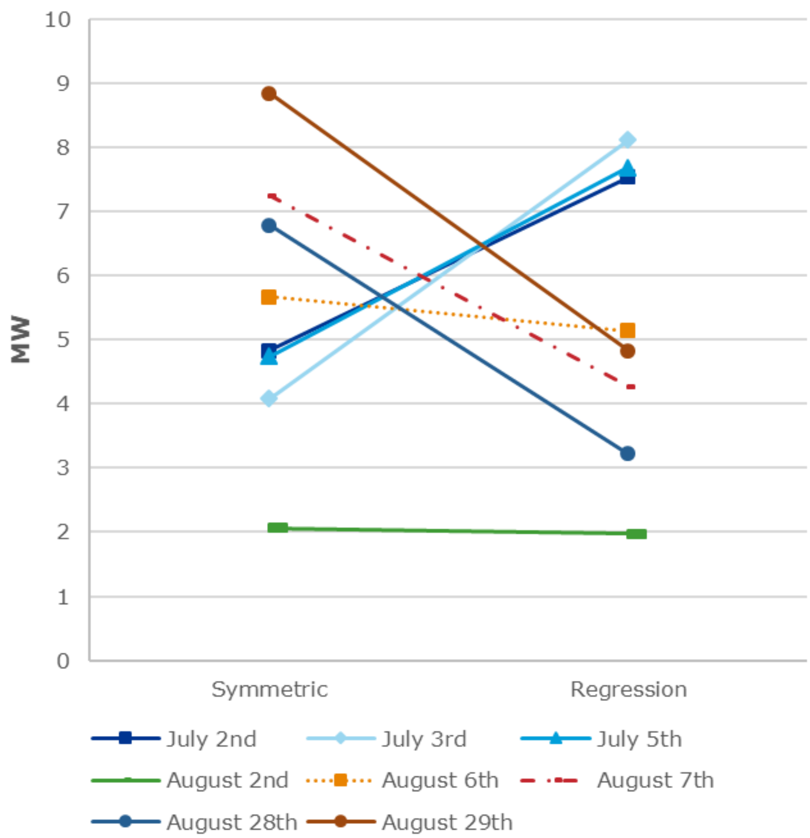
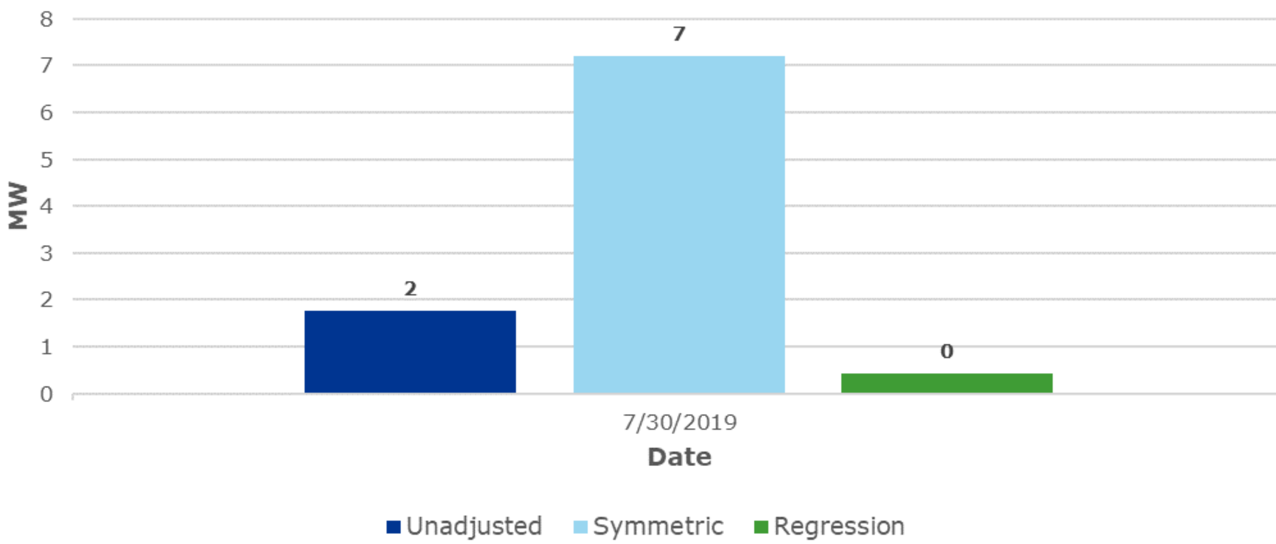


Figure 4-9 shows the event results for Eversource's 2019 demand demonstration in Massachusetts. The significant deviation in load reduction estimated by the symmetrically adjusted baseline compared to the unadjusted and regression baselines is due to a large cluster of related assets with highly variable load. The unadjusted and regression baselines significantly underestimate load for these highly variable assets, whereas the symmetrically adjusted baseline is, by definition, adjusted to observed load during the pre-event hours. The stark differences in load reduction between the adjusted baseline and unadjusted and regression baselines illustrates the power of the same-day adjustment. Note that the accuracy of the symmetrically adjusted baseline ultimately depends on the unadjusted baseline shape.

Figure 4-9. Eversource Massachusetts – 2019 Demand Demonstration Results



National Grid

National Grid called a single event during the summer 2019 season on July 30, the ICAP day. Figure 4-10 is a visual representation of the July 30 event (ICAP day). The figure shows actual load (orange line), the unadjusted 10-of-10 baseline (dark blue line), the symmetrically adjusted baseline (light blue line), the asymmetrically adjusted baseline (cyan line), and the regression baseline (green line). The adjustment window and event period are shaded in gray. The ICAP hour occurred on July 30, from 5 to 6 p.m., and is identified with darker shading. The figure demonstrates that load reduction decreases for later hours of the event. The rolling baselines show that actual load for this group of customers tends to decrease on similar non-event days beginning as early as 2 p.m. As load decreases, it is likely that the achievable load reduction also decreases.

The key takeaways are slightly different than for Eversource:

- The asymmetric baseline is consistently higher than actual load and all other baselines. This is a result of its inherent upward bias.
- The symmetric baseline is identical to actual load during the adjustment period. This baseline diverges less from actual load in hours further away from the adjustment hour, as

opposed to the hours immediately preceding the adjustment period and the hour between the adjustment and event periods. Unlike Eversource, there is a downward step in actual load during the adjustment hour that could indicate greater pre-event load reduction during that hour than load increases like pre-cooling or gaming.

- The regression baseline consistently underestimates actual load throughout the day. While it underestimates load by under 5% it reduces the load reduction estimates by more than 15%.

Figure 4-10. National Grid - Massachusetts - July 30th Event

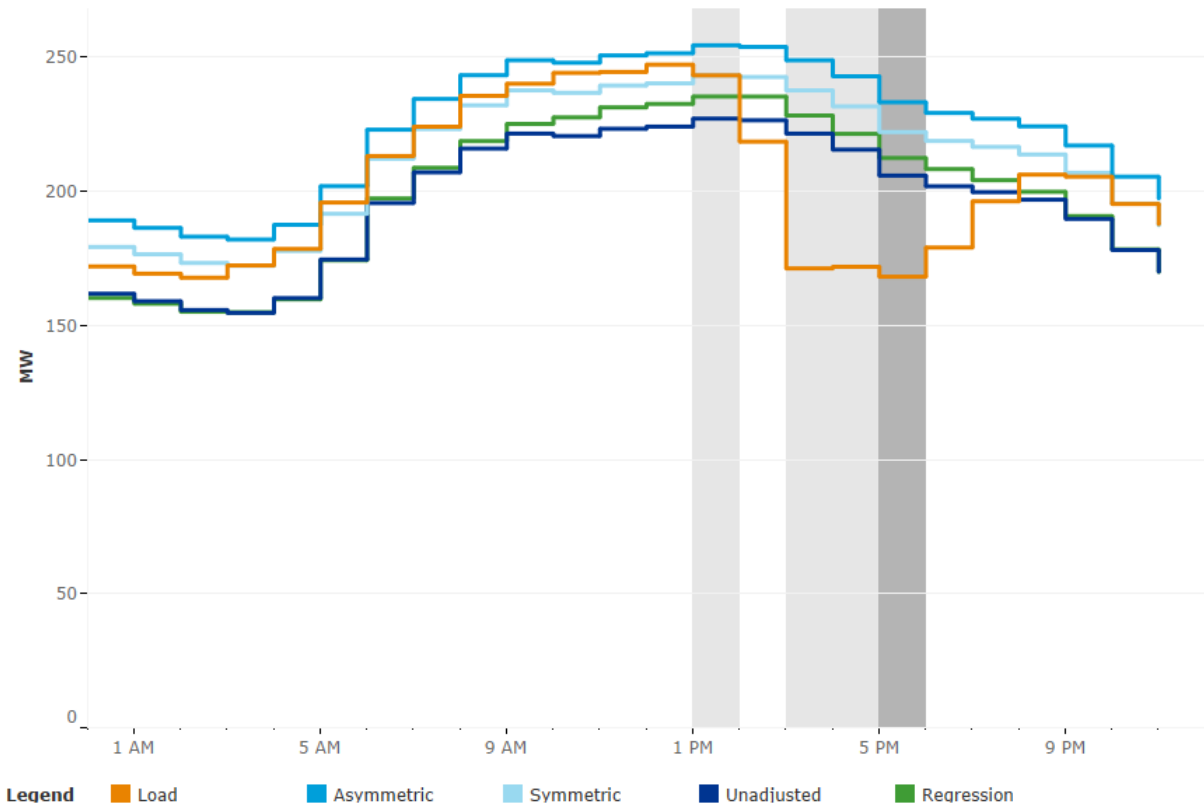


Table 4-8 provides the summary of National Grid's load reduction estimates for Massachusetts. The reported results were only provided on an average event basis, so reported ICAP reductions are not included in the table. All results shown in the table are based on the assets that the evaluation team had sufficient data to estimate load reduction. The number of assets or accounts with sufficient data is shown in the bottom row of the table.

Table 4-8. National Grid Impact Summary – Massachusetts

Result	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
Enrolled Capacity	93,134	93,134
Reported-Asymmetric	71,428	N/A
Evaluated-Validation	71,611	N/A
Evaluated-Unadjusted	42,461	36,090
Evaluated-Asymmetric	69,561	63,190
Evaluated-Symmetric	58,464	52,173
Evaluated-Forecast	57,264	51,266
Evaluated-Regression	48,752	42,538
Accounts	**357**	**357**

The validation line is the first set of estimates in the table provided by the evaluation team. One asset accounts for the difference between the reported asymmetric and validated results. National Grid reported that performance for this asset was disputed. The evaluated-validation results are calculated and provided at the event-level for consistency with the reported results.

4.5 The evaluated-asymmetric results are lower than the validated results reflecting the application of recommended statewide data sufficiency and calculation rules. For National Grid, the difference between the evaluated validation and evaluated asymmetric results is primarily driven by the inclusion of negative load reduction estimates in the evaluated results, rather than setting them to zero. Reported asymmetric results are provided at the event-level for each asset. In instances that an asset is estimated to have negative load reduction, the result is reported as zero, as opposed to the actual value estimated. When the event results are aggregated to the initiative-level, the cumulative effect can be significant. As discussed in Section 3.2.1 Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

4.5.1 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states

- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned
- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs
- Other opportunities for peak demand management

4.5.2 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology (Baselines), negative load reduction estimates are a natural part of baseline estimation. Baseline estimation error is always present but is only visible when negative load reduction is estimated, which is only likely to occur for an asset that did not take any load reduction action. These negatives need to be included to avoid another form of upward bias on the estimates. The evaluated-validation results are calculated and provided at the event level for consistency with the reported results. The evaluated-asymmetric results also implement the use of lookback days, whereas the validated results did not use lookback days in order to replicate the reported asymmetric results.

The evaluated-symmetric load reduction estimates are the best ex post load reduction estimates of the 2019 event days. The final estimates are lower than the evaluated-asymmetric results because of the inherent upward bias in the asymmetric baseline.

The evaluated-forecast results further adjusted the symmetric results to address the likelihood of shutdown. For a forward-looking estimate of expected load reduction, given the empirical evidence available from this summer, this is the most informed estimate possible. In the future, with additional information about participating assets and more empirically based load reduction estimates, improved forecast estimates may be possible.

Table 4-49 provides performance ratios for National Grid based on the results in Table 4-8. These performance ratios are described in the Savings and Realization Rates section of the Impact Evaluation Methodology and Framework in this report.

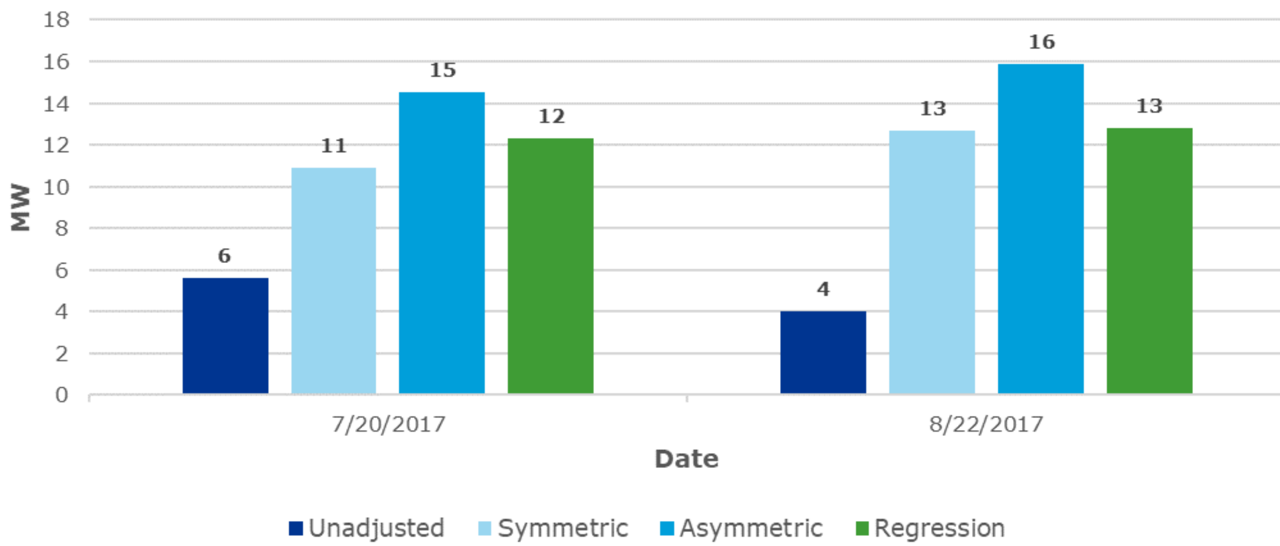
Table 4-9. Summary of Performance Ratios for National Grid

PA and State	Enrollment Ratio (Reported Asymmetric / Enrolled Capacity)	Asymmetric Ratio (Evaluated Asymmetric / Reported Asymmetric)	Retrospective Realization Rate (Evaluated Symmetric / Reported Asymmetric)	Prospective Realization Rate (Evaluated Forecast / Reported Asymmetric)
National Grid MA	76.7%	97.4%	81.9%	80.2%

Prior Year Results

National Grid's Massachusetts demonstration started in 2017. Figure 4-11 shows results from the 2017 demonstration. The results were similar across the two event days. The upward bias of the asymmetric baseline estimates is substantial in both cases. The most surprising aspect of the result is that regression results were similar or higher than the symmetrically adjusted estimates.

Figure 4-11. National Grid - Massachusetts - 2017 Results



Figures 4-12 and 4-13 show results from National Grid's 2018 Massachusetts demand demonstration. Figure 4-12 shows results for the unadjusted, symmetrically adjusted, asymmetrically adjusted, and regression baselines. There were significant missing data issues for the 2018 evaluation. Given this, these results are most useful to compare how load reduction estimates vary across events, relative to the baselines. Similar to Eversource's 2018 Massachusetts demand demonstration, the regression estimates are highest on the July 3 and July 5 events around the July 4 holiday. The degree to which the rolling baselines (unadjusted, symmetric, and asymmetric) are affected by shutdowns around the holiday are one of the drivers of this result. Outside of those two events, the asymmetric baseline estimates the most amount of load reduction. As expected, the symmetrically adjusted baseline generally estimates more load reduction than the unadjusted baseline, with the exception of the August 6 event.

Figure 4-12. National Grid - Massachusetts - 2018 Results

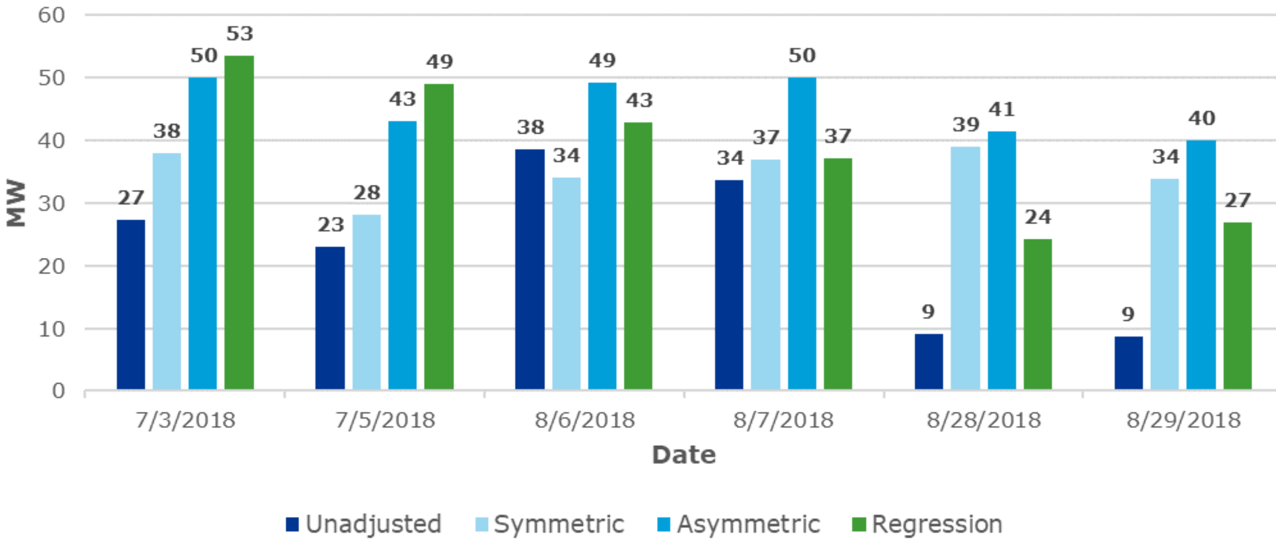
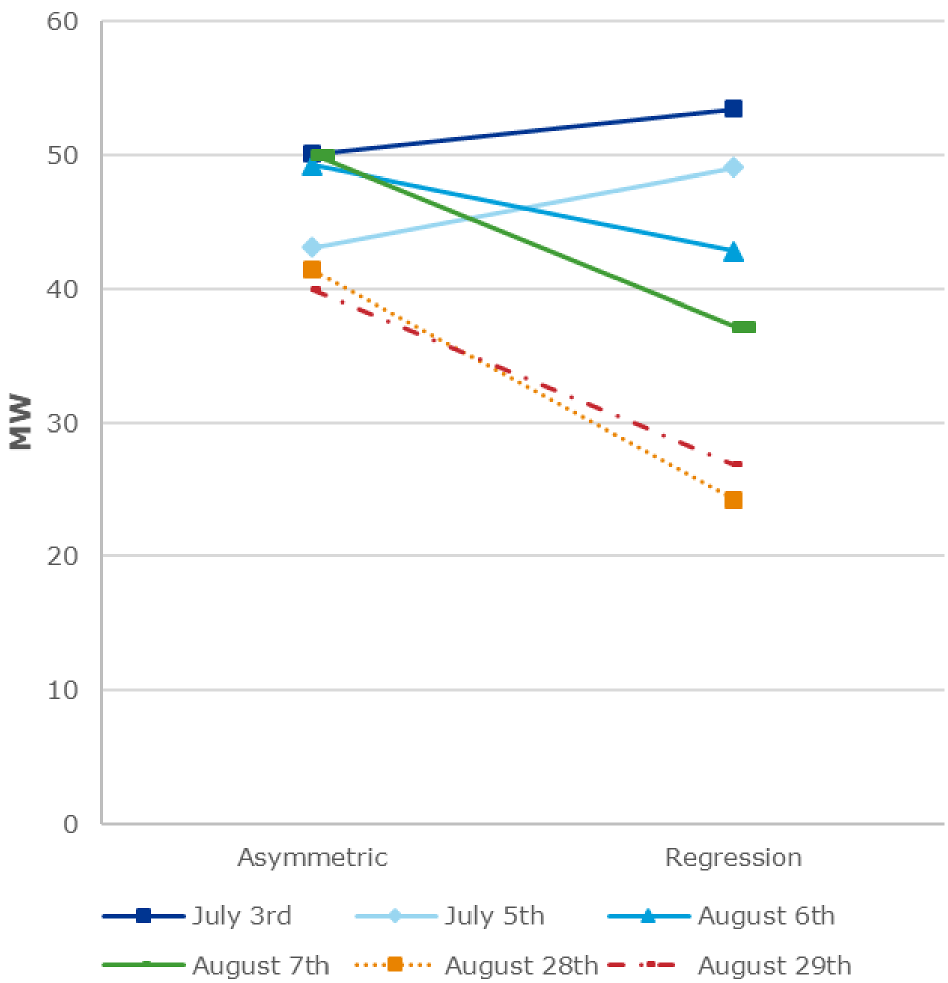


Figure 4-13 only shows the 2018 asymmetric and regression load reduction estimates. The visual makes it easy to compare the magnitude and variability of these baselines relative to each other across events. The regression baseline only estimates more load reduction than the asymmetric baseline on the two events around July 4th. Figure 4-13Figure 4-13. also clearly demonstrates that the regression estimates have significantly more variability than the asymmetric results. This is not the case for the similar figure for Eversource where the settlement baseline was symmetrically adjusted. The apparent lower variation in the asymmetric baseline is an artifact of the upward bias in the algorithm.

Figure 4-13. National Grid - Massachusetts - 2018 Results



4.2.3.3 Unitil

Unitil called a single event during the summer 2019 season on July 30th, the ICAP day. Figure 4-14 is a visual representation of that day for Unitil in Massachusetts. The figure shows actual load (orange line), the unadjusted 10-of-10 baseline (dark blue line), the symmetrically adjusted baseline (light blue line), the asymmetrically adjusted baseline (cyan line), and the regression baseline (green line). The adjustment window and event period are shaded in gray. The ICAP hour occurred on July 30th, from 5 to 6 p.m., and is identified with darker shading. The figure demonstrates that load reduction decreases for later hours of the event. Figure 4-15, below, shows the same information for Unitil in New Hampshire.

In the three following plots, the key takeaways are:

- The asymmetric load is consistently higher than all other baselines and only below actual load for a few morning hours. This is a result of its inherent upward bias.

- There is substantial variability in the actual load shape compared to 10 of 10 baselines. The Massachusetts unadjusted baseline is adjusted by 10% but appears to be approximately the right shape. The NH adjustment is barely detectable. Note, the unadjusted and symmetrically adjusted baseline shapes diverge due to the baseline floor of 0 being utilized in one instance where the adjustment would otherwise be negative during some hours.
- The regression baseline consistently underestimates actual load throughout the day especially in New Hampshire. Like the other PAs, a relatively small underestimate of customer load leads to substantial reductions in estimated load reduction.

Figure 4-14. Unitil - Massachusetts - July 30th Event

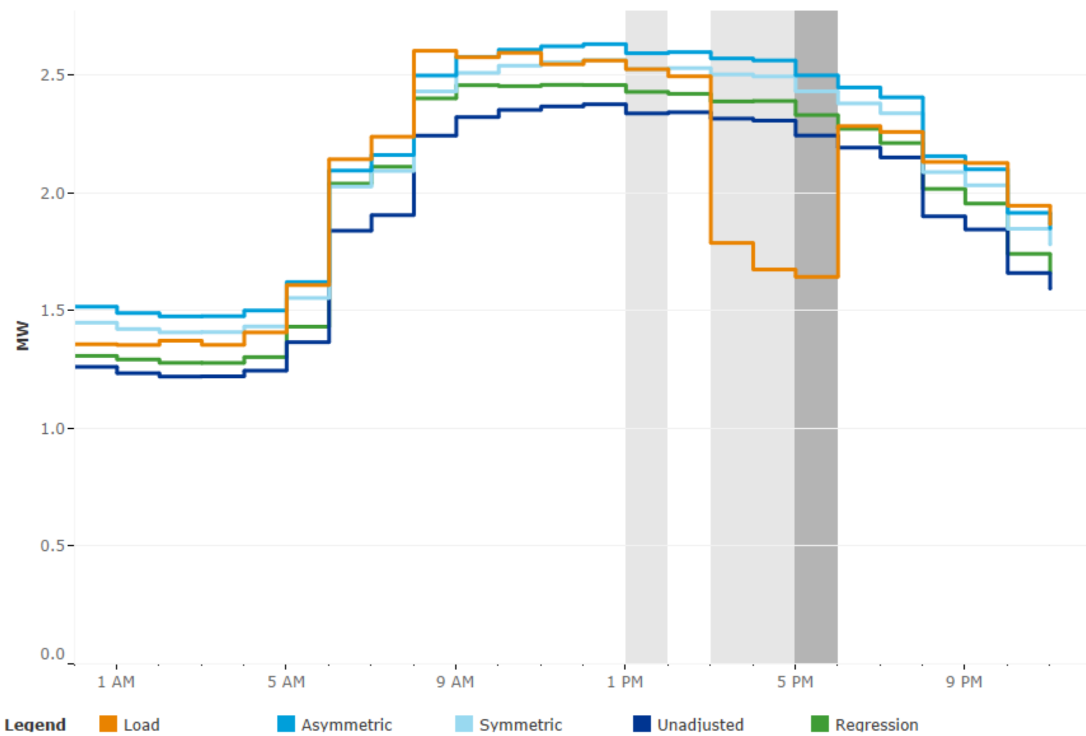


Figure 4-15. Unitil - New Hampshire - July 30th Event

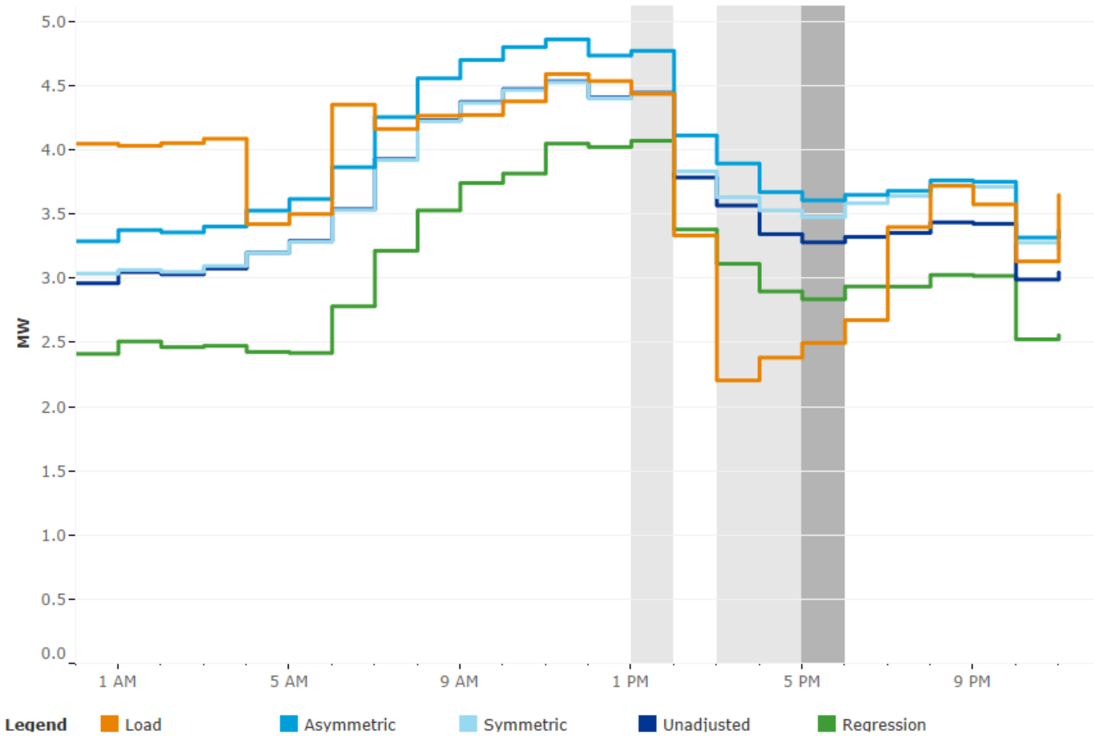


Table 4-10 provides the summary of Unitil's load reduction estimates by state and is structured the same as earlier results tables. Unitil only called a single event on July 30, the ICAP day. The reported results were only provided on an average event basis, so reported ICAP reductions are not included in the table. The evaluated-validation results are calculated and provided at the event level for consistency with the reported results. The evaluation team had sufficient data to estimate load reduction for all assets.

Table 4-10. Unitil Impact Summary

	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
Result	**MA**		**NH**	
Enrolled Capacity	950	950	1,600	1,600
Reported-Asymmetric	853	N/A	1,299	N/A
Evaluated-Validation	843	843	1,363	N/A
Evaluated-Unadjusted	587	601	1,036	784
Evaluated-Asymmetric	843	856	1,363	1,111
Evaluated-Symmetric	775	788	1,185	980
Evaluated-Forecast	775	788	1,153	949
Evaluated-Regression	668	687	588	341
Accounts	**3**	**3**	**7**	**7**

The validation line is the first set of estimates in the table provided by the evaluation team and is based on the number of assets reported in the last row. The validated values for Unitil were very close to the reported asymmetric values. Differences between the validated and reported asymmetric values may be due to differences in the treatment of interval data timestamps with respect to daylight savings time or variation in the underlying data. More details on these discrepancies are provided in Appendix B: Settlement Verification.

The application of recommended statewide data sufficiency and calculation rules does not change the validated results. All assets had sufficient data in the most 10 recent eligible baseline days used to estimate load reduction and the reported results contained negative load reduction, where it occurred.

The evaluated-symmetric load reduction estimates are the best ex post load reduction estimates of the 2019 event days. The final estimates are lower than the evaluated-asymmetric results because of the inherent upward bias in the asymmetrical baseline.

The evaluated-forecast results further adjusted the symmetric results to address the likelihood of shutdown. For a forward-looking estimate of expected load reduction, given the empirical evidence available from this summer, this is the most informed estimate possible. In the future, with additional information about participating assets and more empirically based load reduction estimates, improved forecast estimates may be possible.

Table 4-411 provides performance ratios for Unitil based on the results in Table 4-10. These performance ratios are described in the Savings and Realization Rates section of the Impact Evaluation Methodology and Framework in this report.

Table 4-11. Summary of Realization Rates for Unitil

PA and State	Enrollment Ratio (Reported Asymmetric / Enrolled Capacity)	Asymmetric Ratio (Evaluated Asymmetric / Reported Asymmetric)	Retrospective Realization Rate (Evaluated Symmetric / Reported Asymmetric)	Prospective Realization Rate (Evaluated Forecast / Reported Asymmetric)
Unitil MA	89.8%	98.8%	90.8%	90.8%
Unitil NH	81.2%	104.9%	91.2%	88.7%

4.2.3.4 State-Level

Table 4-12 summarizes the results at the state level. This table illustrates the relative levels of load reduction provided in each state as well as an overall picture of the evaluation results compared to reported results. The more disaggregate tables reported above as well the Settlement Verification Appendix provide more detail on the adjustments that produce the evaluated results. A summary of performance ratios is provided at the PA/state level below in Table 4-13.

Table 4-12. State-Level Impact Summary

Result	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)	Event Average Reduction (kW)	ICAP Hour Reduction (kW)
State	**MA**		**NH**		**CT**	
Enrolled Capacity	131,332	130,767	7,505	7,505	12,519	12,519
Reported - Asymmetric	94,543	N/A	6,455	N/A	13,085	N/A
Evaluated - Validation	93,982	N/A	7,149	N/A	12,558	N/A
Evaluated - Unadjusted	49,609	44,062	4,238	3,473	7,951	7,275
Evaluated - Asymmetric	90,316	85,805	7,024	7,092	12,158	11,564
Evaluated - Symmetric	76,671	73,519	6,332	6,927	11,462	10,921
Evaluated - Forecast	74,562	71,435	6,106	6,623	11,330	10,779
Evaluated - Regression	62,169	58,610	5,035	4,164	9,594	9,824
Program Administrators	**3**	**3**	**2**	**2**	**1**	**1**

Table 4-13. Summary of Performance Ratios

PA and State	Enrollment Ratio (Reported Asymmetric / Enrolled Capacity)	Asymmetric Ratio (Evaluated Asymmetric / Reported Asymmetric)	Retrospective Realization Rate (Evaluated Symmetric / Reported Asymmetric)	Prospective Realization Rate (Evaluated Forecast / Reported Asymmetric)
Eversource MA	59.8%	89.4%	78.3%	74.2%
Eversource NH	87.3%	109.8%	99.8%	96.1%
Eversource CT	104.5%	92.9%	87.6%	86.6%
National Grid MA	76.7%	97.4%	81.9%	80.2%
Unitil MA	89.8%	98.8%	90.8%	90.8%
Unitil NH	81.2%	104.9%	91.2%	88.7%
MA – All 3 PAs	72.0%	95.5%	81.1%	78.9%
NH – Eversource, Unitil	86.0%	108.8%	98.1%	94.6%
CT – Eversource	104.5%	92.9%	87.6%	86.6%

Figures 4-16 through 4-18 are visual representations of the July 30 ICAP Day event for the ADR initiative in Massachusetts, New Hampshire, and Connecticut, respectively.

Figure 4-16. ADR Initiative – Massachusetts - July 30th Event - ICAP Day

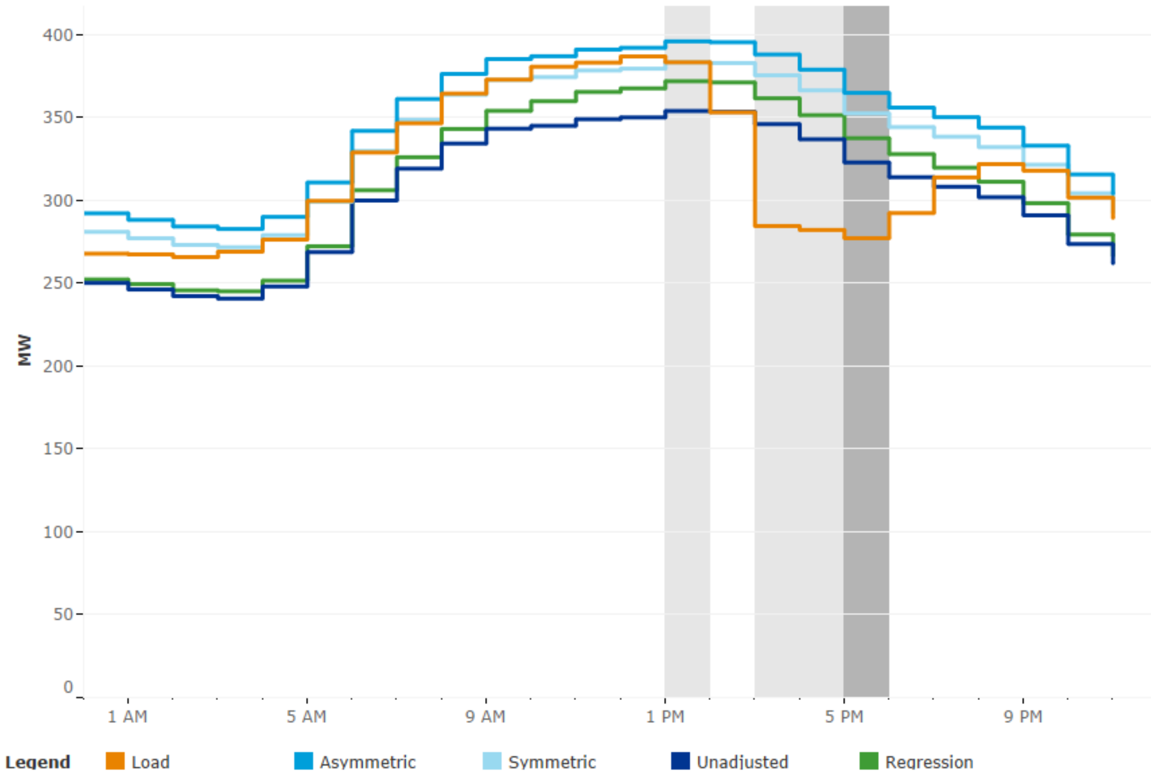


Figure 4-17. ADR Initiative - New Hampshire - July 30th - Event - ICAP Day

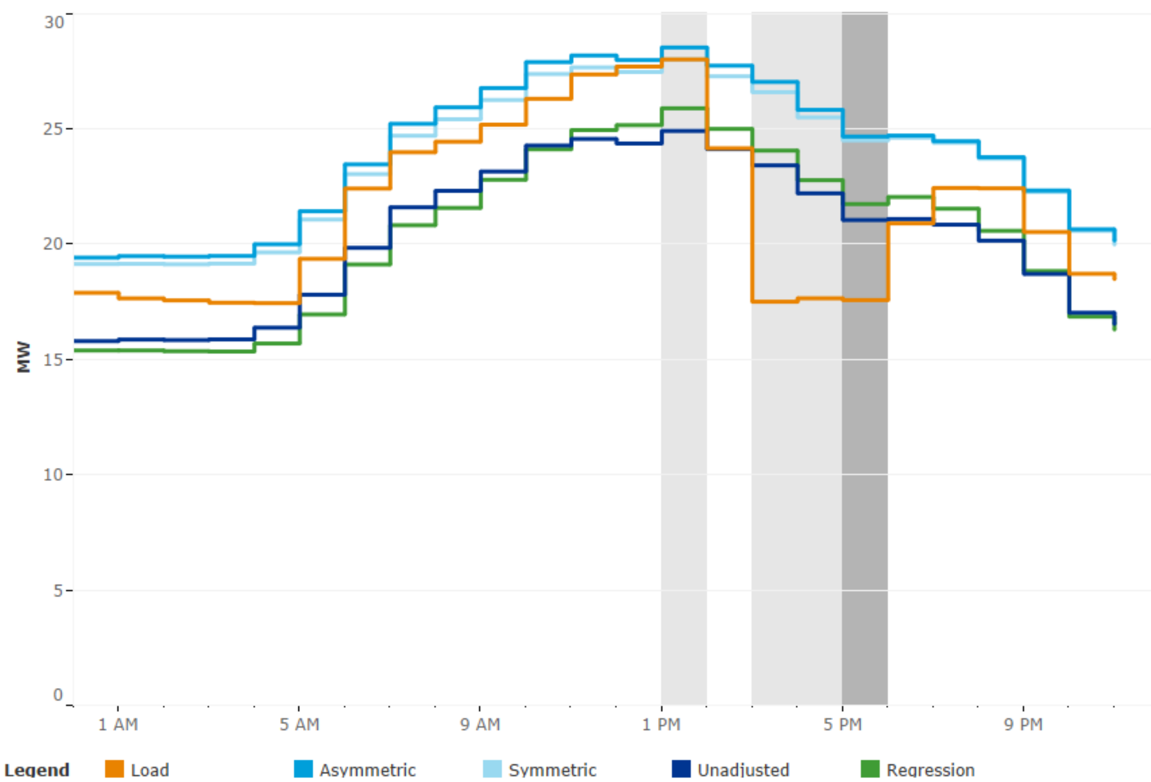
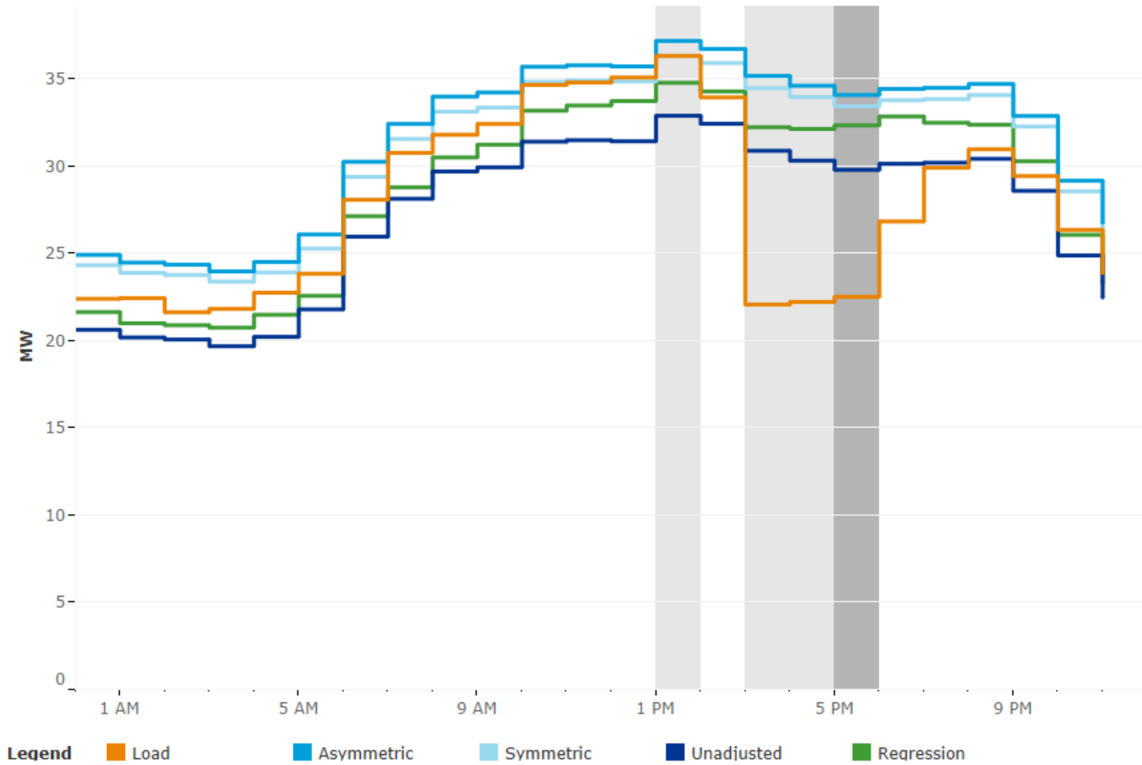


Figure 4-18. ADR Initiative - Connecticut - July 30th Event - ICAP Day



4.5.3 Targeted Battery Storage

There were two targeted battery storage participants in the Summer 2019 initiative. Both participants are Eversource customers in MA. Both customers are Eversource customers and were called three times during the summer season. They were also called for a voluntary weekend event, and both systems were dispatched during the weekend event.

Battery 1

Battery 1 is a 1,500 kW, 3,000 kWh lithium-ion battery installed in a manufacturing facility in southern Massachusetts. This system was installed and operational as of early July 2019 and was dispatched from 4 p.m. to 7 p.m. daily on non-holiday weekdays from early July through September. The facility also participated in the Eversource ConnectedSolutions initiative as a targeted battery storage resource. Table 4-14 summarizes the performance of Battery 1 over the 2019 summer season.

Table 4-14. Battery 1 Results Summary

	PA Event Average Reduction (kW)	Weekend PA Event Average Reduction (kW)	ISO-NE ICAP Hour Reduction (kW)	Net Energy Impact (kWh)
Committed	400	N/A	N/A	N/A
Reported	283	347	800	N/A
Evaluated	295	703	800	84,122*
Hours	**9**	**3**	**1**	**-**

N/A – Not available

*The ratio of discharging energy to charging energy for the 2019 summer season was found to be 0.674.

On days when daily dispatch and targeted events coincided, battery dispatch kW was allocated between the two event types to delineate performance and avoid double-counting. The first 500 kW of battery dispatch was attributed to daily dispatch and battery dispatch exceeding 500 kW was attributed to the targeted dispatch performance. For example, if the total battery dispatch is 550 kW, 500 kW would be attributed to daily dispatch while 50 kW would be attributed to targeted dispatch. All three targeted weekday events overlapped with the daily dispatch schedule (2 events overlapped for all 3 hours; 1 event overlapped for 2 out of 3 hours).

Figure 4-19 shows the performance of the battery on an event day (7/19/2019). The battery performed as expected during the event window from 4 pm to 7 pm.

Figure 4-19. Battery 1 Performance on a Peak Event Day 7/19/2019

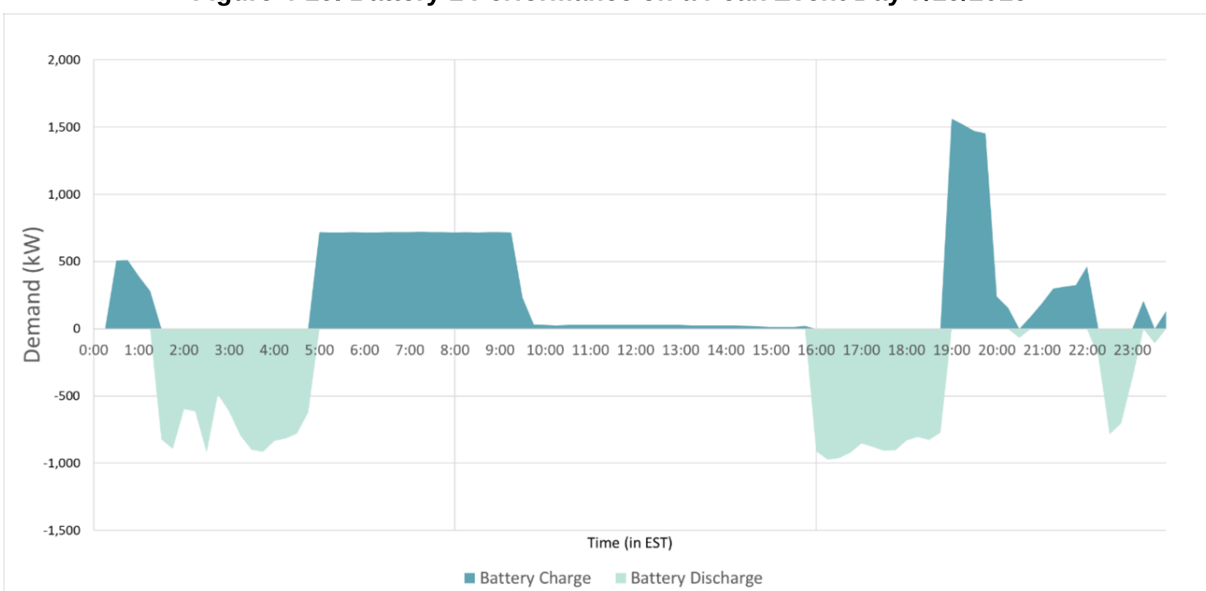


Figure 4-20, below, shows the battery performance on the ICAP day. Although the event was called from 3 pm to 6 p.m. on the ICAP day, the battery system was found to be charging from 3 to 4 p.m. The battery discharged from 4 pm to 6 pm, thereby successfully performing during the ICAP hour (5 to 6 p.m.).

Figure 4-20. Battery 1 Performance on the ICAP Day - 7/30/2019

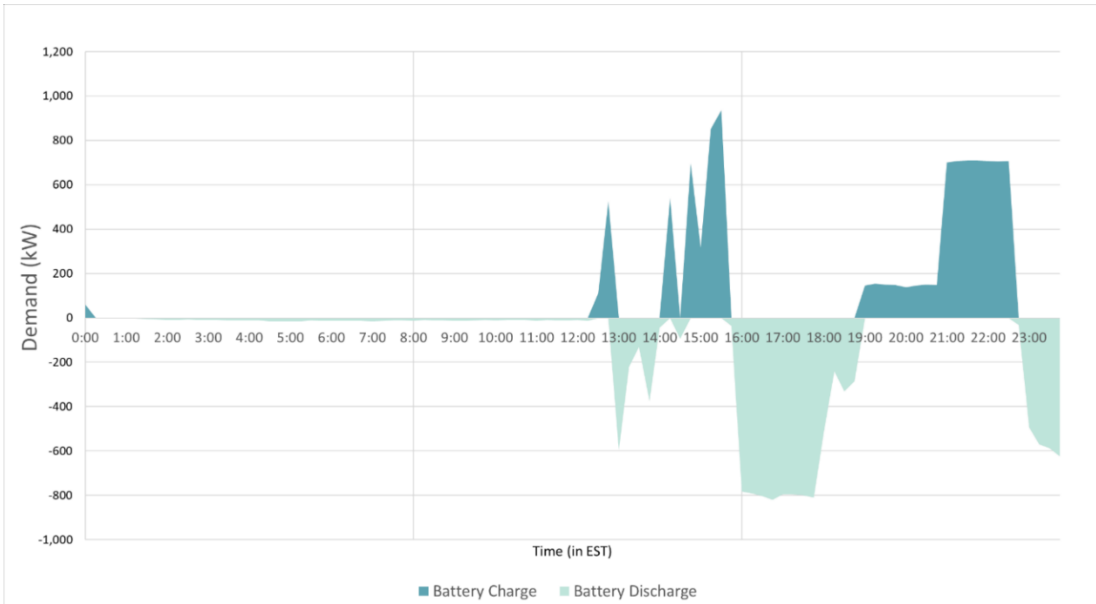
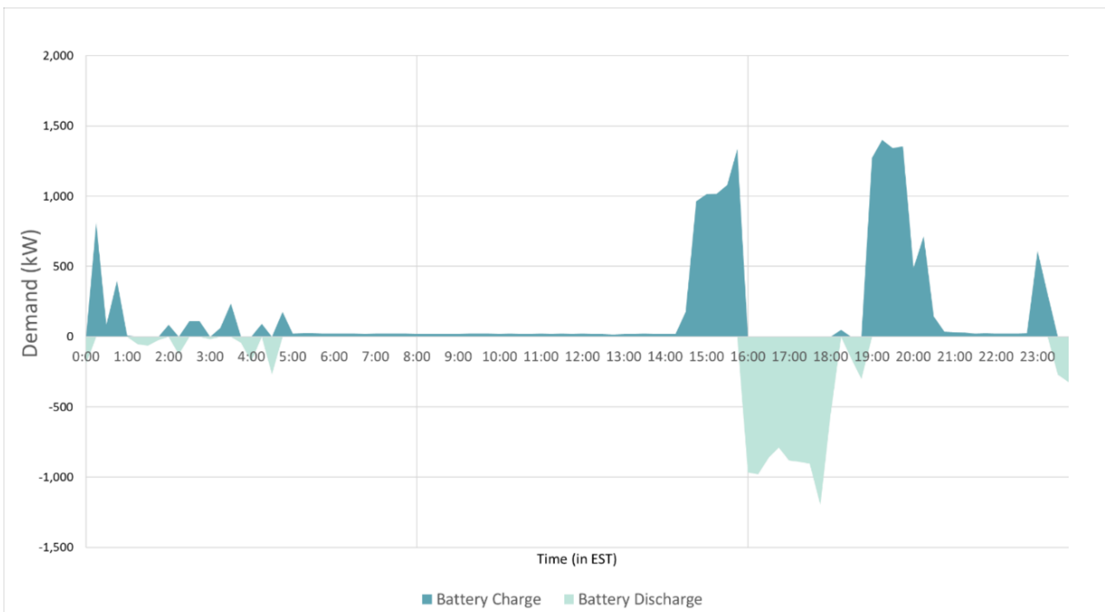


Figure 4-21 shows the battery performance on a weekend peak event day which occurred on 7/20/2019. The battery was successfully dispatched during the weekend event hours.

Figure 4-21. Battery 1 Performance on a Weekend Peak Event Day - 7/20/2019



Battery 2

Battery 2 is a 1,300 kW, 5,600 kWh lithium-ion battery installed in a university in Massachusetts. The facility participated in the Eversource ConnectedSolutions initiative as a targeted storage resource. The vendor for Battery 2 provided the evaluators with the calculated reduction at the end of the season but not the committed reductions. The reported results were found to be accurate. Table 4-15 summarizes the performance of Battery 2 over the 2019 summer season.

Table 4-15. Battery 2 Results Summary

	PA Event Average Reduction (kW)	Weekend PA Event Average Reduction (kW)	ISO-NE ICAP Hour Reduction (kW)	Net Energy Impact (kWh)
Reported	1,033	N/A	1,290	7,000
Evaluated	1,033	1,162	1,290	N/A*

*The ratio of discharging energy to charging energy for the 2019 summer season was reported to be 0.897. Net energy evaluation was not possible due to absence of beginning and end states of charge and charging data for the days preceding the first event day.

Figure 4-22 shows the performance of the battery on an event day (7/19/2019). The battery performed as expected during the event window from 4 pm to 7 pm.

Figure 4-22. Battery 2 Performance on a Peak Event Day - 7/19/2019

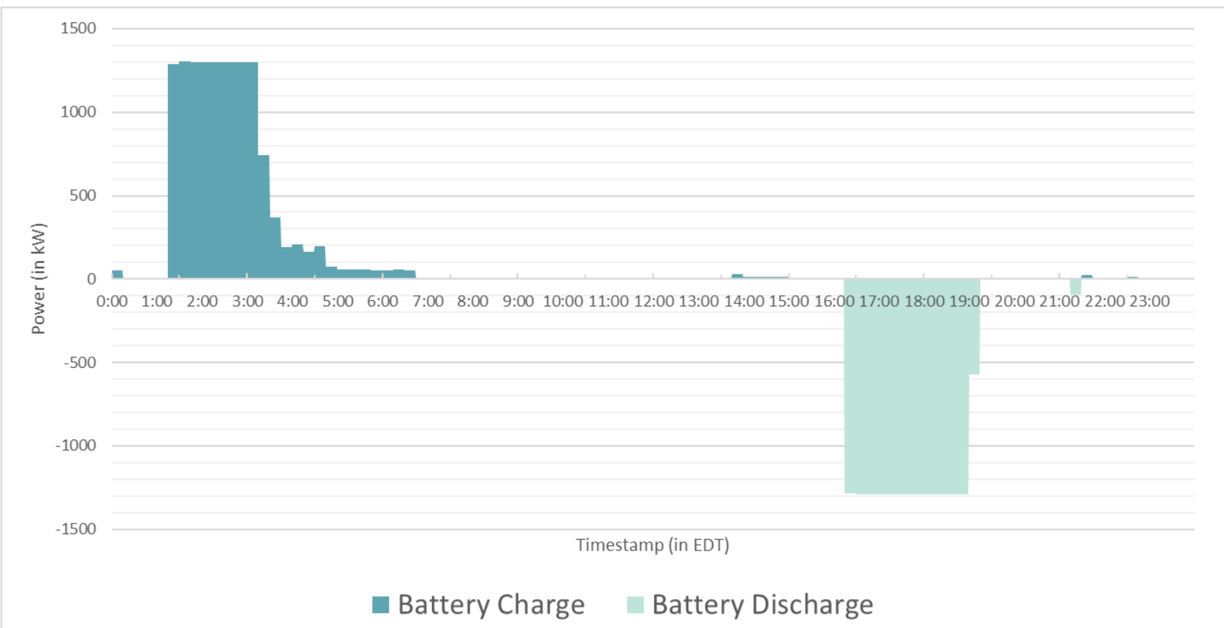


Figure 4-23, below, shows the battery performance on the ICAP day. Although the event was called from 3 pm to 6 pm on the ICAP day, the battery system was not dispatched from 3 pm to 4 pm. The battery discharged from 4 pm to 6 pm, thereby successfully performing during the ICAP hour (5 pm to 6 pm).

Figure 4-23. Battery 2 Performance on the ICAP Day - 7/30/2019

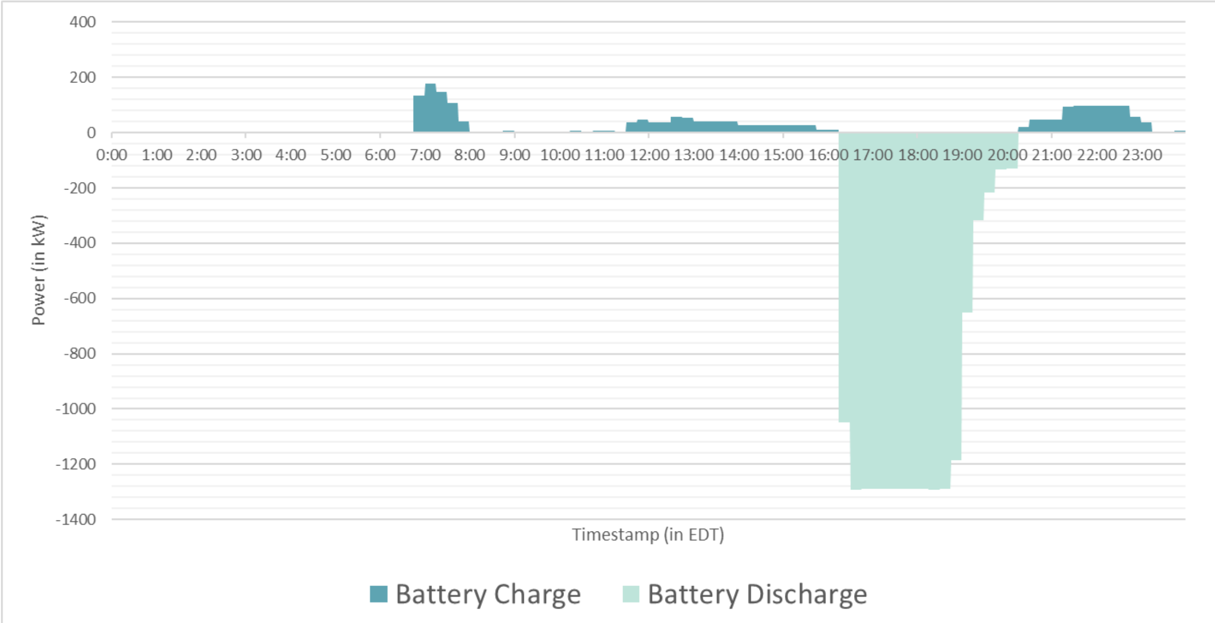
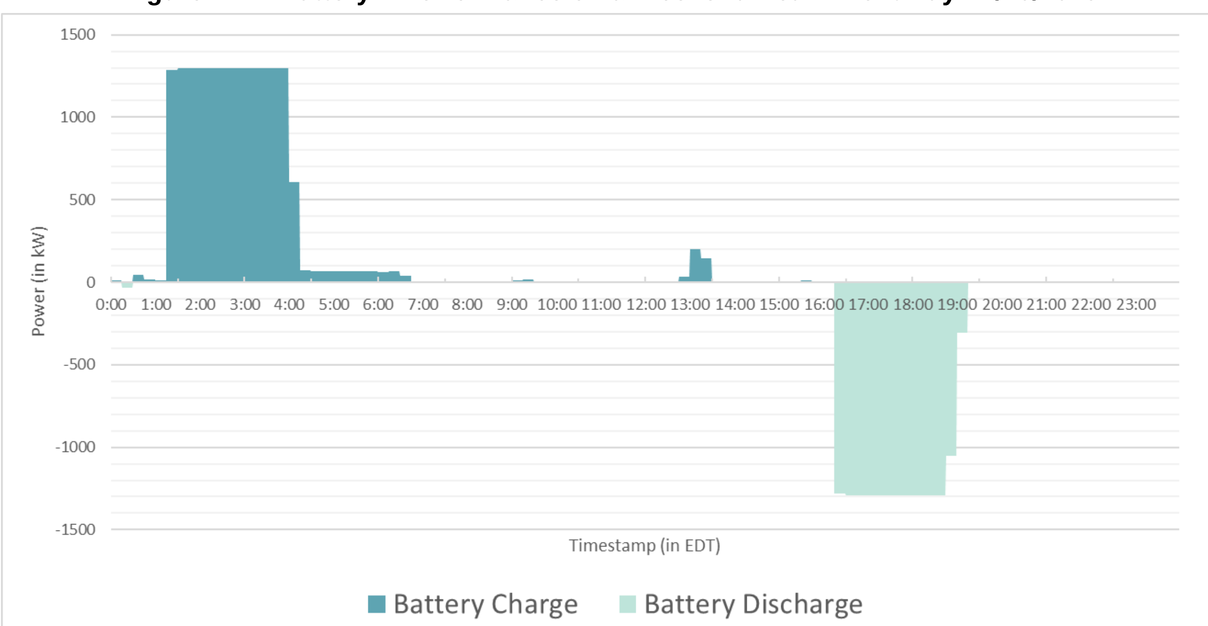


Figure 4-24 shows the battery performance on a weekend peak day which occurred on 7/20/2019. The battery was successfully dispatched during the weekend event hours.

Figure 4-24. Battery 2 Performance on a Weekend Peak Event Day - 7/20/2019



The performance of the two batteries resulted in a realization rate of 1.01, meaning that the evaluated performance was 1% higher than the reported performance.

4.6 Integrated Impact and Process Evaluation Findings

Challenges to Reliability: This section discusses whether the ADR Initiative and prior demonstration projects provide substantial evidence that it is reasonable to expect the program to meet its load reduction targets. We acknowledge that variability across a number of dimensions may raise doubts about the reliability of DR as a resource. Understanding the different sources of variability helps dispel these doubts and supports the claim that the initiatives can provide a reliable DR resource.

The results included in this report exhibit variability across a number of aspects of the programs. The primary evaluation baselines (Evaluated – Symmetric, Evaluated – Forecast, and Evaluated – Regression) vary amongst themselves and across PAs for the 2019 ICAP day event. For Eversource, the one PA with multiple events during summer 2019, results vary by event day. In addition, there has been variation relative to the demonstration projects that came before the current ADR initiative.

There are different kinds of variability at play in a load reduction estimate. Some degree of variability is inherent in the underlying load reduction process of DR programs. Load reduction requires special actions during limited timeframes – the load to be reduced needs to be active and the means to reduce it needs to be successfully applied. As a result of this, the amount of actual, underlying load reduction will naturally vary. Variation in weather contributes to load reduction variability, though in the context of a peak-targeting programs, load reduction tends to be higher on the more extreme days when system peak events are most likely to take place.[47] In addition to the natural underlying variability of load reduction, the act of quantification of load reduction adds another source of variability. Due to the variability of site-level loads, caused by weather sensitivity, processes with variable demand and/or strategic response to the DR program itself, the estimation of load reduction will have its own evaluation-related variability.

All of these recognized sources of variability could lead to a basic question of the reliability of DR as a resource. In this context, it is first essential to identify the standard by which reliability should be assessed and then understand what determines reliability therein. The primary stated goal of the initiative is reducing capacity and capacity-related costs for all customers through the reduction of demand at the system peak hour. Given that the ADR initiative is implemented through CSPs that have their own independent contracts with the individual resources and

[47] The 2018 heatwave that occurred during the week of July 4th is the exception that proves the rule. Despite substantially lower participant load levels due to shutdowns, system load was still high enough to trip ICAP day algorithms. Overall participants contribution to peak was substantially lower on those days despite a lower level of calculated load reduction. Those high load days were likely driven by high residential HVAC load with residential customers at home on vacation.

control the curtailment process, the primary standard against which variability should be assessed has to be program level load reduction targets.[48] Currently, this is the primary basis of load reduction expectations and basis for expectations of, for example, system operators.

There are four important dimensions of reliability identified as meeting load reduction targets:

1) **Can the CSPs recruit and manage sufficient load curtailment resources?** CSPs sign up for contractual targets well in advance of the program implementation. As a result, meeting contractual targets is necessarily a longer-term basis for assessing reliability. It is important, however, because it is the only pre-event prediction of load reduction that the evaluation team is aware of. CSPs could likely provide estimates of load reduction a day ahead of any event that are substantially more accurate than their long term contractual targets (given natural variability in weather, expected shutdowns, etc) and this could be a useful basis for assessing reliability in the dynamic context of peak load events.

2) **Is the system peak day successfully identified?** For summers 2018 and 2019, load forecasting algorithms successfully identified the ICAP day, leading to program implementation on each day. In 2017, the peak day came early in the season and some operators missed it. Forecasts for identifying the ICAP day in advance are probabilistic and only as good the historical data and the applied forecasting algorithms. The algorithms could miss the ICAP day or produce so many false positives that the limit on event days per season has been met by the time the actual ICAP event is identified. Large exogenous shocks to the system (e.g., Coronavirus) will complicate the forecasting challenge considerably. This aspect of the DR program reliability is difficult to assess even though correctly identifying the ICAP day is essential to meeting the primary goals of the program.

3) **Is the true underlying customer performance occurring?** This process is out of the hands of the PAs and evaluators. The CSPs take this into consideration in their recruitment, determination of targets/enrolled capacity and implementation strategy. The CSPs have every economic interest to make sure that they can deliver the enrolled capacity. The CSPs have limited control over shutdown days (e.g. around July 4th). If an event is called and the expected hot weather does not materialize, load reduction may be lower than expected. These are all marginal effects, though, compared to missing the

[48] This is in contrast to enrolled capacity, for instance. It is clear that CSPs enrolled substantially more capacity that targeted so as to meet their commitments. The implicit de-rating process is part of the CSP's implementation process and not an issue with respect to assessing reliability of the program. To the contrary, it is one of the ways the CSPs minimize the risk of not meeting program targets.

ICAP day. This kind of variability also supports the idea of event day-specific forecasts of load reduction as an alternative basis for assessing reliability.

4) **Is there a consensus way to measure load reduction?** Part of the perceived variability of load reduction estimates is a function of the array of baselines that are used to quantify the load reduction. Quantifying load reduction is a prime example of George Box's aphorism, "all models are wrong, but some are useful". This report includes discussion regarding how each baseline is useful in different ways for estimating load reduction. Focusing on a single baseline or at least just one of them at a time has the advantage of eliminating a major source of variability, one due to measurement methods. The remaining variability is due to actual differences in load reduction between participants, PAs, states and days. Variability in load reduction estimates across baselines for a particular event reflect the underlying definition of the counterfactual.

Taking into consideration these four dimensions, the ADR initiative and prior demonstration projects provide substantial evidence that it is reasonable to expect the program to meet its load reduction targets at the PA level. There is variability, across states and, for Eversource, across events, and confidence would be greater if the evaluation was based on more 2019 events, but the CSPs appear to have the flexibility and visibility into the situation such that they can meet targets, even those set a long time ago. Over the last two years they have successfully recruited customers, identified the ICAP day and met overarching targets for load reduction by PA.

The initiative is not currently organized to facilitate better estimates of load reduction capabilities across a range of conditions or on a geo-targeted basis. However, with minor changes to operating procedures, effectively a re-emphasis of these secondary goals of the initiative, this further grounding of reliability would be possible. For example, the initiative could increase the number of events called over the summer and test geo-targeted subsets of customers.

The CSPs and PAs currently demonstrate their ability to meet the targets based on the asymmetric baseline, which is the agreed upon, but biased, measure of load reduction. We discuss the balance between program rules that promote ease of program development and those that support the most clear, reliable and unbiased load reduction estimates. If the initiative continues to use the asymmetric baseline, then the realization rate reflecting the ratio between asymmetric and symmetric baselines will be important inputs for reliability considerations at the applied system operations level.

To date, the ADR initiative has demonstrated that resources can be activated and an accurate estimate of load reduction can be produced despite ongoing challenges with interval data systems. The PAs could easily produce symmetric baseline estimates of load reduction for the purposes of system operators and, frankly, future contracts with CSPs should be based on this metric even if customer settlements continue to be based on the asymmetric baseline.

The targeted nature of the evaluation, to date, means that a number of important issues have not been pursued to their full extent.

- The incidence of only one summer event for Unitil and National Grid and three for Eversource added marked variability to the 2019 evaluation and tempers evaluators' ability to comment on DR capabilities and confidence in reliability of DR resources in the future. If the reliability of the initiative continues to be a concern, the PAs should consider putting more emphasis on developing their DR capabilities and calling more events in support of that effort. If the next season is likely to have as few events as 2019, this would increase program cost and customer disruption. It would aid impact evaluation, but the more important benefit of doing so is that the program would generate more data to assess the ability of CSPs to accurately identify the available load reduction across a range of conditions and improve confidence in reliability for future years.
- Evaluators can provide the PAs cross-CSP, geographically targeted analysis of load reduction capabilities. Development of these capabilities are a goal of the program and have not been fully explored.
- Evaluators can explore the interaction with ISO-NE in much greater depth. We can increase our understanding of how resources take part in the ISO-NE market, developing tools to identify those activities based on load data and ISO-NE energy market prices alone. These efforts should make it possible to quantify the degree to which overlap occurs and the implications for both PAs and ISO-NE.
- Evaluators could fully explore the calculation of benefits proposed to justify the ADR initiative.

4.6.1 Shutdown Days

Customers or CSPs are failing to report shutdown days to the utilities.

All three PAs (Eversource, National Grid, and Unitil) included a shutdown day allowance in their ADR initiative participation rules. This rule allows customers to not participate in a certain number of events because of facility shutdown days.[49] The days are also subsequently dropped from their baseline. For a shutdown day to be counted, that customer must notify their PA at

[49] The MA initiative rules state there is a limit of 10 shutdown days per season. Eversource's evaluation team has said that the same rule applies in CT. In the NH order, pilot rules state that a customer cannot have more than 14 reported shutdown days in a season.

least a week in advance of the shutdown. The customer can report a scheduled shutdown to their CSP or communicate it directly to the PA.

National Grid staff noted that the shutdown day allowance was included in the ADR initiative rules so that customers would be assured that if there was a need to shut down a facility, it would not impact performance calculations or their baseline. The shutdown day rule could save customers from an event performance that is lower than expected if the event was called on a shutdown day. Also, as noted by Eversource PA staff, shutdown days could have an impact on customer payments if they fell within their baseline and went unreported.

4.7 Per interviewed PA staff, neither customers nor CSPs reported any shutdown days in the 2019 summer season. National Grid staff also noted that, to date, there has not been a single shutdown day reported during the initiative's three-year lifespan. However, one surveyed participant reported that their "plant was shut down on potential peak demand day(s) prior to the actual event." In addition, the analysis of the meter data revealed a number of data points that indicate facilities were likely shut down in the baseline period. The methodology used to identify shutdown days is described in the Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

4.7.1 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned

- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs
- Other opportunities for peak demand management

4.7.2 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology (Baselines) as part of the **Evaluated-Forecast** result.

Responses from the CSPs suggest that this lack of reporting is a function of many factors, including:

- **Only one interviewed CSP has a protocol for communicating with customers throughout the summer DR season about shutdown days.** This CSP first discusses anticipated shutdown days with customers during onboarding. They also proactively remind customers to report potential shutdown days, especially for days on which shutdowns are more common (e.g., directly before or after a holiday). For example, this CSP calls certain customers prior to Thanksgiving to say that similar businesses are usually closed on the Friday after Thanksgiving and to ask whether the customer would like them to report that day as a scheduled shutdown.

  One other CSP reported asking customers at the beginning of the DR season to alert them of any shutdowns they had planned for the summer. The remaining interviewed CSPs said that they did not have any customers who scheduled or reported any shutdown days over the summer. One explained that, if they had, the CSP would have reported those days during the settlement process after the season ended.

- **Customers forget to report and/or have unexpected shutdown days.** The CSP that has a protocol for communicating regularly with customers about shutdowns noted that even with their approach, they were unable to capture all shutdown days. Sometimes customers forgot to report a shutdown day in advance, and sometimes there were shutdowns that the customers did not anticipate.

- **Reported shutdown events did not fall within any baseline periods.** The CSP that instructs customers at the beginning of the season to let them know of planned facility shutdowns said that dozens of customers had reported a shutdown event over the summer; however, none of those reported events fell within the baseline periods. As such, the CSP did not report those shutdown days to the PAs.

Alternatively, lack of reporting may be due to the use of an asymmetric adjustment in the settlement baseline. An asymmetric adjustment decreases the risk, to the customer and the CSP, of not reporting a shutdown. That is, a customer may be credited with substantial load reduction regardless of shutdowns during the baseline period or event day. Requiring and incentivizing advance notice (or conversely, penalizing unreported planned shutdowns) could support this end, but it would be difficult to enforce. In contrast, the symmetrically adjusted baseline shifts the risk to the customer, making it more likely that they will be compliant.

Collectively, these findings suggest that CSPs are not appropriately interpreting the shutdown day allowance rule. For a shutdown day to be excluded from an event or baseline calculation days, the customer or CSP must report that shutdown date to the PA seven days in advance of its occurrence. It will not be known whether a customer's planned shutdown will fall on an event day or within their baseline period until an event is called either on that day or within ten eligible days after. If a CSP reports only customer shutdown days that fall within a baseline period or waits until the season ends to report shutdowns, the CSP is not giving the PA the notice currently required per the ADR initiative rules to account for the shutdown. CSPs should report *all* customer-announced shutdown days to the PA seven days prior to the day. The customers' performance will not be negatively affected, and the day will be excluded from the baseline, when appropriate.

4.7.3 Evidence of Pre-Cooling, Gaming, and Snapback

This section substantiates the evaluation's The performance of the two batteries resulted in a realization rate of 1.01, meaning that the evaluated performance was 1% higher than the reported performance.

Integrated Process and Impact Evaluation Findings described in the Executive Summary. As described above, in the Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

4.7.4 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned

- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs
- Other opportunities for peak demand management

4.7.5 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology (Baselines), the baseline adjustments rules create opportunities for savvy customers and/or CSPs to inflate customer baselines by artificially increasing site load in the hours leading up to a DR event, resulting in a more favorable adjustment and higher payment. This would be considered "gaming" because customers/CSPs would be using this rule not in the way it was intended but to benefit them in a way that doesn't accurately reflect the actual demand reduction provided. The challenge with detecting gaming is that it is impossible to know the intent of the customer. Pre-cooling is a valid form of preparing for an event by shifting some cooling load early. Similarly, production could be ramped up in advance of an event in the hopes of minimizing production shortfalls. The load signature of these legitimate pre-event preparations and that of gaming are practically indistinguishable from one another. Both the impact and process evaluation teams set out to explore the issue of participant gaming.

Through the customer surveys, the evaluation team asked customers whether building operations adjustments were made in the hours leading up to or in the hours after DR events, or both.[50] If customers indicated adjustments were made, they were asked to describe the adjustments. Altogether, a quarter of participants who responded to the survey (15 out of 58) noted taking action before the event (six participants reported making adjustments prior to DR events, while nine reported[51] making adjustments both before and after DR events). Of note, most respondents said the adjustments they made prior to DR events involved either shutting down equipment or preparing systems to cycle down. One respondent stated they "shutdown as early as possible," and another remarked, "We won't turn on some equipment if we know that it turns off slowly, and non-essential equipment is left off or brought down."

[50] Though actions taken in the hours following a DR event would not have an impact on customer performance, the evaluation team asked customers about post-event building adjustments in an attempt to "disguise" the question about pre-event adjustments and give it an innocuous undertone.
[51] There are indications that some respondents might have misinterpreted this question. Some respondents indicated that they made building adjustments both before and after DR events, but when describing the adjustments made, seemed to list their curtailment activity right at the start and end of the events. For example, one participant wrote, "Shut down refrigeration and turn it back on." Another noted, "Turn off/on escalators etc. and adjust/return BMS schedules to previous/programmed schedules."

Table 4-16. Customer Responses: When Did Curtailment Occur?

Under your plan with your CSP, were there [manual/automatic] adjustments that you or others made on DR event days either BEFORE the actual DR event began or AFTER the actual DR event finished?	
Manual Adjustments	**n=35**
Yes – Made in the hours **prior** to when the DR event(s) began	5
Yes – Made in the hours **both prior to and after** the DR event(s)	8
Yes – Made in the hours **after** the DR event(s)	1
No	17
Don't know	4
Automatic Adjustments	**n=13**
Yes – Made in the hours **prior** to when the DR event(s) began	1
Yes – Made in the hours **both prior to and after** the DR event(s)	1
Yes – Made in the hours **after** the DR event(s)	-
No	10
Don't know	1

These results should be interpreted and quantified with caution since a minority (less than half) of participating customers completed the customer survey. Additionally, if participants knowingly engaged in gaming, it is unlikely they would admit to it in the customer survey. It is still notable that none of the respondents discussed any actions that could have increased consumption in the pre-event hours, including reasonable responses related to pre-event pre-cooling or production shifts intended to reduce the business impact of load reduction.

The impact evaluation investigated whether there was evidence of pre-event load increases that could be explained either by pre-cooling, load shifting, or gaming. Post-event, the impact evaluation investigated whether there was evidence of post-event load increases that could be explained by snapback.

Figure 4-25 shows event day load (orange line), the unadjusted and symmetrically adjusted 10-of-10 baselines (dark blue and light blue lines, respectively), and the regression baseline (green line) for all of Eversource's Massachusetts participants, on the July 30, 2019, event. The adjustment hour and event period are shaded gray. Note, that the July 30 event captured the ISO-NE ICAP hour, which occurred from 5 to 6 p.m. and is shown in dark gray.

Figure 4-25. Eversource Massachusetts – July 30, 2019, Event

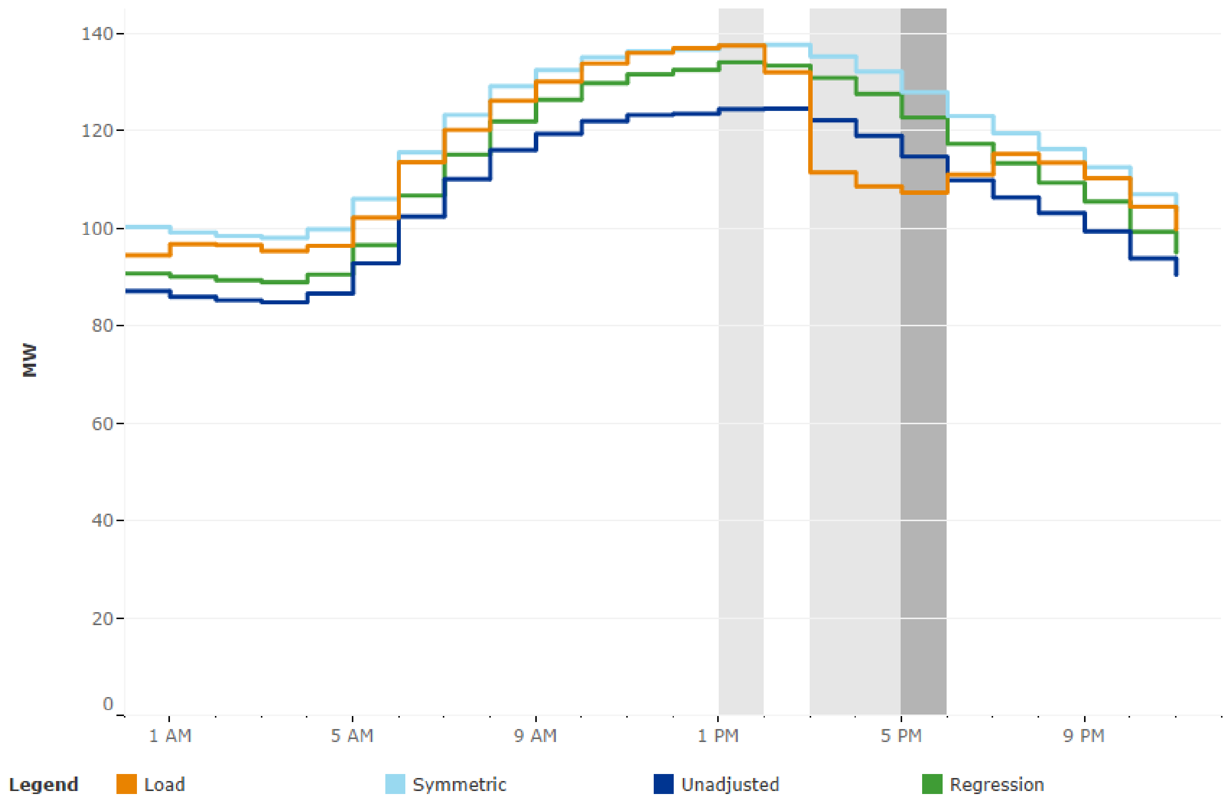


Figure 4-25 allows us to compare the actual load in the pre-event period with the 10-of-10 and the regression load shapes. The relationship between the actual load and 10 of 10 shapes does not reflect a typical gaming scenario. The symmetrically adjusted baseline is adjusted to pre-event load and is higher than actual load across the remaining pre-event hours. The difference between these two shapes is greater in the morning hours but decreases for many hours prior to the adjustment period.[52] This is the signature of a baseline shape that has too flat a shape for the event day load rather than gaming. In contrast, gaming is expected to increase load just around the adjustment hour, lifting the entire baseline a similar amount across all earlier hours. It is common for the adjusted baseline to reflect a flatter shape from the more moderate baseline days compared to the more extreme system peak day.

The shape of the regression baseline relative to actual load further supports this hypothesis. The regression baseline explicitly controls for weather across all hours. It demonstrates a greater ramp from minimum to maximum load across the day than does actual load. Given that the

[52] Pre-cooling and, particularly, gaming are expected to only take place in the last couple hours prior to an event. The goal of the former is to get the internal temperature a few degrees cooler; the goal of the latter is to artificially increase load just during the adjustment hour. In neither case is there motivation for these pre-event behaviors to start the night before and affect early event day load.

regression estimates hourly load relative to the unique weather of the event day, we could expect the regression baseline to provide a more appropriate overall shape than the 10 of 10. Furthermore, to the extent the regression does not perform as well on extreme days like this, we would expect it to underestimate the magnitude of the overall ramp of the day. That the regression has a greater ramp than the 10 of 10, despite this likely underestimation, is further evidence that the 10 of 10 is too flat for the true counterfactual load on the event day.

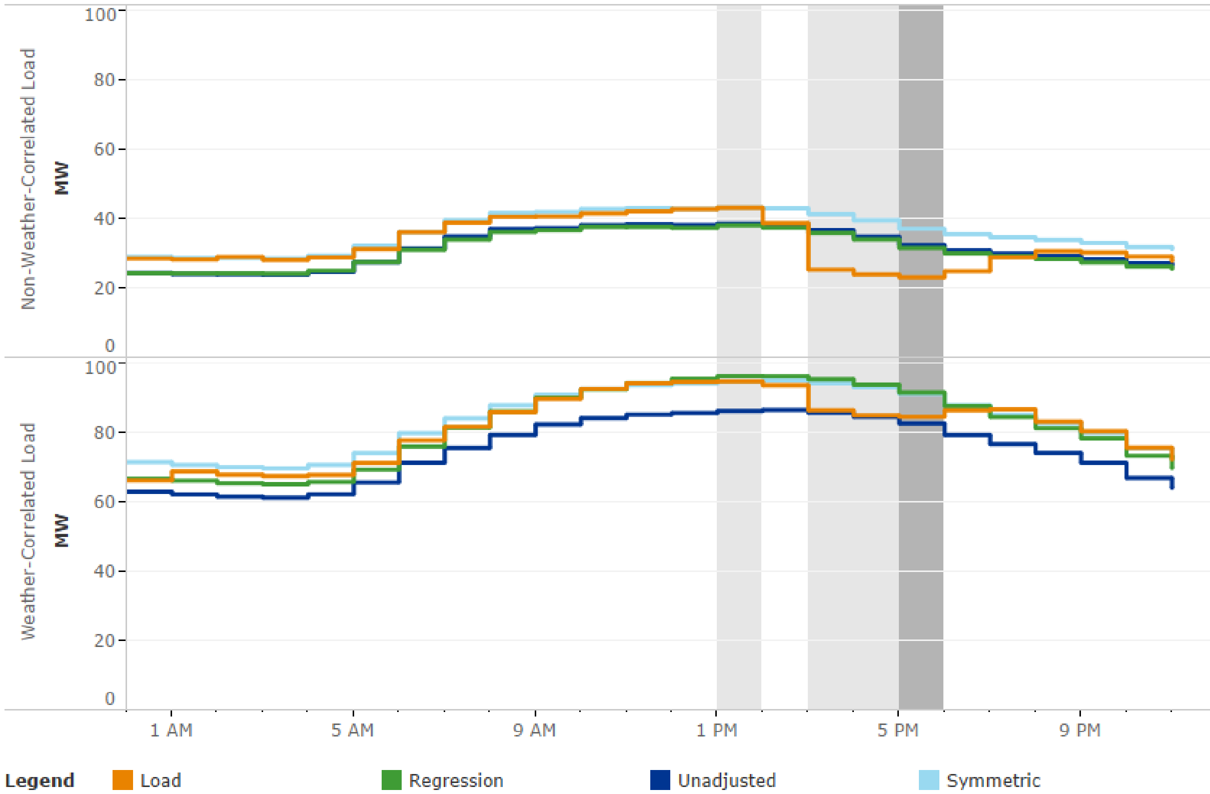
The post-event period can also be a period of increased load due to the event, frequently called snapback. Snapback is effectively a shift of load as opposed to shedding of load. HVAC-based DR is frequently a shifted load causing snapback, though for events late in the day, higher post-business hour setpoints might make the snapback effect much smaller. The 10-of-10 adjusted baseline will frequently not show snapback due to its too-flat shape. In Figure 4-25, the actual load rises back above the regression baseline in the post-event period indicating that there may actually be some snapback.

Figure 4-26 provides the same plot for the Eversource MA population split into customers who have well-behaved regressions with evidence of weather-correlated load[53] and those who do not. These plots further illustrate the discussion about pre-event load increases (pre-cooling or gaming) and snapback. In the second panel, the regression load shape is almost identical to the adjusted 10-of-10 during the adjustment and event hours. In the early part of the day, the 10 of 10 is higher than actual load while the regression baseline is lower. The regression with its hourly adjustment for weather illustrates how the 10 of 10 can carry too flat a shape from the baseline days. For these weather-correlated customers, the regression is a surprisingly good estimate of actual load across the day, suggesting no pre-cooling/gaming and modest snapback.

The non-weather-correlated plot shows a regression that is almost identical to the unadjusted 10-of-10. This is to be expected in non-weather-correlated loads as both baselines are effectively mean processes. The regression baseline produces a much lower estimate of load reduction and, because it is so low, show a small amount of snapback. When adjusted, the 10-of-10 is almost identical to actual load in all pre-event hours. As with the overall plot above, this adjusted 10 of 10 shape is not indicative of any kind of pre-event activity. It also does not show evidence of snapback. These baselines indicate that the non-weather-sensitive baselines happened to have higher load across the whole event day due to non-weather-related factors.

[53] A customer with a well-behaved regression and evidence of weather-correlated load is defined as having an adjusted r-squared greater than or equal to 0.90 with an average p-value less than 0.05 for hourly CDD coefficients for the hours ending 13 thru 18.

Figure 4-26. Eversource Massachusetts – July 30, 2019, Event – Weather Correlation



Figures 4-27 and 4-28 lead to similar conclusions for Eversource NH. Figure 4-27, a plot of the whole population, appears to exhibit pre-event increases in load. However, when split out, the plots show that almost all of the pre-event period load increase is motivated by non-weather-correlated loads and is consistent from 5 AM through the adjustment hour. This doesn't fit the profile of an artificial load increase. For the weather-correlated group, the regression baseline does a much better job of matching the actual load outside of the event period. By contrast, the 10 of 10 is too flat, consistently underestimating load outside of event hours.

Note, Figures 4-25 and 4-26, above, and Figures 4-27, and 4-28, below, are examples. Plots of the other PA states support similar conclusions and can be found in Appendix D: Weather Correlation Plots.

Figure 4-27. Eversource New Hampshire – July 30, 2019, Event

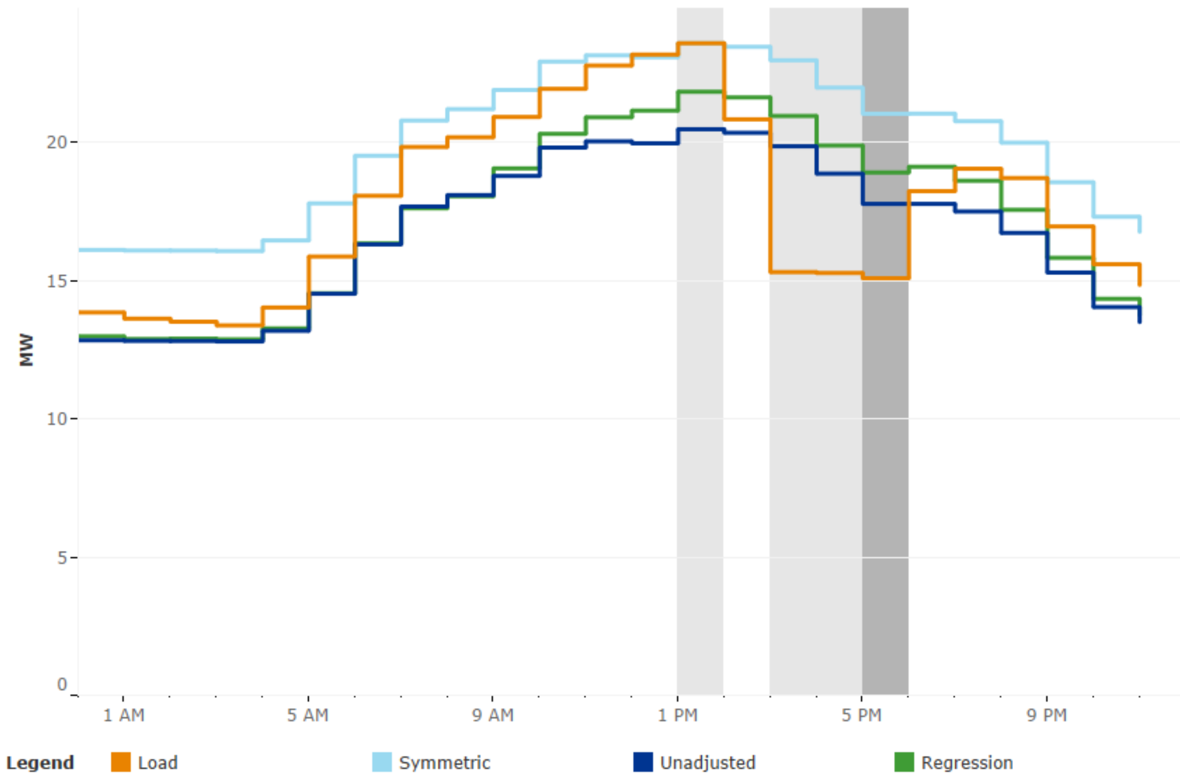
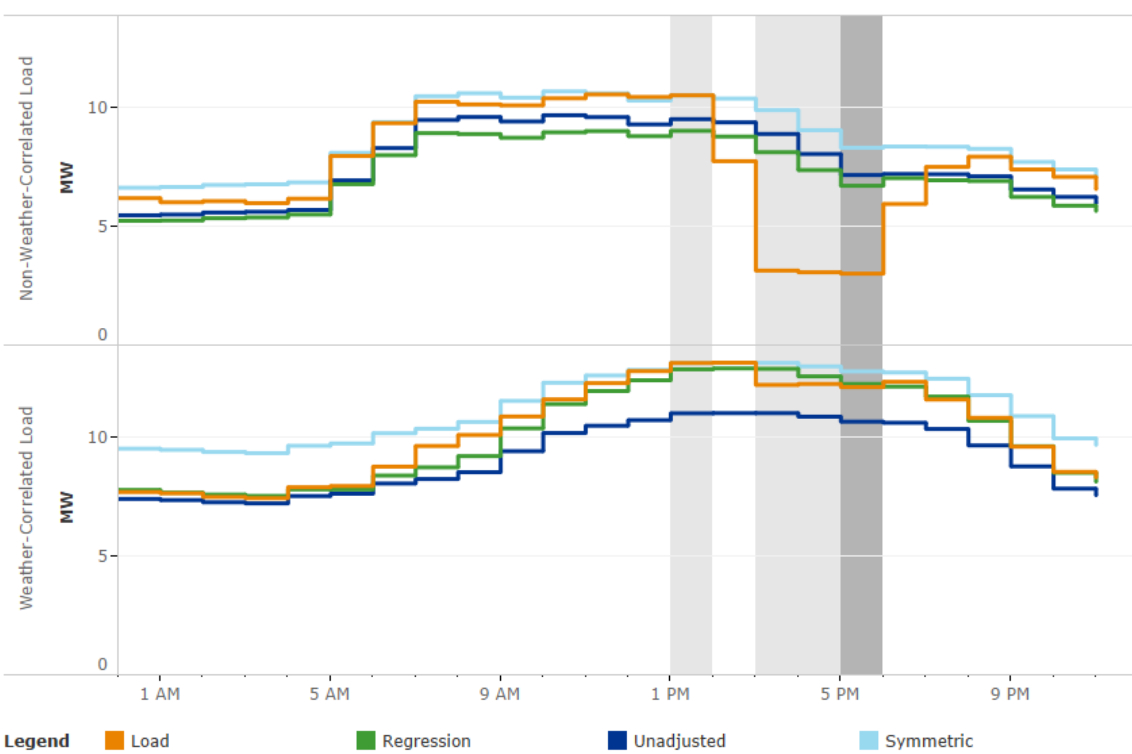


Figure 4-28. Eversource New Hampshire – July 30, 2019, Event – Weather Correlation



As mentioned previously, DR initiatives with day-ahead notification are more vulnerable to upwardly biased load reduction estimates due to either strategic activity or gaming. The advanced notice allows a customer to either prepare for the event by pre-cooling or increasing production in the pre-event hours, or, more nefariously, target the adjustment hour for artificial load increases that will, in turn, increase the baseline. One reason for considering a regression baseline is that it is not reliant on the same-day adjustment. In addition to being an alternative baseline, the regression baseline is also the most obvious basis on which to assess customers for gaming. An unexpectedly large difference between the regression and actual load during only limited pre-event periods could be an indication of gaming.

Figures 4-29 through 4-31 provide an illustration of regression errors compared to the unadjusted 10-of-10 errors during the adjustment period (1 to 2 p.m.) across the summer of 2019 for each PA. The plotted values represent the aggregate adjustment (or error) for both the regression and unadjusted 10-of-10 baseline during the adjustment hour. The values are baseline minus load. The positive values are situations when the baselines were higher than load; the negative values are situations when the baselines were lower than load. In these plots, gaming would manifest as an unexpectedly large negative value on the event day. The event days (July 30 for all PAs, with additional events on July 19 and August 19 for Eversource) are marked in red.

The figures' plots show the extent and variability of the errors across the summer. The majority of errors are in a band of +/- 10% and, in at least some places, appear to have an underlying wave-like structure. Highpoints around the week of July 4 coincide with a higher concentration of shutdowns that week. Low points frequently coincide with periods of hotter weather.

The plots also show that the regression, if it can successfully control for a substantial amount of variability, should have lower errors (be closer to zero) than the unadjusted 10-of-10 errors in most cases.[54] This is the case for the majority of days, for all three PAs.

Finally, these plots highlight the challenge of identifying gaming amongst natural variations in load. There is no evidence that the regression errors on event days are of a greater negative magnitude than expected. To the contrary, none of the events represent extreme errors in general, and in the non-event days immediately around the event days there are many negative errors with greater magnitudes. This evidence points to the conclusion that at the aggregate level there is little evidence of gaming and that any gaming that might be present is much smaller in magnitude than typical adjustments necessary to align the 10-of-10 with actual load.

[54] To counter this, the 10-of-10 has the advantage of a shorter and more recent set of data. To the extent that customers are not specifically weather-correlated, the 10-of-10 may have an advantage.

Figure 4-29. Eversource - Adjustment Hour Baseline Errors

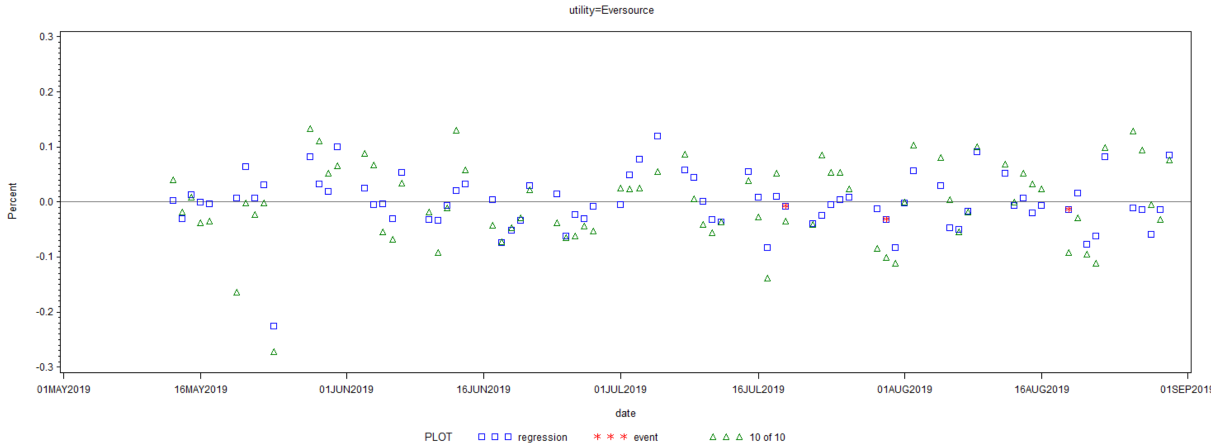


Figure 4-30. National Grid - Adjustment Hour Baseline Errors

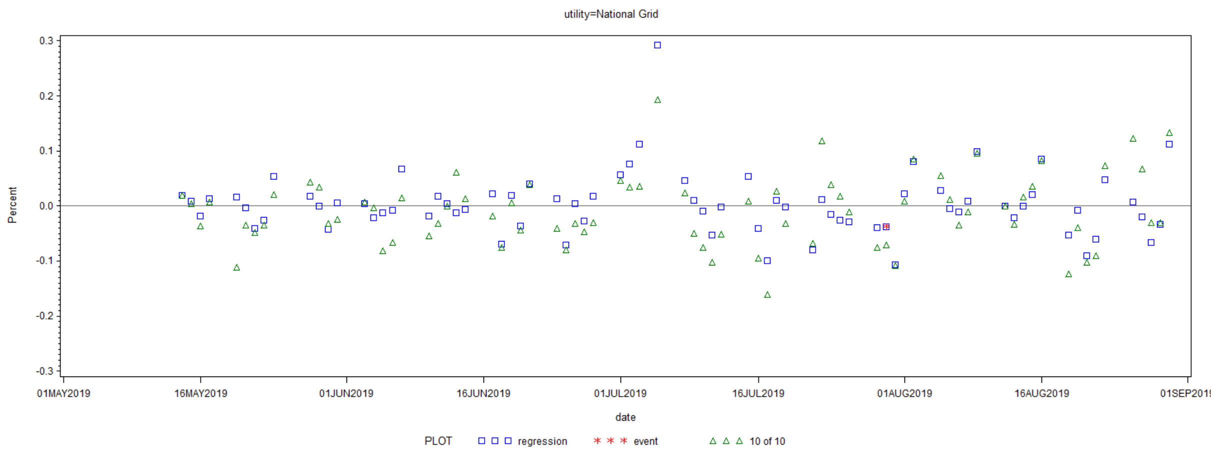
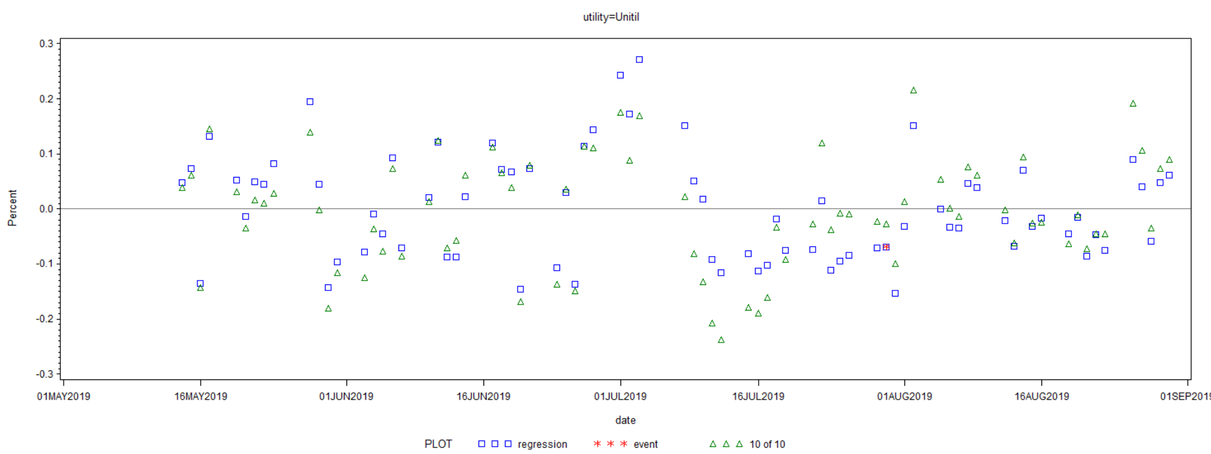


Figure 4-31. Unitil - Adjustment Hour Baseline Errors



5 Conclusions, Considerations, & Recommendations

Based on the process evaluation interviews and surveys, impact evaluation data collection and analyses, and document review, the evaluation team generated the following conclusions, recommendations, and considerations:

- **Conclusion 1: Requiring participants to sign the PA ADR application after they have signed the agreement with a CSP slows down and complicates the sales process.** Several CSPs noted that asking a customer to sign the additional PA contract hinders the recruitment process. Customers questioned why they needed to sign another agreement when they had already signed an agreement with the CSP. One CSP observed several large prospects refusing to sign the PA's contract after they signed the agreement with the CSP. CSPs also said they are unclear about the purpose of the PA contract for those customers who participate through a CSP, since the customer has already provided the same information and agreed to a contract with the CSP.

  Consideration 1-1: Explore consolidating PA and CSP terms and conditions into a single document or at least presenting the two agreements at the same time during recruiting. National Grid stated that it is standard practice for CSPs to present customers with both the CSP contract and PA application at the same time, but this is not what was reported by CSPs. National Grid staff should check that CSPs are adhering to this practice.

- **Conclusion 2: Feedback from nearly all sources suggests that settlement and payment is a significant administration challenge.** ADR initiative materials (e.g., flyers on National Grid's website, Eversource's Active Demand Reduction application) indicate that payments for the DR summer season will be made by the fall. At least one PA understood that CSPs had communicated the likelihood of a later payment period to participants, but this did not come up in CSP or participant interviews. By December 2019, a minority (less than one-third) of surveyed participants across the three PA initiatives reported that they had received event payments for the 2019 summer season. For Eversource, the settlement calculation and payment processes were difficult due to the interval meters failing to record or transmit information and the decentralized tracking systems. National Grid staff stated that revised procurement requirements created payment delays, and they also reported that a small number (less than 5%) of payments were delayed due to metering and data issues. When asked if they were satisfied with the event payments, the majority of participants (61%) in the Eversource sample and a notable proportion (31%) in the National Grid sample were uncertain (stated "don't know") since many of them noted that they had not yet received payment. The National Grid effort was in operation with evaluation cycles for two years before this current study, which likely explains why their respondents were more familiar with payment processes. It is also important to note that the survey results did not indicate that the participants were dissatisfied, just that they were not yet satisfied.

Recommendation 2-1: Continue to seek solutions to accelerate the incentive payment process. Eversource is building a centralized program tracking and management system, which will pull customer interval data directly from the CSPs in close to real time. If the system performs as expected, it will shorten the time it takes the PA to complete the settlement calculation and issue payment. National Grid is starting to allow CSPs to access their online day-after data and daily performance summaries. This access should help CSPs more quickly identify faulty meters or reconcile data discrepancies, which affect the payment turnaround time.

Consideration 2-1: Monitor participant satisfaction with event payments by periodically surveying customers after payments are sent out. Consider adjusting initiative materials to more strongly convey the likely payment schedule.

- **Conclusion 3: Fear of facility disruptions, even when minor, is a barrier to participation.** This barrier may be more prominent for the manufacturing facilities than any other customer types. Manufacturing facilities were the majority of Eversource and National Grid participants who experienced slight or temporary disruptions when the events were called and thus reported performing partially or not at all during the DR event(s). CSPs explained that industrial facilities will not always be able to participate due to their inability to decrease production when an event is called.

- **Conclusion 4: All the participants across all three PAs who responded to the surveys stated that their opinion of the PAs was either positively impacted or unaffected by the ADR initiatives.** The initiative is improving the participants' opinions of their PAs. This is despite the fact that, at the time of questioning, many participants had not received their payment for curtailment, which can heavily influence levels of satisfaction with overall engagement.

- **Conclusion 5: The ADR initiative is not consistently branded across the three PAs' marketing materials and initiative administration documentation.** For example, National Grid, Eversource, and Unitil use the "ConnectedSolutions" brand to refer to this initiative in MA, but Unitil does not use the "ConnectedSolutions" brand in NH. Additionally, the Eversource website refers to the initiative as ConnectedSolutions, but the Eversource DR initiative application is titled the "Active Demand Reduction" application. This can create confusion in the marketplace.

- **Conclusion 6: Customers and CSPs are failing to report shutdown days to the PAs.** Initiative rules include a limited number of announced shutdown days, communicated in advance. The benefit to participants and PAs is that the shutdown day can and should be excluded if it occurs during the baseline period, to avoid an inappropriate reduction in baseline load and claimable load reduction. Disclosing it in advance of what turns out to be an event day is beneficial as well, as it may prevent the penalty of no/low-DR performance and helps PAs and CSPs manage event day expectations. Not knowing shutdown days also interferes with evaluators' effectiveness in using regression-based baselines, even if they occur outside of the baseline periods that affect settlement. Per interviewed PA staff, neither customers nor CSPs reported any shutdown days in the 2019 summer season. Yet one surveyed participant reported their facility was closed prior to the actual DR event, two CSPs had customers report facility shutdown events, and interval meter data showed evidence of shutdowns. CSPs noted three reasons for not reporting shutdown events to the PA: 1) facility shutdown events were unexpected, 2) customers forgot to report it, or 3) shutdown events do not fall within any baseline periods.

- **Recommendation 6-1:** Remind and educate the CSPs of the shutdown allowance and reporting rule. The PAs could ask for pre-planned shutdown information during the application/enrollment process.
- **Recommendation 6-2:** Adapt the shutdown rule to account for unexpected facility shutdown events. To exclude a facility shutdown day from a customer's baseline calculation, that customer or their CSP must notify their PA at least seven days in advance of the shutdown. It is difficult to do this when a facility shutdown event is unexpected. Consider allowing customers or CSPs to report the shutdown to the PA 24 hours before an event is called.

- **Conclusion 7: Data quality and analysis requirements are not coordinated or consistent across the PAs.** There is little evidence of coordination across PAs with regards to data sufficiency rules and other load reduction calculation details. Each PA's approach would produce different load reduction estimates if applied to the same group of assets. Within a given state (e.g., Massachusetts), load reduction should be calculated in a consistent fashion across administrators. Inconsistencies of this nature jeopardizes the ability of the PAs to aggregate their estimates at a state level with confidence.

  - **Recommendation 7-1:** Formally standardize all rules related to data quality, baseline calculation methods, and aggregation.
  - **Recommendation 7-2:** Establish data quality rules with clear outcomes for poor quality and/or insufficient data. The evaluation team developed several rules as part of this study, which are described in the Data Sufficiency section of this report; these rules may be a useful starting point to develop consistent rules. The data issues

encountered in this study were not anticipated. Establishing firmer expectations or providing incentives are two possible means of motivating and ensuring clean and complete data in future initiative cycles.

- **Conclusion 8: The quality of the settlement baseline process can be improved.** A demand response offering is only as good as its settlement baseline process. The PAs should consider changing the initiative to a baseline that is more in line with ISO-NE's baseline. The ISO's baseline has evolved over time to reflect the best baseline methods for attaining an accurate estimate of load reduction. Important aspects for consideration include a symmetric adjustment, no day-ahead notification, and meaningful reporting of shutdown days. While not fully examined, the current initiative baseline has theoretical limited downside risks for assets with highly variable load that do not intend to take load reduction actions but will receive payments when random load variation produces a positive load reduction relative to baseline. A change in notification from day-ahead to day-of would improve the quality of the settlement load reduction estimate by eliminating customers' ability to increase load during a known adjustment hour in anticipation of event hours to follow later that day, with the potentially significant trade-off to the initiatives of more expensive recruitment and/or likelihood of lower MW enrollment, due to customer preference or need for advance notice to participate.

  Consideration 8-1: The PAs should consider revising the program design and baseline to be more in line with ISO-NE. There are several options that could move rules closer to ISO-NE rules, including:

  - Pay incentives to participants based on the symmetrically adjusted 10 of 10 baselines instead of asymmetric. Incentives may need to increase correspondingly.[55]
  - Provide an additional incentive for customers to be dispatched without prior-day notification.
  - Increase payment for delivered kW, while adding a penalty or de-rating factor for partial delivery. Delivery below some benchmark (e.g., 80% of enrolled) would be paid on a sliding scale down to some benchmark with no payment below that level.

- **Conclusion 9: Evaluators found that the 10-of-10 baseline with symmetric adjustment was the most appropriate measure of impact for the 2019 summer season and is the basis for the retrospective realization rate. Evaluated forecasts can be utilized to estimate reductions for future summers.** In each result table, the evaluators provide

[55] The symmetric baseline will produce lower estimates of load reduction for assets that shutdown early for events. These assets may require additional payment to choose a symmetric baseline.

evaluation symmetric and forecast values that reflect our best estimate of the level of reduction achieved and that might be expected on a forward-looking basis given the use of a symmetrically adjusted baseline with an additional shutdown adjustment.

- **Recommendation 9-1:** Use the retrospective realization rate to determine past season performance.
- **Recommendation 9-2:** Use the prospective realization rate to estimate future load reduction.

- **Conclusion 10: Lack of coordination between ISO-NE markets and the PA's ADR initiatives could have unintended consequences.** The lack of coordination and communication between the PA DR and ISO-NE offerings could affect the ISO-NE baseline calculations and could result in the over-estimation of available reserves. Both ISO and PA staff expressed a willingness to discuss overlap concerns and solutions.
  - **Recommendation 10-1:** In the short-term, representatives from ISO-NE, the PAs, and, if feasible, the CSPs should come together at a Demand Resources Working Group (DRWG) meeting and brainstorm mutually beneficial design solutions that would minimize the impact of one entity on the other.
  - **Consideration 10-1:** The PAs could consider incorporating language into their rules that instruct the CSPs to either make themselves unavailable or modify their bids so as not to clear on the ISO platform during PA ADR events.
- **Conclusion 11: There is a natural tension between program design for program implementation and program design for the most comprehensive assessment of reliability.** The ADR initiative design minimizes customer burden while maximizing value to the customer by minimizing the number of events called, offering day-ahead notification, and using the asymmetric baseline. A program designed to produce the most reliable and consistent estimates of load reduction would call more events, forego day-ahead notifications, and use a symmetric baseline for the basis assessing contractual targets. If prioritized, these changes would likely increase customer burden while increasing program costs and/or decreasing the amount of available load reduction. The optimal combination of these program characteristics will flow from a clear recognition of reliability expectations for these programs.

Appendix A: State-Level Detailed Results

Differences in Delivery by State

Eversource and Unitil staff both noted that there were no differences in how they administer the initiative to customers located in different states. Eversource serves customers in Massachusetts, Connecticut, and New Hampshire. Unitil serves customers in Massachusetts and New Hampshire only. For National Grid, the evaluation team only assessed their Massachusetts initiative.

Participant Experience by State

The team also examined participant experience across the three states to assess whether any aspects of the initiative might be different or are perceived differently in different jurisdictions. Analysis revealed no differences in participant experience by state (see subsequent sections).

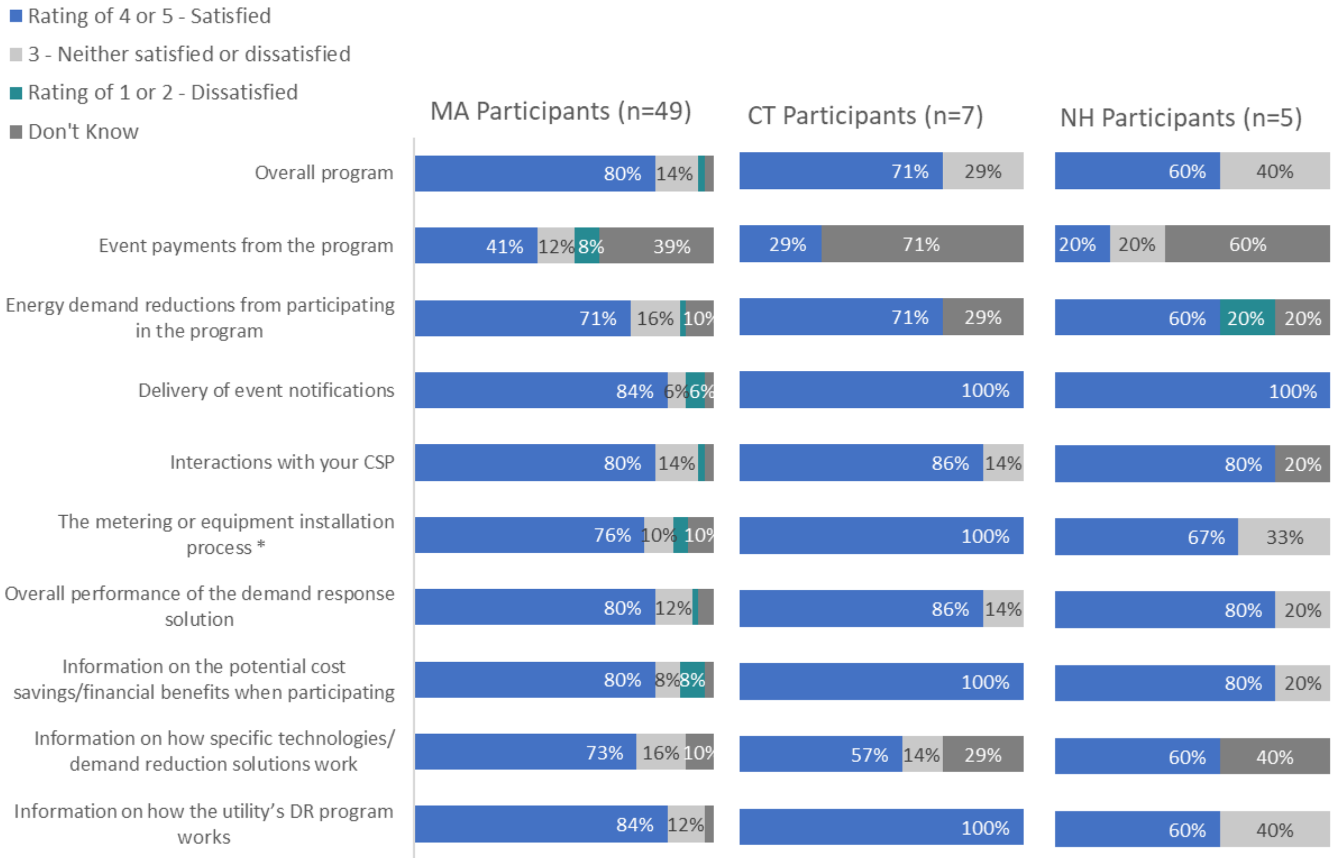
Note that participant populations in Connecticut (N=20) and New Hampshire (N=13) were small and thus very few Connecticut (n=7) and New Hampshire (n=5) participants responded to our survey. These responses, when reported, should be interpreted with caution. Small sample size is associated with low statistical power and a higher margin of error.

Participant Satisfaction

In all three states, participants reported moderate to high satisfaction with nearly all initiative elements, except the payment process (Figure A-1).

The majority of participants in Massachusetts, Connecticut, and New Hampshire did not know how to rate the payment process (i.e., provided a "Don't Know" response or a rating of "neither satisfied or dissatisfied"), which is expected since less than a third of respondents in the overall sample reported they had received payments. at the time of the survey.

Figure A-1. Participant Satisfaction on the Following Initiative Elements



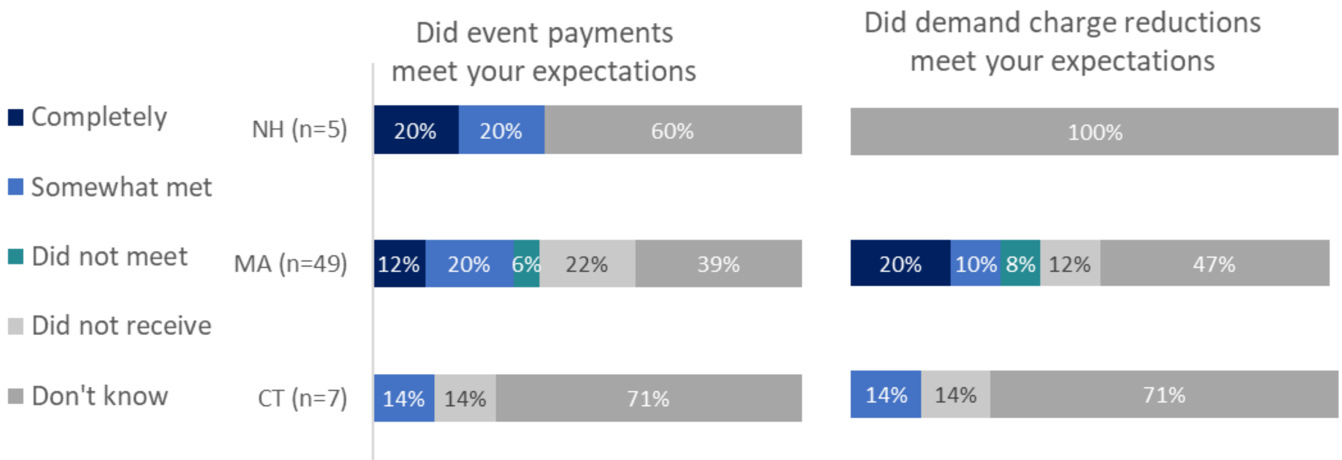
* Only those who had metering installed reported satisfaction on this item, which is 21 respondents in Massachusetts, two in Connecticut, and three in New Hampshire.

Those who indicated dissatisfaction were nearly all from Massachusetts. Section 4.1 documents these customers' reasons for dissatisfaction.

Participant Perceptions of Financial Benefits

Considering less than a third of respondents in the overall sample reported receiving payments at the time of the survey, it is also not surprising that a substantial proportion of surveyed participants in all three states gave a "Don't Know" and "Did not receive" response when asked whether event payments and demand charge reductions met their expectations (Figure A-2).

Figure A-2. Participant Responses on Financial Benefits



The minority who said their expectations were not met were respondents from Massachusetts. They explained:

- Not enough benefits (Savings from participation were low and the penalties “neutralized the winnings”, the ICAP tag reduction benefit was small, and no bidding into the ISO-NE since initiative rules were unclear about it)
- Unclear information on demand rates for the next year and need to conduct the demand cost analysis to assess benefits from the summer participation
- Limited battery availability to reduce demand
- Displeased with communication on results, data, and payments even after the results were received (No explanation on what results meant)
- Failed to achieve target demand reduction due to reluctance to affect the ongoing operations at the facility
- Need to refine the strategy to achieve more demand reductions

Facility Disruptions and Opting Out

In all three states, the majority of participants reported experiencing slight or temporary disruptions when participating in an event (Table A-1).

About 23% (or 7 of 31) of those who experienced slight, temporary, or other disruptions in Massachusetts said these challenges prevented them from participating in one or more summer DR events. None of those in Connecticut and one of four New Hampshire participants who experienced slight, temporary, or other disruptions said these challenges prevented them from participating in one or more summer DR events.

Table A-1. Participation Challenges (Multiple responses allowed)

Challenges			
	MA (n=49)	CT (n=7)	NH (n=5)
Slight or temporary disruptions to your company's core operations	24 (48%)	5 (71%)	3 (60%)
Complaints from facility occupants	9 (18%)	-	-
Major disruptions	1 (2%)	-	-
Other	2 (4%)	-	1 (20%)
Don't know	2 (4%)	-	-
No disruptions	18 (37%)	2 (29%)	1 (20%)

DR Actions Done to Curtail Load

Participants predominantly relied on manual actions to curtail load.

Under the plan with their CSP, 58% (28 of 48)[56] of surveyed curtailment participants in Massachusetts reported being responsible for making manual adjustments to curtail load. About one-half (3 of 7) of Connecticut's and one of five New Hampshire curtailment participants reported being responsible for making manual adjustments to curtail load.

Notably fewer Massachusetts participants reported their CSP planned to make automatic adjustments – about 23% (11 of 48) in Massachusetts. Other respondents, three of seven in Connecticut, and two of five in New Hampshire said they let their CSP make automatic adjustments. Among all those (n=16) who opted for automatic adjustments, only four reported also relying on manual adjustments to curtail load.

Table A-2 documents types of manual actions participants had done to curtail load. Among those who curtailed load manually, the majority (20 of 28 Massachusetts' and 2 of 3 Connecticut's customers) reported taking these actions every time they were alerted. In Massachusetts and Connecticut, participants reported being alerted about two times, on average.

[56] One respondent in Massachusetts used a battery to participate and was not asked this question.

Table A-2. Manual Curtailment Actions (Multiple responses allowed)

Count of actions they were responsible to make when the event was called			
	MA (n=28)	CT (n=3)	NH (n=1)
Turn off inessential lights	14	1	-
Run fewer chillers or reduce the chiller load in the building	12	-	-
Ramp down production activity	12	-	-
Turn off non-critical roof top units	10	1	-
Reduce the speeds on the air handler fans, and make a temporary adjustment to variable frequency drives on fans and pumps to meet temperature requirements	8	-	-
Raise the temperature set point	7	1	-
Raise the chilled water temperature set point	7	-	-
Turn off inessential split system units	6	-	1
Turn off unnecessary electric space conditioners	6	1	-
Switch off unnecessary return fan in areas served by multiple return fans	6	-	-
Turn electric reheat in-air handling units or variable air volume units completely off	4	-	-
Turn off unnecessary electric domestic hot water heaters	3	1	1
Run some cooling tower fans at lower speeds	3	-	-
Shut down non-critical elevators or escalators	2	-	-
Run a generator for the duration of the event	2	-	-
Other	4	1	1

Table A-3 documents types of automatic actions participants had opted to do to curtail load. Among those who curtailed load automatically, none opted out from automatic adjustments or from participating in the event(s).

Table A-3. Automated Curtailment Actions (Multiple responses allowed)

Count of actions they opted to automate and conduct when the event was called			
	MA (n=11)	CT (n=3)	NH (n=2)
Turn off inessential lights	6	1	-
Raise the temperature set point	3	1	-
Run fewer chillers or reduce the chiller load in the building	2	-	-
Ramp down production activity	2	1	1
Raise the chilled water temperature set point	2	-	-
Reduce the speeds on the air handler fans, and make a temporary adjustment to variable frequency drives on fans and pumps to meet temperature requirements	2	-	-
Run a generator for the duration of the event	2	-	-
Turn off non-critical roof top units	1	-	-
Turn off inessential split system units	1	-	-
Turn off unnecessary electric space conditioners	1	-	-
Switch off unnecessary return fan in areas served by multiple return fans	1	-	-
Turn electric reheat in-air handling units or variable air volume units completely off	1	-	-
Turn off unnecessary electric domestic hot water heaters	1	-	-
Run some cooling tower fans at lower speeds	1	-	-
Other – set all roof top units back 3-5 degrees	1	1	1
Other – automated a whole sequence of actions	1	-	-
Other – set all systems on "unoccupied" setting	1	-	-

Among those who engaged in manual and automatic actions to curtail load, most did these actions at the time of the event (Table A-4). Those who did something prior to the DR event(s) (n=13) reported what actions they took prior to the event(s). The most common response was either shutting down or preparing systems, equipment, or schedules to cycle down (nine responses).

Table A-4. When Did Manual Curtailment Occur

Under your plan with your CSP, were there [manual/automatic] adjustments that you or others made on DR event days either BEFORE the actual DR event began or AFTER the actual DR event finished?			
Manual Adjustments	**MA (n=28)**	**CT (n=3)**	**NH (n=1)**
Yes - made in the hours **prior** to when the DR event(s) began	5	-	-
Yes - made in the hours **both prior to and after** the DR event(s)	6	2	-
Yes - made in the hours **after** the DR event(s)	1	-	-
No	12	1	1
Don't know	4	-	-
Automatic Adjustments	**MA (n=11)**	**CT (n=3)**	**NH (n=2)**
Yes - made in the hours **prior** to when the DR event(s) began	1	-	-
Yes - made in the hours **both prior to and after** the DR event(s)	1	-	-
Yes - made in the hours **after** the DR event(s)	-	-	-
No	8	3	2
Don't know	1	-	-

Appendix B: Cross-cutting and PA Detailed Findings

Cross-cutting

Application and Contracts

Requiring that participants sign a PA application in addition to a separate CSP-customer agreement, both with their own T&Cs, complicates and slows the sales process. Three CSPs noted that the application is a roadblock in the sales process. Two also reported not understanding the purpose of the application for customers who participate in the initiative through a CSP (as opposed to independently). They explained that the PA contract captures the information that the customer has already provided upon signing a contract with the CSP. Asking the customer to sign an additional contract, a PA contract, can give the customer pause. One CSP explained:

> "Many times, it goes to the lawyers, and the lawyer says, 'why are we signing this? We already signed an agreement with you, [CSP]. Why are you making me sign another agreement with the PA?"

One CSP reported several large prospects refusing to sign the PA contract and thus not participating in the DR initiative.

The Unitil staff recognizes that the PA application complicates the sales process for CSPs. In one case, they accommodated a customer's request to modify the application T&Cs. Unitil also allowed one customer in NH to participate without filling out an application. They suggest that the utilities collaborate to revise the T&Cs to make the terms very agreeable and more relevant to customers.

Eversource staff also noted that with all procurement, contract T&C's are frequently modified to accommodate customer-specific issues, though there was no indication this happened for the 2019 summer season of the DR initiative.

Eversource DR Initiative Goals and Performance

The Eversource staff reported two initiative-specific goals:

- **Reduce peak period demand to maximize system and customer benefits.** Staff planned to call events during the ICAP hour, summer monthly peaks, and as many of the 20 highest hours of system demand during the summer as possible. This strategy would ease future system capacity needs and utility costs while generating savings for the customers.

- **Reduce carbon emissions.** On the days with the highest demand, the marginal generating units are usually inefficient, simple cycle gas peaker plants. Reduced operation of these plants would result in a beneficial emissions impact.

Eversource staff noted either meeting or exceeding their summer 2019 enrollment targets and achieving their demand reduction targets in nearly all their territories. Staff explained that in Connecticut and New Hampshire, Eversource surpassed their 2019 targets of 6 MW and 5 MW of reduction, respectively. In Massachusetts, the utility set a target of 30 MW of average participation but reported 23 MW reduction instead. The evaluation team's attempts to validate Eversource's estimates of load reduction, presented in Section 4.3 mostly corroborate the reported amounts. Remaining differences likely reflect discrepancies in the data available to the evaluation team and the utility implementation team, as discussed in Appendix D.

Eversource Marketing and Enrollment Process

The initiative is marketed through Eversource's website and by Account Executives (AEs) who are incentivized and responsible for recruiting customers to participate in Eversource's energy efficiency programs, including the DR initiative. Considering the utility met their enrollment goals, marketing and/or outreach strategy appears to be adequate.

CSPs also market and generate leads, at times together with Eversource AEs. Two CSPs met with AEs at Eversource headquarters at the start of the 2019 sales season to discuss recruiting, prospective customers, and their approach to DR. They wanted AEs to be prepared to answer any customer questions on the services of a particular CSP. The one CSP who had not met with the AEs expressed interest in collaborating with AEs on recruitment in the future, noting that this would help them recruit customers. Most of the customers this CSP had enrolled were existing clients that participated with them in National Grid's DR initiative and also had sites in the Eversource territory.

Eversource Contracting Challenges

The CSPs had notable difficulties with the utility contracts in the 2019 season. Three CSPs reported that Eversource changed the Terms and Conditions (T&C) on the application several times throughout the sales season (January – May). At certain points in time, there were multiple different versions of the T&Cs in circulation. These CSPs reported having a constant back-and-forth with Eversource (that lasted well into the summer) to negotiate the T&Cs and clarify rules. One CSP had to reach out to a few customers who had already signed the utility application, which at that point had been voided, and ask them to sign a new application with the revised T&Cs. Another CSP noted that Eversource had worked off their standard energy efficiency program application to develop their DR initiative contract and had failed to remove certain T&Cs that did not apply in the context of DR. This created confusion for some of the

CSP's prospective customers, and, in at least one case, a customer chose not to participate because they recognized that some of the T&Cs were irrelevant or extraneous.

Eversource Application Submissions, Data Tracking, and Program Management

Implementation staff described a number of initiative administration challenges:

- To enroll in the DR initiative, customers must sign and submit an application form through their CSP. This past summer, customers submitted applications as PDFs, which staff then had to code into a spreadsheet, making application processing a time-consuming task. To ease this burden, Eversource is building a platform to process applications electronically. Staff noted their goal is to have this platform ready to facilitate enrollment for the 2020 summer season.
- The utility staff also used multiple platforms for reporting, monitoring, and management. Event notifications were sent manually to CSPs (by email). To monitor dispatch, Eversource staff had to log in to CSP websites and access multiple tools. Concurrent with making the application process electronic, Eversource is designing infrastructure to enable reporting, monitoring, and dispatching of event communications within one Distributed Energy Resources Management System (DERMS). This DERMS platform will also facilitate tracking and customer segmentation.

The utility wants to grow DR initiative enrollment and enable all customer types to participate. To fully assess how the initiative can serve various customers, the utility will need better data on how each customer will participate in DR. Currently, Eversource's application asks customers to indicate their building type and describe their active demand reduction measures in long-form response. Moving forward, Eversource staff want to redesign the application to ensure customers report types of technologies and actions they plan to take to curtail load, as well as facility load shapes and site characteristics (i.e. hours of operation). Utility staff expects that this information will help the implementation team better predict actual delivered customer load. The DERMS platform will facilitate this data gathering process.

Eversource Participants' Reported Actions to Curtail Load

Table B-1 documents the types of manual actions participants took to curtail load. Among those who curtailed load manually, the majority (8 of 11) reported taking these actions for all or most of the times when they were alerted of an event. Participants reported receiving three DR event notifications on average.

Table B-1 also documents the types of automatic building adjustments CSPs planned to implement at participant sites to curtail load. Among participants whose load was curtailed automatically, none opted out from automatic adjustments or from participating in the event(s). Of note, the vast majority of customers reduced load through curtailment; only one participant reported dispatching a generator.

Table B-1. Manual and Automated Curtailment Actions (Multiple responses allowed)

Responses	Manual Count (n=10)	Automatic Count (n=7)
Lighting Actions		
Turn off inessential lights	5	3
Cooling Actions		
Raise the chilled water temperature setpoint	4	-
Raise the temperature setpoint	4	1
Run fewer chillers or reduce the chiller load in the building	3	-
Turn off non-critical rooftop units	3	-
Turn off inessential split system units	1	-
Turn electric reheat in-air handling units or variable air volume units completely off	1	-
Turn off unnecessary electric space conditioners	2	-
Other – set all rooftop units back by 3 to 5 degrees	-	1
Other – shut down refrigeration	1	-
Fan and Motor Actions		
Reduce speeds on the air handler fans and make temporary adjustment to VFDs on fans and pumps to meet temp. requirements	3	-
Run some cooling tower fans at lower speeds	1	-
Switch off unnecessary return fans in areas served by multiple fans	1	-
Other – turned off the kiln and motors manually	1	-
Process		
Ramp down manufacturing or production activity	-	1
Other – curtailed all operations	1	-
Other- initiated the curtailment in the control system manual, the rest was automated	1	1
Dispatching Generator		
Run a generator for the remainder of the event	-	1
Other		
Shut down non-critical elevators or escalators	1	-
Turn off unnecessary electric domestic hot water heaters	1	-

Eversource Participant Suggestions for Improvement

When asked what could be improved, participants suggested:

- Provide more clarity on program details, such as the payment turnaround and start and end of a season (two responses).
- Communicate results faster so a customer can adapt their strategy (one response).
- Provide additional clarity on state-level DR policies or requirements (one response).

- Shorten event time (one response). This participant noted they can substantially reduce more energy in 10-minute intervals than in two- to four-hour windows, and are thereby considering actively participating at the grid level.

All surveyed participants said they will continue participating in the initiative.

National Grid DR Initiative Goals and Performance

Interviewed National Grid staff described one initiative-specific goal: to reduce long-term distribution, transmission, and capacity costs.

National Grid has been delivering the DR initiative in its Massachusetts territory since the summer of 2017.[57] The planned MW reduction goal for the 2019 summer season was set at 62 MW. In the last three years, National Grid has seen customers reduce load 20% less, on average, than what they or their CSPs enrolled. National Grid's implementation team uses this metric when recruiting participants and enrolling capacity to meet the delivered MW reduction goal for each DR season.

National Grid staff reported surpassing their enrollment and reduction goals for targeted dispatch every year since the start of the DR initiative.

National Grid Marketing and Enrollment Process

Considering National Grid has surpassed its summer enrollment goals in the past three years, their marketing and/or outreach strategy for the DR initiative appears to be adequate. Like Eversource, National Grid promotes the initiative through their AEs, whose responsibility is to identify energy efficiency program offerings and solutions that are best suited for individual customers. AEs introduce the DR initiative to customers, provide them with the list of approved CSPs, and encourage them to reach out to a CSP. There have been some cases where CSPs have recruited customers without involving the utility's AEs. Sometimes, CSPs will encourage prospective customers to reach out to National Grid for reassurance about the initiative.

Two CSPs were highly satisfied with National Grid's marketing of the DR initiative. One mentioned that there was an effective collaboration between their and National Grid's sales teams. Another noted that National Grid staff were always willing to facilitate introductions between AEs and their staff, which was helpful. They strategized together with AEs on how to approach certain customers with a coordinated message. CSPs report that this collaboration brings assurance to customers upfront, both about the initiative and the CSP.

[57] National Grid administers DR programs in Massachusetts, Rhode Island and New York electric service territories. This evaluation examined only the DR initiative in Massachusetts.

Some CSPs suggested that National Grid's DR initiative promotional efforts could use some improvement. One said that some AEs provided new customer leads only to those CSPs with whom they had developed a closer working relationship, rather than introducing those opportunities to all the approved CSPs. Another CSP mentioned difficulty navigating the website to find information about the DR initiative.

The evaluation team also observed that the application form accessed via the National Grid website[58] allows you to select which PA you are a customer of, or through which PA you will be participating. This gives the impression that the application can be used by Eversource, National Grid, and Unitil customers. While Unitil did use the same application as National Grid, Eversource has a separate application with different T&C's. Although this was not brought up in any customer surveys or CSP interviews, it may confuse prospective participants.

National Grid Delivery and CSP Performance

National Grid staff were satisfied with CSP performance across years. Enrollment is currently double what it was in the summer of 2017, the CSPs have surpassed enrollment forecasts established before the initiative launched, and MW targets have been met for the past three seasons.

When asked about key lessons learned from the 2019 summer season, National Grid staff reported that since the summer peak came close to occurring on a Saturday, the utility now wants to build weekend events into the initiative. One CSP mentioned that they had already attended a productive meeting with National Grid about how best to incorporate weekend participation into the initiative rules.

National Grid Participants' Reported Actions to Curtail Load

Table B-2 documents types of manual actions participants took to curtail load. Among those who curtailed load manually, nearly all (21 of 22) reported taking these actions every time or most of the times they were alerted of an event. Participants reported being alerted two times, on average, of a DR event.

Table B-2 also documents the types of automatic building adjustments CSPs planned to implement at participant sites to curtail load. Among participants whose load was curtailed automatically, none opted out from automatic adjustments or from participating in the event(s).

[58] https://www.nationalgridus.com/media/pdfs/bus-ways-to-save/mademandresponseapplication.pdf

Table B-2. Manual and Automated Curtailment Actions (Multiple responses allowed)

Responses	Manual Count (n=22)	Automatic Count (n=7)
Lighting Actions		
Turn off inessential lights	10	4
Cooling Actions		
Run fewer chillers or reduce the chiller load in the building	9	2
Turn off non-critical rooftop units	8	1
Turn off unnecessary electric space conditioners	5	1
Turn off inessential split system units	5	1
Raise the temperature setpoint	4	3
Raise the chilled water temperature setpoint	3	2
Turn electric reheat in-air handling units or variable air volume units completely off	3	1
Other – set all rooftop units back by 3 to 5 degrees	-	1
Fan and Motor Actions		
Reduce speeds on the air handler fans and make temporary adjustment to VFDs on fans and pumps to meet temp. requirements	5	2
Switch off unnecessary return fans in areas served by multiple fans	5	1
Run some cooling tower fans at lower speeds	2	1
Process		
Ramp down manufacturing or production activity	9	1
Other – shut down the whole plant or production	2	-
Other – shut down manufacturing induction furnaces	1	-
Other – send limited school staff in the summer home a little early	1	-
Other – turn everything on “unoccupied” setting	-	1
Dispatching Generator		
Run a generator for the remainder of the event	2	1
Other		
Turn off unnecessary electric domestic hot water heaters	3	1
Shut down non-critical elevators or escalators	1	-
Other – shut down unnecessary power draw	1	

National Grid Participant Suggestions for Improvement

When asked how the DR initiative could be improved to make it better for their organization, four participants indicated that they would like more frequent and/or better communication around event performance and payments after events.

One CSP noted they provide scorecards to all customers after each event that show performance. Other CSPs indicated that customers who have CSP meters have access to an online portal where they can see event performance and projected payments. One CSP mentioned frequently encouraging customers to access the portal to see their performance in real-time during events. Even with these resources, customers still expressed a need for better communication about their performance and payments.

All but three participants (29 of 32) will continue to participate in the utility's DR initiative. The remaining three said they are unsure as to whether they will continue to participate.

Unitil DR Initiative Goals

The interviewed Unitil staff described two initiative-specific goals:

- **To provide customer savings.** Unitil is focused primarily on shedding as much load as possible during the three peak hours of the year (including especially the ICAP hour), as this is where customers get the most dollar savings.
- **To achieve system and utility grid benefits.** The utility wants to achieve demand reduction benefits (in $'s). Implementing DR on other high demand days (besides the ICAP day) also benefits the grid.

Unitil achieved its 2019 MW reduction target in Massachusetts, but not in New Hampshire. The utility set a goal to enroll at least one customer and achieve 200 kW of delivered reduction per participant in Massachusetts, and staff reported achieving 853kW of demand reduction (on average 284 kW per participant). In New Hampshire, the utility set a goal of 1.8 MW of reduction and staff reported achieving a reduction of 1.3 MW. The initiative[59] was not approved by the New Hampshire commission until April 2019, which indicates that the utility had limited time to market it before the start of the summer season and enroll enough participants to achieve its goal of 1.8 MW of reduction.

In Massachusetts, the goal for 2020 increases to 400 kW of delivered reduction. In New Hampshire, Unitil's delivered reduction goal will increase to nearly 3 MW.

Unitil Delivery and CSP Performance

Unitil staff were highly satisfied with the CSP's performance during the summer 2019 season. Staff noted that the CSP had made the enrollment process easy and all paperwork had been submitted in a timely fashion. The CSP was also quick to respond to any correspondence and helpful in reconciling data issues that came up when settlements were calculated.

In 2020, Unitil staff will call more than one event over the summer season. This way, if customers are unable to perform during one event or curtail during the ICAP hour, they will have other opportunities to participate and receive at least some financial benefits.

[59] Note that the DR initiative in New Hampshire is currently a demonstration program.

Appendix C: Settlement Verification

This section details the actions taken to duplicate the settlement baseline and load reduction estimates reported by each utility, in addition to issues encountered during the verification process.

In addition to reported and validated load reduction, load reduction for the evaluated asymmetric baseline are presented for comparison. Evaluated baselines require that each similar day included in the baseline pool has load values for at least half of the intervals for each hour from 1 p.m. through 7 p.m. (includes all adjustment period and event period times evaluated in this study). If more than half of the intervals are missing for any hour from 1 p.m. through 7 p.m., the day is considered ineligible for inclusion in the baseline pool and the next most recent similar day will be considered for baseline eligibility. The maximum lookback for baseline pool days is 42 calendar days. Ultimately, no more than 25% of intervals for the baseline or the event day during the adjustment and event time periods may be missing for the load reduction estimate to be considered valid.

Eversource

Eversource uses a DERMS vendor as the primary source of event load reduction estimates for settlement. Eversource uses vendor CSP estimates as a secondary source, as needed. CSP estimates may be used when the vendor has interval meter data that Eversource does not. To estimate the settlement baseline, the DERMS requires that each similar day included in the baseline pool has at least one interval for each hour of the day. If all intervals are missing for any single hour, the day is considered to be ineligible for inclusion in the baseline pool and the next most recent similar day will be considered for baseline eligibility. As the initiative is not punitive, negative event load reduction is considered to be zero. While not punitive, no performance for an event will lower the season performance average, which is used in customer settlement. The evaluation team attempts to validate the asymmetric baseline used by the DERMS.

Table B-5-1., Table B-5-2., and Table B-5-3. compare the reported asymmetric load reduction with evaluated load reduction for Eversource in MA, NH, and CT, respectively. The tables show the evaluated load reduction for the validation and for its own version of the asymmetric baseline, which is common across utilities.

Table B-5-1. Eversource - MA - Settlement Validation

Event Date	7/19/2019	7/30/2019	8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Reported - Asymmetric	18,155	24,990	22,137
Evaluated - Validation	19,500	24,346	18,847
Evaluated - Asymmetric	15,557	23,877	18,656
Accounts	**160**	**151**	**135**

Table B-5-2. Eversource - NH - Settlement Validation

Event Date	7/19/2019	7/30/2019	8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Reported - Asymmetric	4,988	5,552	1,814
Evaluated - Validation	4,796	7,012	1,679
Evaluated - Asymmetric	4,619	6,902	1,659
Accounts	**40**	**40**	**19**

Table B-5-3. Eversource - CT - Settlement Validation

Event Date	7/19/2019	7/30/2019	8/19/2019
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Reported - Asymmetric	12,575	13,135	13,085
Evaluated - Validation	12,331	12,663	12,271
Evaluated - Asymmetric	11,898	12,347	11,822
Accounts	**95**	**96**	**95**

There are several possible reasons for differences between reported load reduction and the evaluation team's validation of reported load reduction.

- Eversource interval meter data for an account is unavailable. When Eversource interval meter data is unavailable or the DERMS is unable to estimate load reduction, Eversource reports the vendor's estimate of load reduction. The evaluation team cannot validate load reduction for accounts that interval meter data has not been received or where the underlying data is different.

- For some accounts, the evaluation team received multiple meters. Differences in the aggregation of account meters will lead to different estimates of load reduction.
- For some accounts, DEL meters (electricity delivered by Eversource), NET meters (net metered electricity), or both DEL and NET meters were received. For accounts with valid data for both DEL and NET meters, NET meters were used by the evaluation team. If only DEL meters were received or the NET meters were missing event data, DEL meters were used by the evaluation team. Differences in the use of DEL or NET meter type will lead to different estimates of load reduction.

National Grid

National Grid calculates reported load reduction and validates vendor load reduction estimates internally. To estimate the settlement baseline, National Grid includes the most recent similar days for the baseline pool. Though no data requirements were applied to the most recent days for inclusion in the baseline pool, National Grid works closely with vendors to acquire data for accounts that National Grid is missing interval meter data. Accounts for which National Grid is ultimately unable to acquire data are reported and paid based upon their enrolled capacity. Note that National Grid updates account enrolled capacity each year based upon performance in the previous year. As the initiative is not punitive, negative event load reduction is considered to be zero. While not punitive, no performance for an event will lower the season performance average, which customer settlement is determined by. The evaluation team attempts to validate the asymmetric baseline used by National Grid. Table B-5-4. compares reported load reduction with evaluated load reduction for National Grid in MA during its only summer 2019 weekday event. The table shows National Grid reported load reduction split into two rows: one for accounts that the evaluation team did not receive interval meter data and one for those that the evaluation team did. Note that National Grid's performance report confirmed that data was missing for all accounts that the evaluation team did not receive interval meter data for. The table also shows the evaluation team's estimates of load reduction for the validation and for its own version of the asymmetric baseline, which is common across utilities.

Table B-5-4. National Grid – July 30th, 2019 – Settlement Validation

Event Date	7/30/2019
Result	Average Hourly Reduction (kW)
Reported - Asymmetric	71,428
Evaluated - Validation	71,611
Evaluated - Asymmetric	69,561
Accounts	**357**

There are several possible reasons for differences between reported load reduction and the evaluation team's validation of reported load reduction. One asset drives the difference between the reported asymmetric and validated results. National Grid reported that performance for this asset was disputed.

Unitil

Unitil provided interval meter data and performance calculation workbooks for all accounts. It is unclear if Unitil applies any data requirements to baseline days, however, interval meter data was complete for all accounts during the only event and associated baseline pool days. The evaluation team attempts to validate the asymmetric baseline used by Unitil.

Table B-5-5. compares reported load reduction with evaluated load reduction for Unitil in MA and NH during its only summer 2019 event. The evaluation team received complete data for all accounts.

Table B-5-5. Unitil - July 30th, 2019 - Settlement Validation

State	MA	NH
Result	Average Hourly Reduction (kW)	Average Hourly Reduction (kW)
Reported - Asymmetric	853	1,299
Evaluated - Validation	843	1,363
Evaluated - Asymmetric	843	1,363
Accounts	**3**	**7**

There are several possible reasons for differences between reported load reduction and the evaluation team's validation of reported load reduction.

- Interval meter data appears to be adjusted for daylight savings time for some accounts, while not adjusted for daylight savings time for others.

Interval meter data included in the performance calculation workbooks does not appear to match the interval meter data received for all accounts. As the interval meter data and performance calculation workbooks were received at two distinct times, but not necessarily the same time for an account, the interval meter data and performance calculation workbooks may not contain the same data. At a minimum, the data is sometimes in different granularities, however, than in and of itself, does not constitute a reason for why differences exist.

Appendix D: Weather Correlation Plots

This appendix section provides the remaining weather-correlation plots. The plots support similar conclusions as those derived for plots from Eversource in the Integrated Impact and Process Evaluation Findings.

Figure 5-1. Eversource - Connecticut - July 30th Event

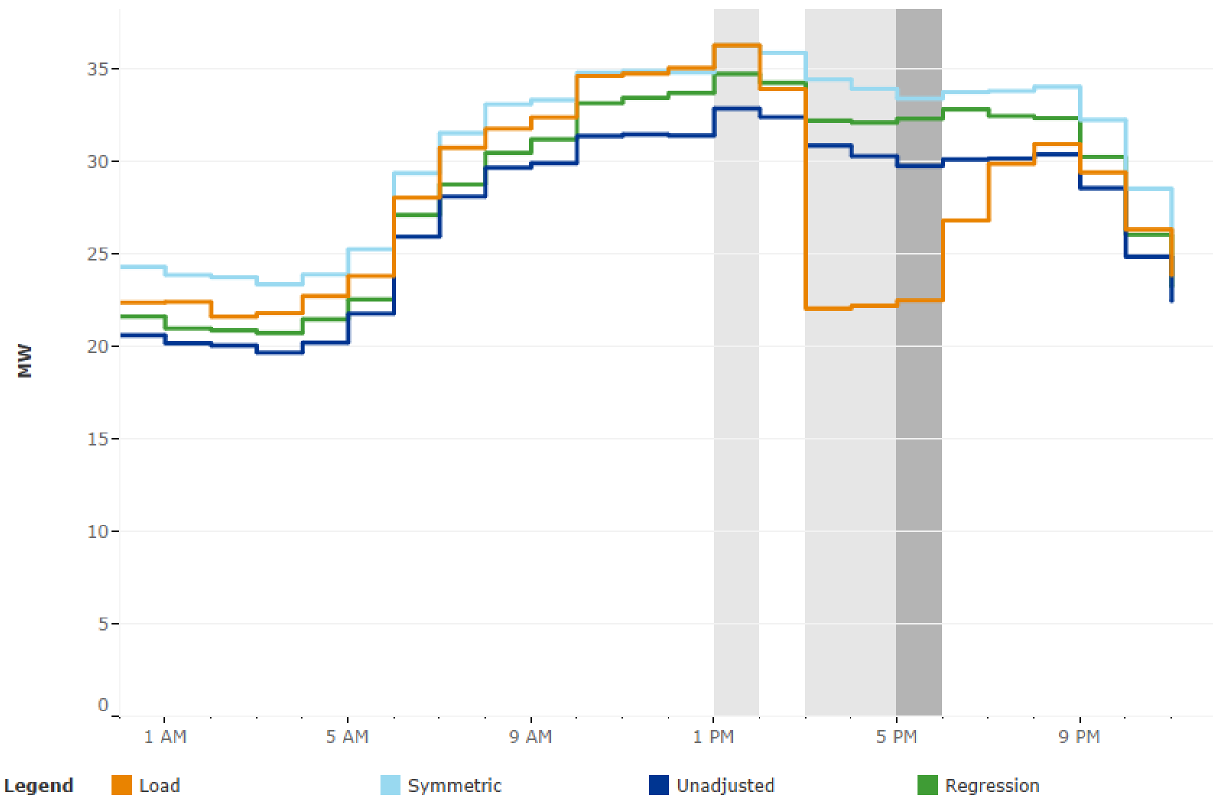


Figure 5-2. Eversource - Connecticut - July 30th Event – Weather Correlation

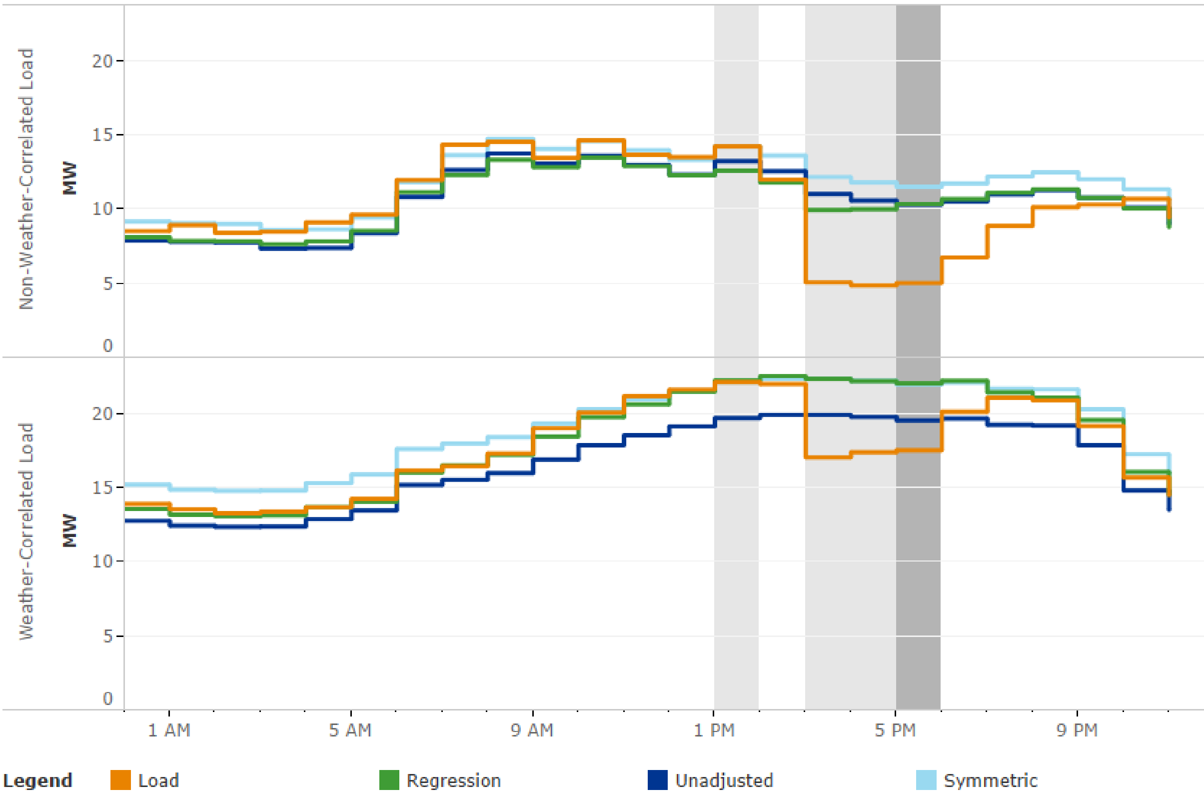


Figure 5-3. National Grid - Massachusetts - July 30th Event

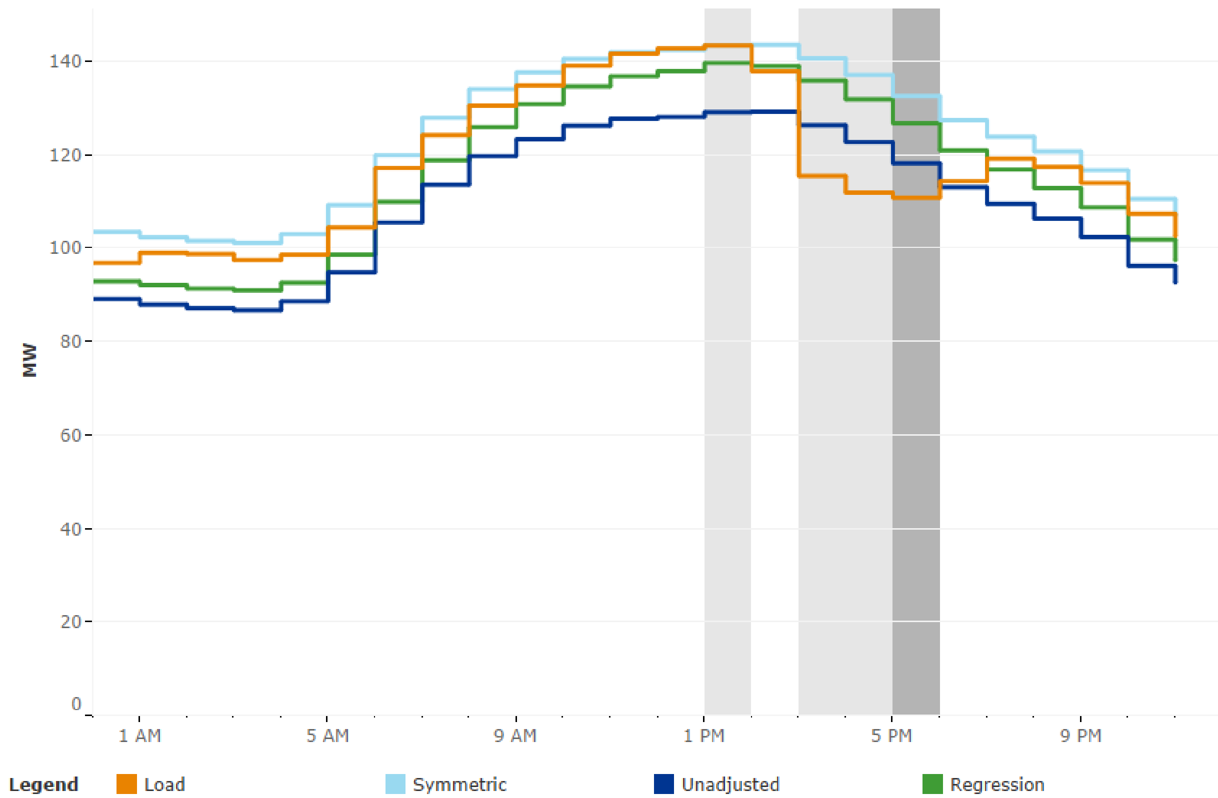


Figure 5-4. National Grid - Massachusetts - July 30th Event – Weather Correlation

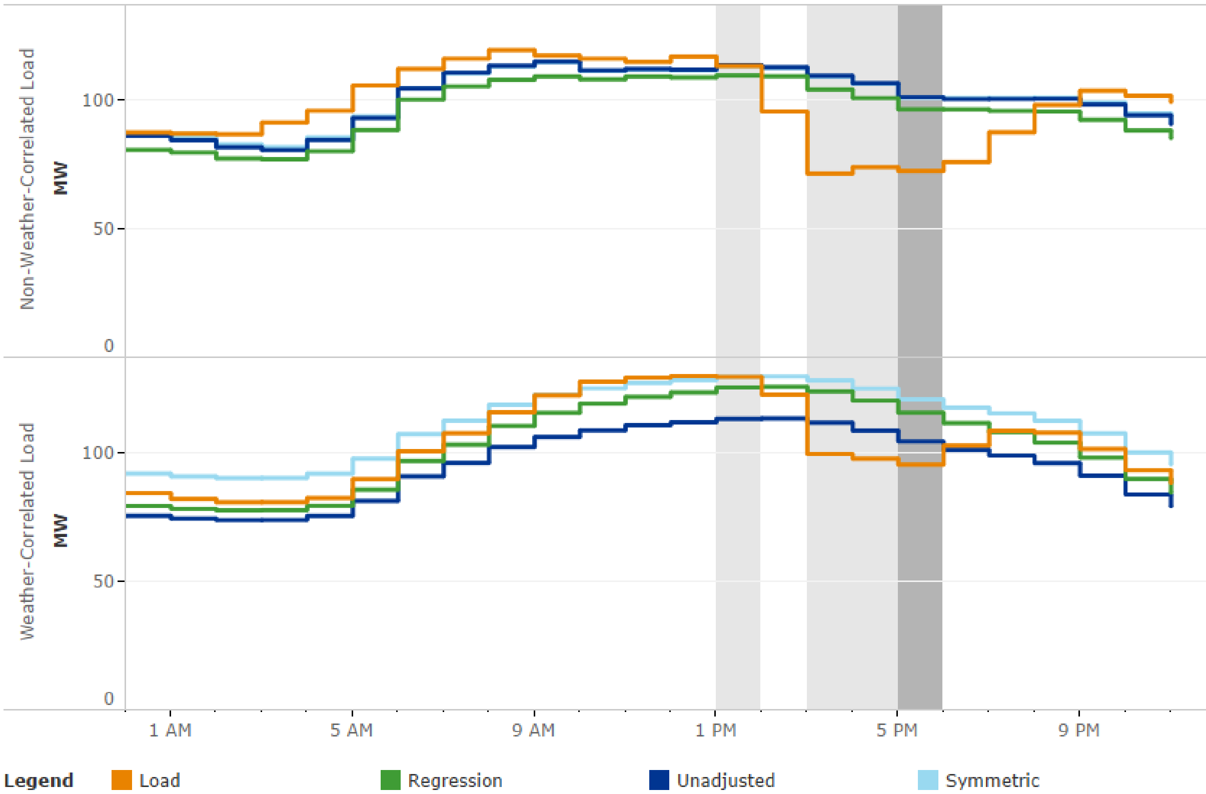


Figure 5-5. Unitil - Massachusetts - July 30th Event

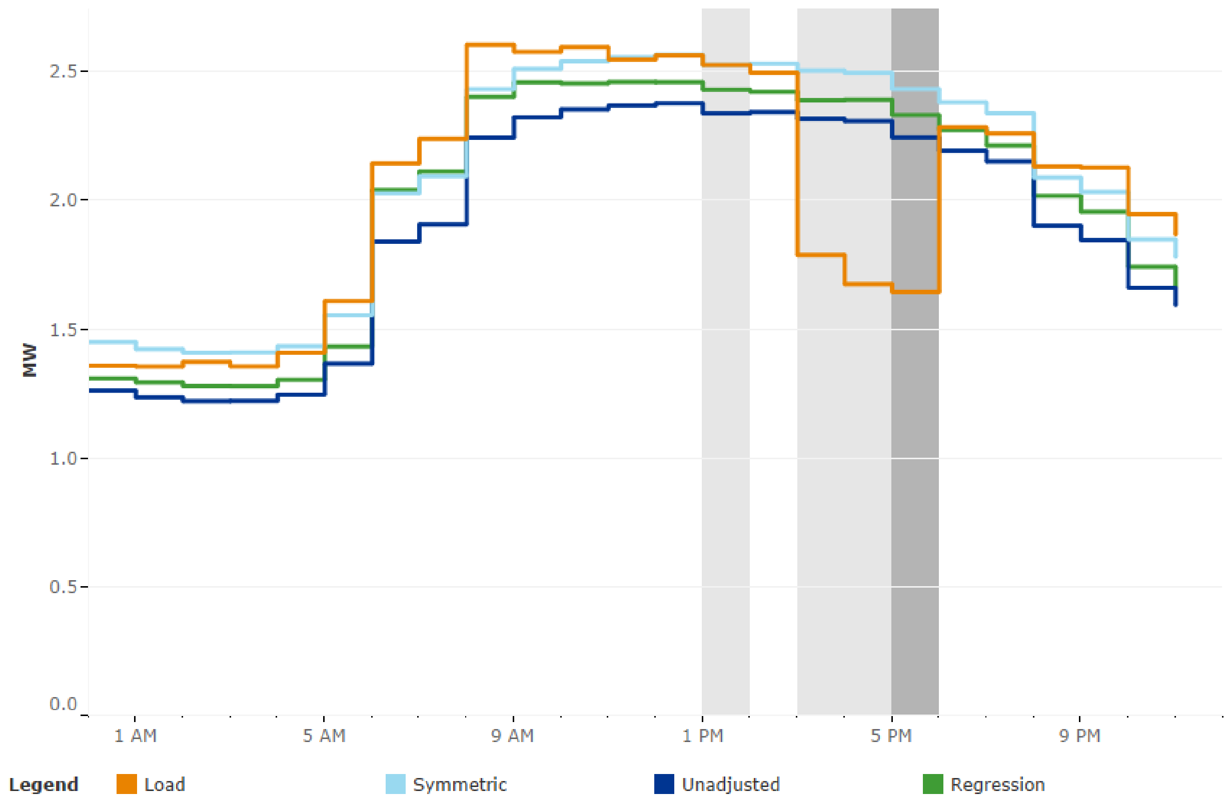


Figure 5-6. Unitil - Massachusetts - July 30th Event – Weather Correlation

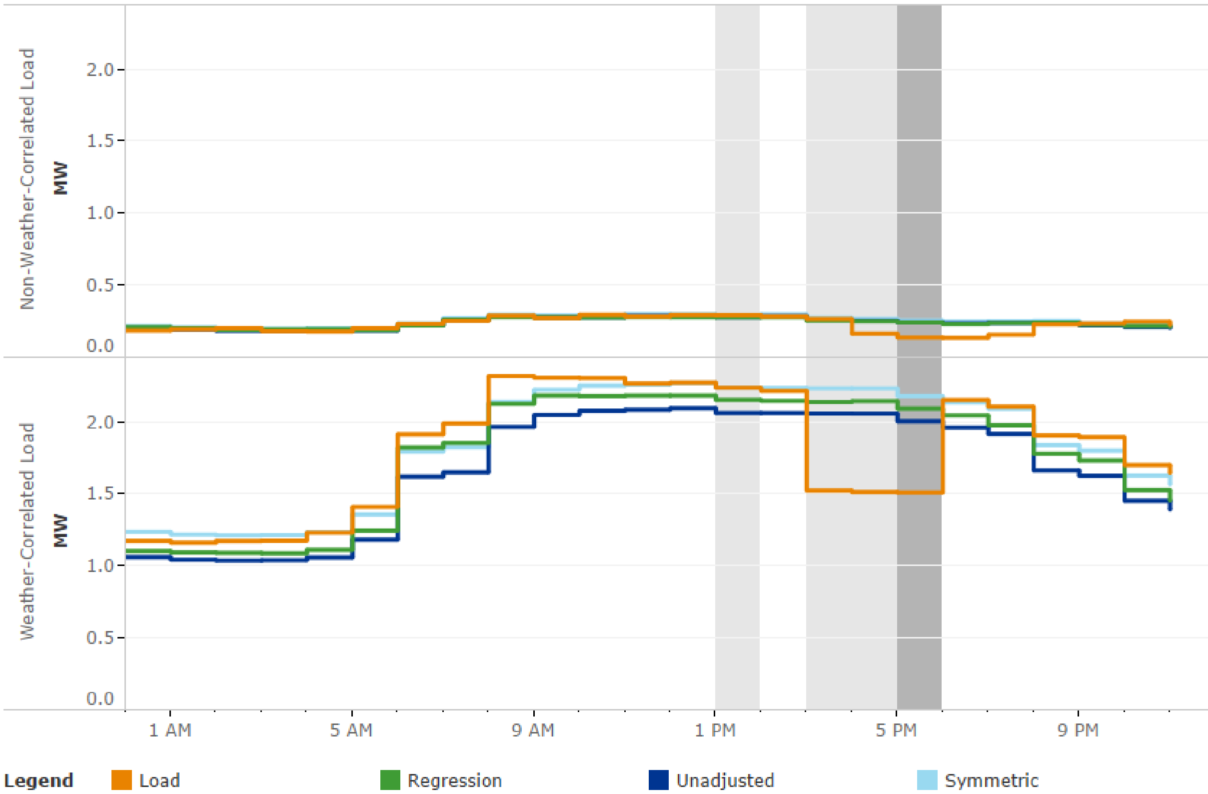


Figure 5-7. Unitil - New Hampshire - July 30th Event

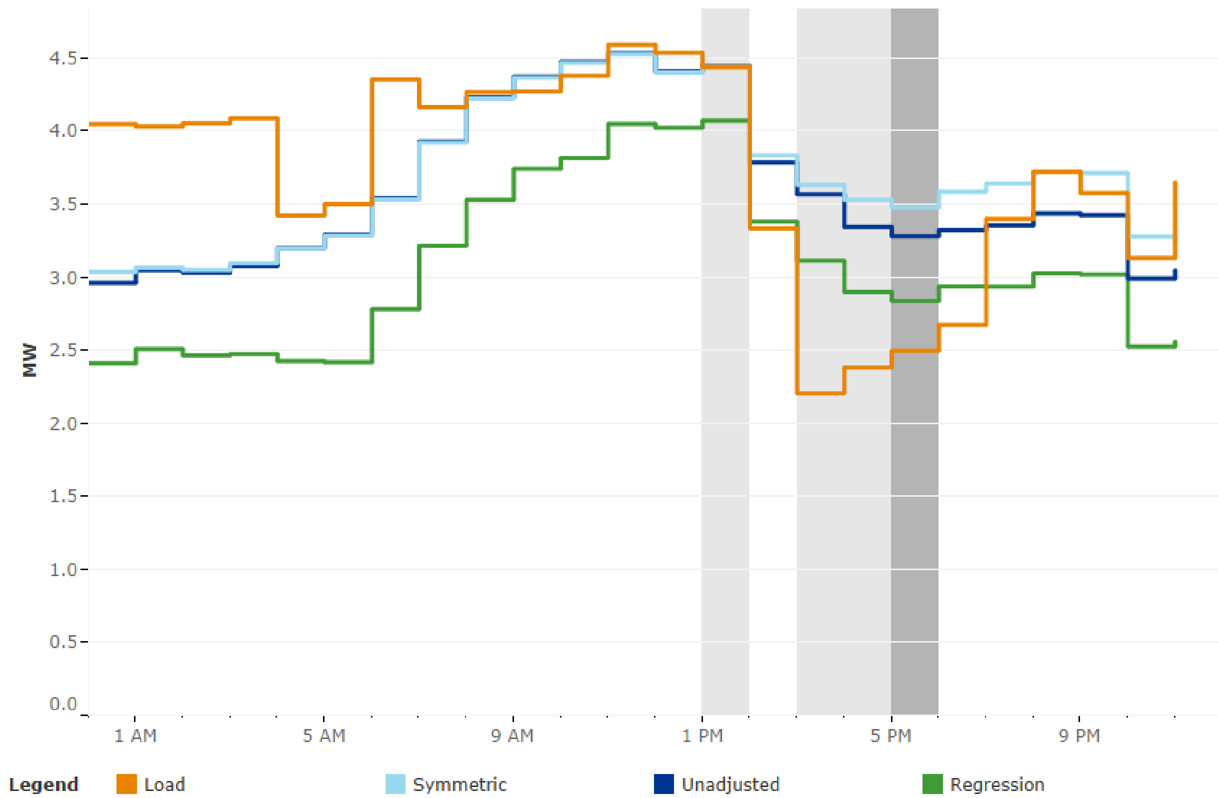
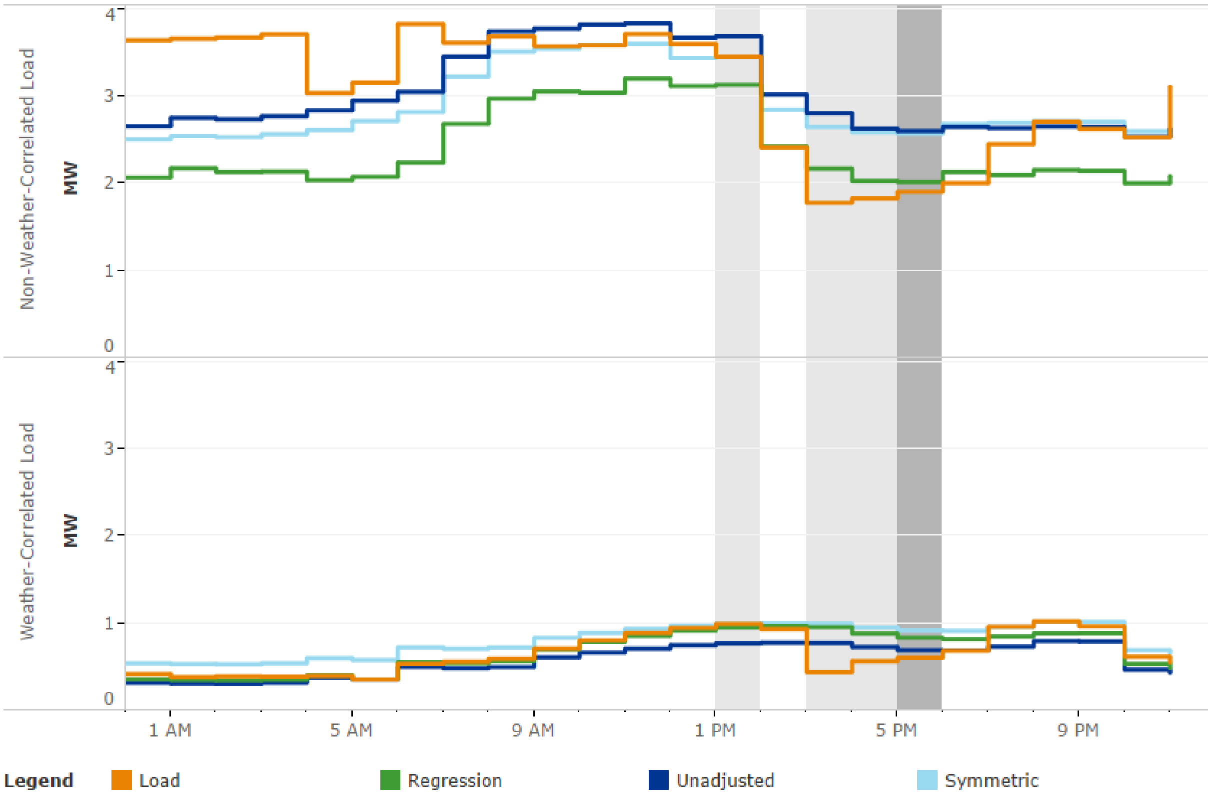


Figure 5-8. Unitil - New Hampshire - July 30th Event – Weather Correlation

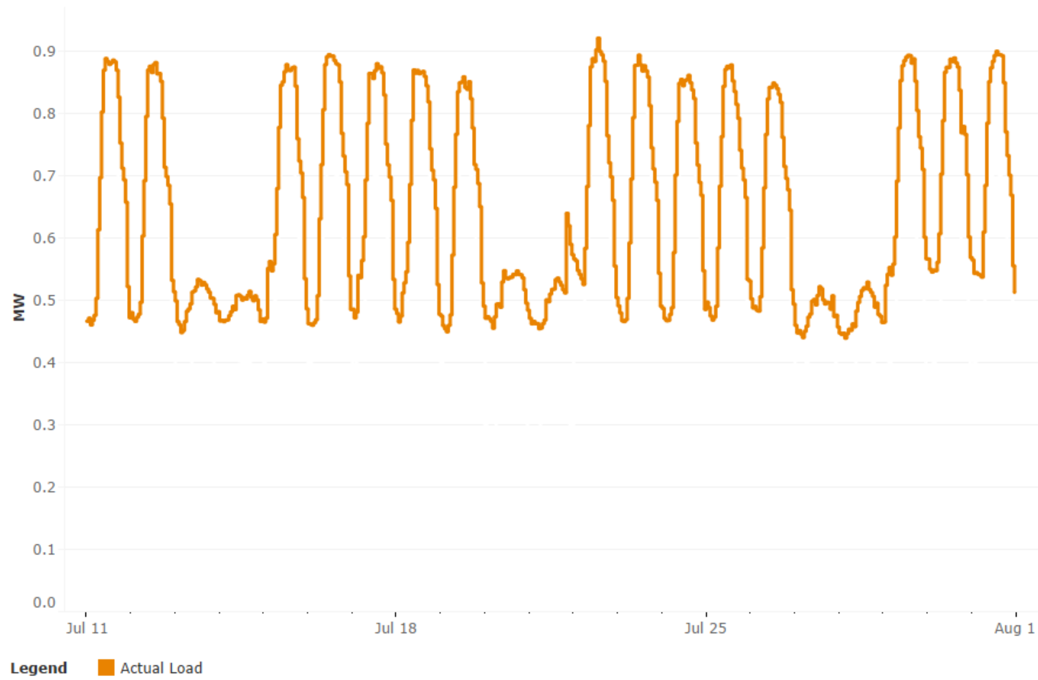


Appendix E: Fundamentals of DR Evaluation

This section is designed to provide a high-level introduction to the use of baselines to estimate load reduction in a demand response program.

Settlement baselines are designed to facilitate simple, transparent, and accurate measurement of load reduction. Under reasonable assumptions, they perform their job admirably. We will explore how baselines work under reasonable assumptions and how they start to fail as those assumptions do not hold. An initial assumption, for most baselines, is that the underlying load has a regular pattern. Figure 5-9. Example Customer 1 - Summer Period Load provides an example of 3 weeks of load data with a reasonably regular pattern. The weekdays, the sets of five higher load days, have a consistent shape and magnitude throughout this series.

Figure 5-9. Example Customer 1 - Summer Period Load



Most settlement baselines use data from prior days to establish an average shape for the event day. The ADR initiative uses a 10-of-10 baseline. This means the ten most recent eligible days are averaged, interval by interval, to produce a single average shape representing those days. Eligible days are non-holiday, weekdays that have not been flagged in advance for a shutdown. Figure 5-10. Example Customer 1 - Baseline Days for July 29th provides the ten days that would constitute the baseline day for an event that would occur July 29th.

Figure 5-10. Example Customer 1 - Baseline Days for July 29th

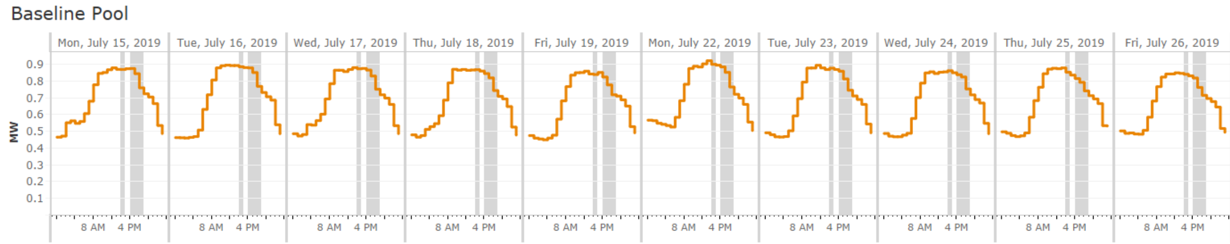
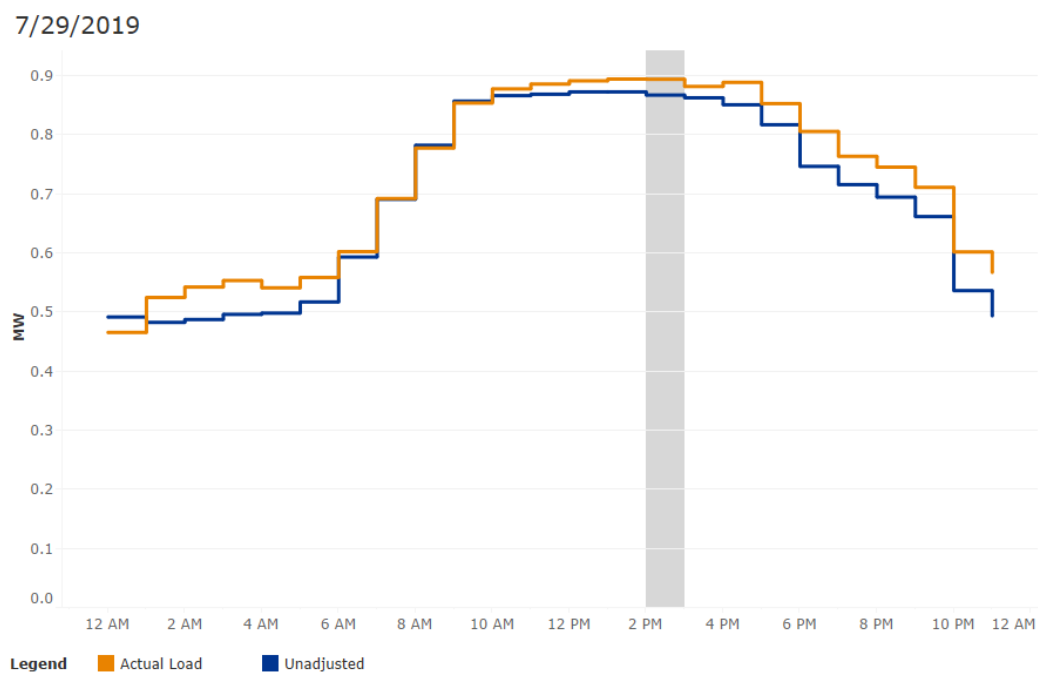


Figure 5-11. Example Customer 1 on July 29th overlays the average load shape from the ten baseline days over the actual load on July 29th. This baseline is referred to as the unadjusted 10-of-10 baseline. The 10-of-10 is a widely used baseline but baselines constructed of as few as 5 days and as many as 20 have been used. Also, some baselines will select within the initial set of baseline days. Baselines with names like "middle 8-of-10" or "top 5-of-10" are self-explanatory in terms of construction. In terms of motivation, these x-of-y baselines generally attempt to make the resulting baseline more accurate or representative of the event day. The former drops the extreme days perhaps assuming there are extreme load days that could dominate the average shape while the latter chooses the higher load days on the assumption that the event day will likely be a higher load-type day.

Figure 5-11. Example Customer 1 on July 29th



Even for this customer with regular load, the unadjusted baseline is not a perfect estimator of load on the day after the baseline days. In this example, the actual load is flatter than the baseline, with similar maximum loads in the middle of the day but higher relative loads at either end of the day.[60] In this example, the load diverges between 4 and 7 p.m., the expected event period.

Because baseline are frequently not ideal estimators of load during the key event periods, it is common to adjust baselines to load just prior to the event period. Figure 5-12. Example Customer 1 on July 29th illustrates a shaded adjustment period between 2 and 3 p.m. where the baseline load is adjusted to match the actual load. The load increment added to the baseline to make baseline load equal to actual load in that hour is applied to the baseline for all hours of the day.[61] The baseline is described as symmetrically adjusted because the adjust can be added or subtracted, moving the baseline either up or down, depending on the level of the adjustment period load. Under programs without day-ahead notification, the pre-event period is very likely to be a true indicator of how the event day load compares to the average load of the prior ten days. The baseline shape might not be perfect, but errors can only get so great over the course of a few hours. In contrast, with day-ahead notification, adjustment is still usually the most accurate baseline, but it may reflect customers' responses to the upcoming event, including pre-cooling, gaming etc.

[60] This is somewhat unusual as the opposite is generally the case with customers with weather-correlated load. HVAC loads are generally lower on more mild days producing flatter shapes. Events are more likely to occur on more extreme days when HVAC load gives the shape a more convex shape. It is typical that baseline days are milder than event days, thus producing a baseline that is flat compared to actual load. Mitigating this issue is the motivation behind the high 5-of-10 baseline.

[61] This is an additive adjustment. Alternatively, the ratio of load to baseline in that adjustment hour can be applied to each baseline hour. Referred to as a multiplicative adjustment, this approach has the intuitive characteristic of scaling the adjustment based on the load level. The reason multiplicative adjustments are used less frequently is small loads during the adjustment period can cause extreme adjustments. Where a multiplicative adjustment is used it is usually accompanied by a prescribed minimum and a maximum adjustment of, perhaps, 20%.

Figure 5-12. Example Customer 1 on July 29th

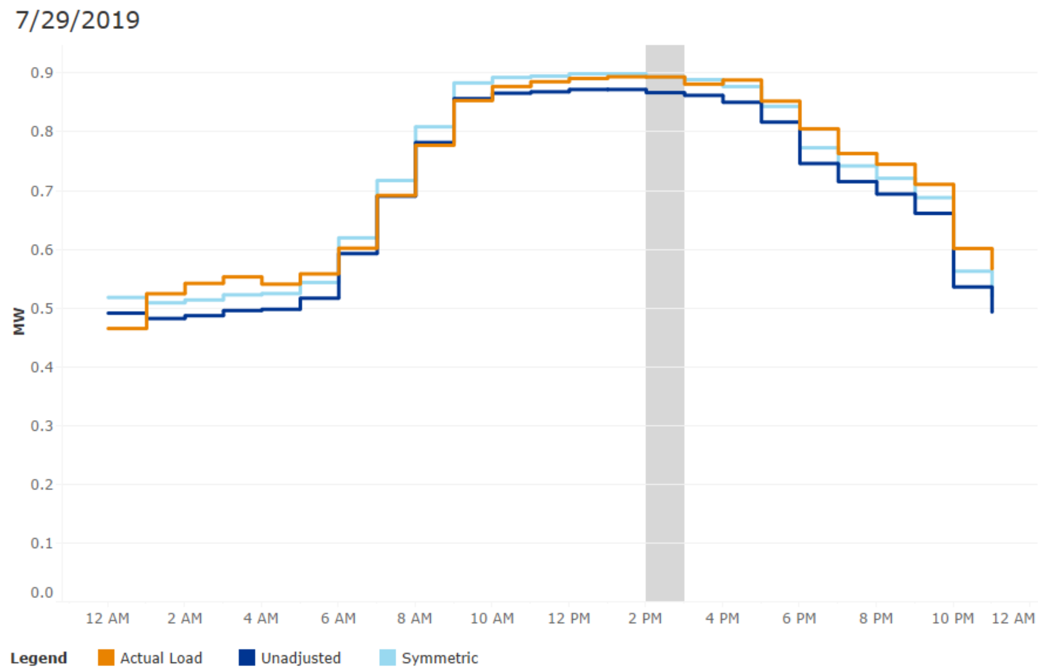


Figure 5-13. Example Customer 1 on July 29th adds a regression baseline to the plot. In this example, the regression does a particularly poor job of fitting the actual load data. The regression, unlike the 10-of-10, attempts to model all of the data from the summer. When a customer's load is weather-correlated and otherwise consistent across the summer, the regression can characterize weather-correlated loads reasonably well without resorting to an adjustment. If the data are highly variable, then the regression may do a worse job of characterizing the data than the recent and relatively short duration data in the 10-of-10 baseline.

Figure 5-13. Example Customer 1 on July 29th

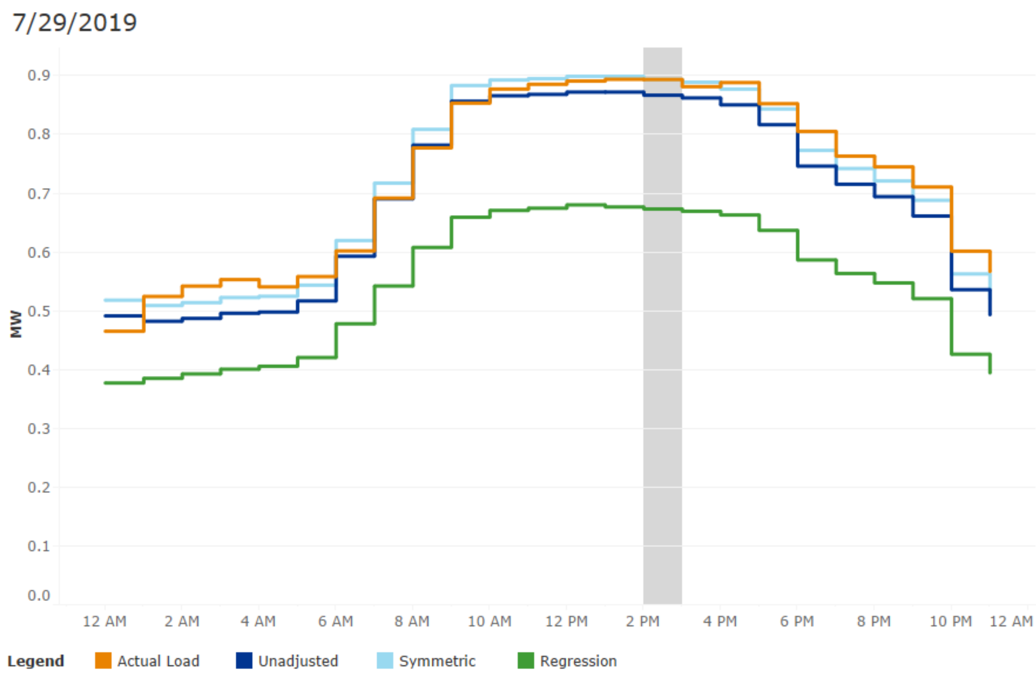


Figure 5-14 illustrates why the regression provides such a poor estimate. The plot provides the full summer of data for the same customer with the three weeks of load data that were discussed in prior plots circled. Across the summer, the data switches between two modes, one of which is effectively double the other. The two 10-of-10 baseline examples come from a time period of consistently high load. The regression, in contrast, is attempting to characterize the two different levels with a single structure. The light green line in the background illustrates how the regression produces baseline estimates that fall between the two load levels. Thus, on July 29th, the estimate is too low, but had the event fallen at the end of June, the baseline would have been way too high.

Figure 5-14. Example Customer 1 - Full Summer Load

Figure 5-15 provides the plot for the ICAP event day, July 30th. This plot looks remarkably like the plot from the prior day (Figure 5-13), even though it was an actual event day. This customer clearly failed to offer any load reduction to the program on this event day. The adjusted 10 of 10 baseline is difficult to see because it is perfectly aligned with the actual load through the event period.

Figure 5-15. Example Customer 1 on ICAP Event Day

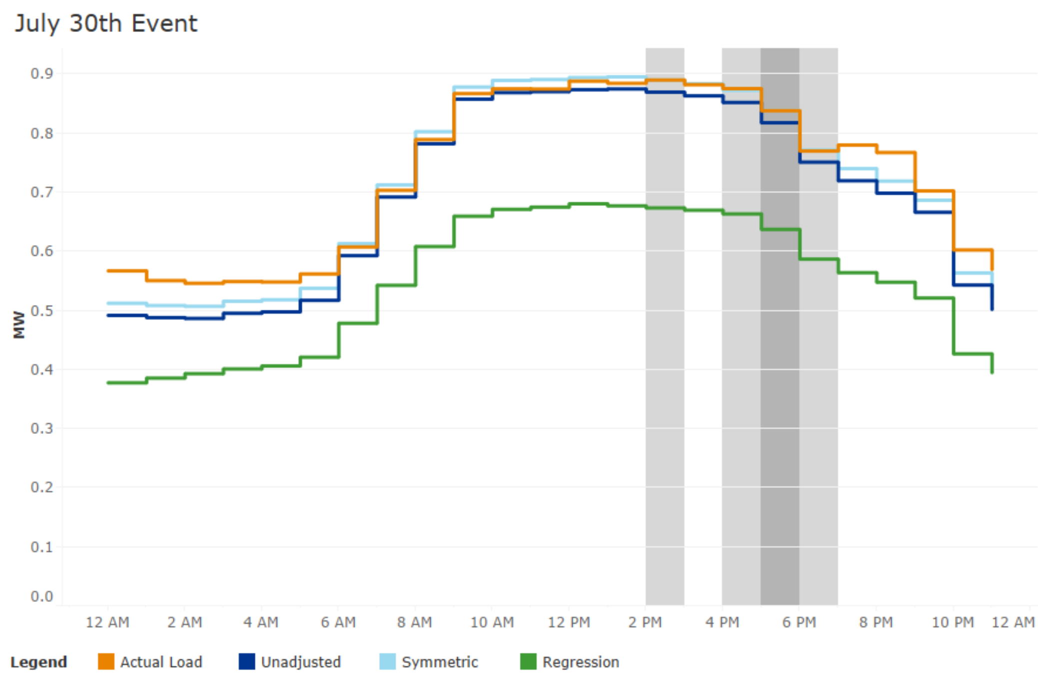


Figure 5-16. Example Customer 1 on August 1st provides that same plot for August 1st, two days after the ICAP event. As Figure 5-14 indicates, the load dropped to the lower mode on August 1st, and in August, the regression does a reasonably good job of characterizing the customer's load. The unadjusted 10 of 10 baseline, still primarily informed by the high load period in July is more than double actual load.

Figure 5-16. Example Customer 1 on August 1st

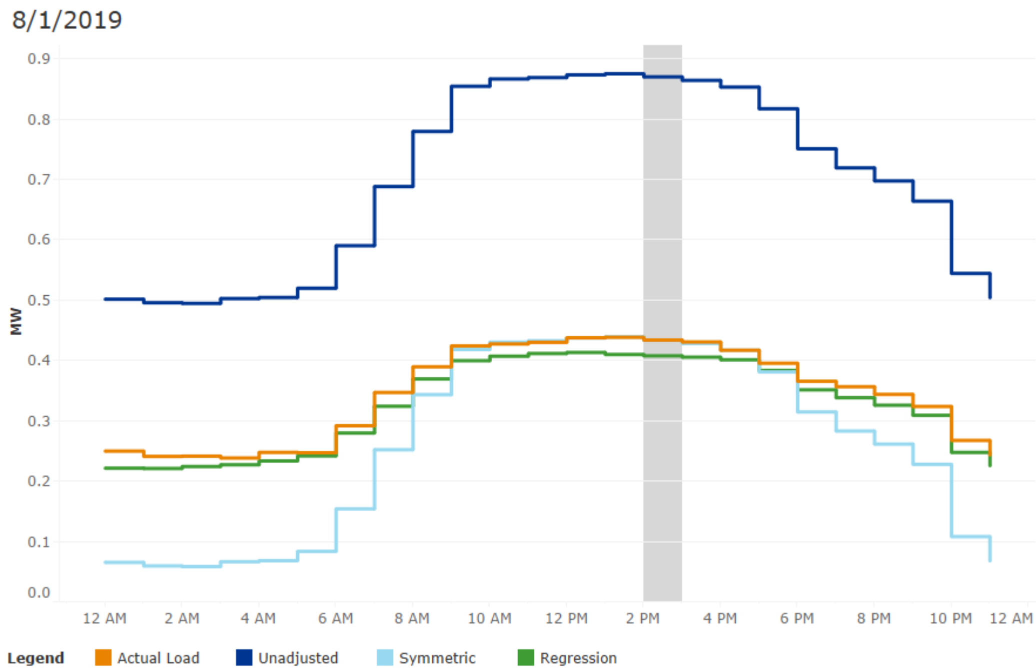


Figure 5-17. Example Customer 2 - Weather-Correlated Summer Load plots the full summer load of a more conventionally weather-correlated load. The light blue line in the plot represents the temperature, and the weekday load clearly tracks with the temperature.

Figure 5-17. Example Customer 2 - Weather-Correlated Summer Load

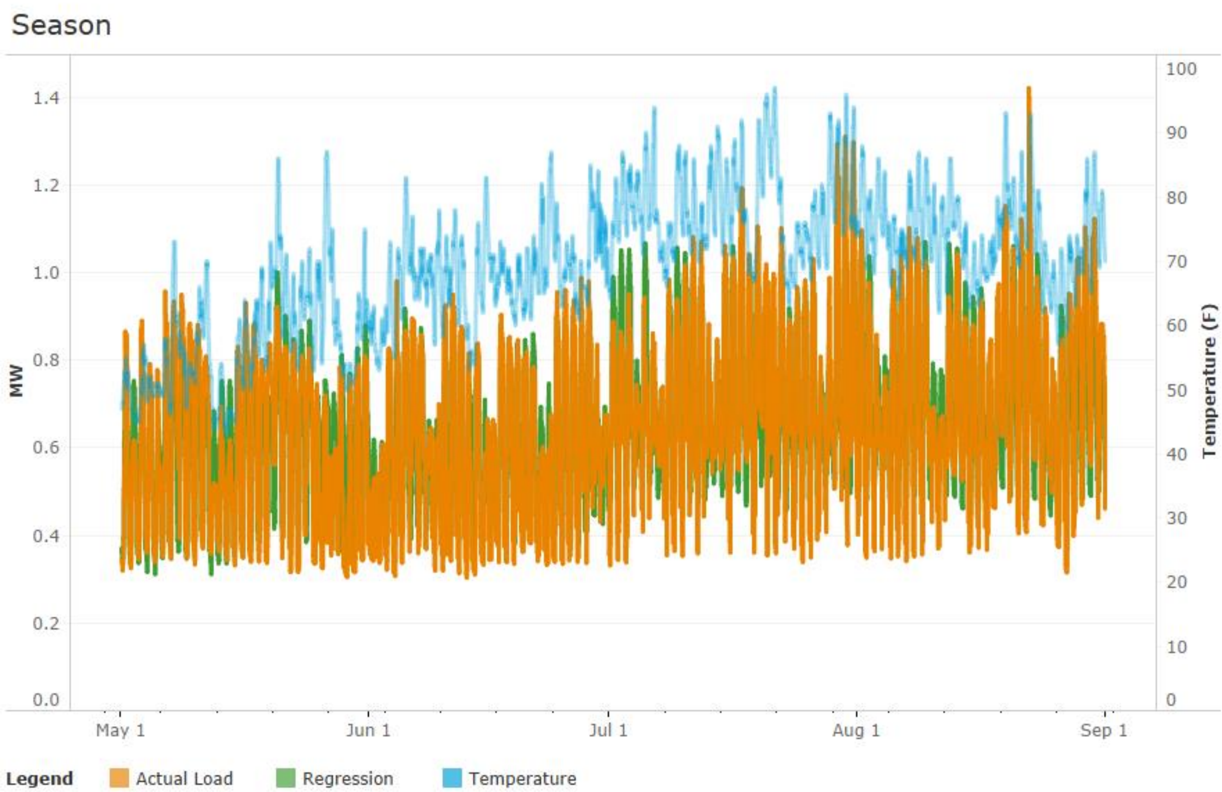


Figure 5-18. Example Customer 2 - Weather-Correlated Load on ICAP event day shows the ICAP event day for this weather-correlated customer. While on most days, the regression tracks this customer's load well, for the three days this week, three of the hottest weekdays of the summer, the customer's load was well beyond the regression's prediction. As a result, the load reduction is reduced by approximately a third. The regression is equally low on the hot days on either side of the event day so there is no evidence event day gaming. These are simply hot days for which the regression does a poor job of estimating the load.

Figure 5-18. Example Customer 2 - Weather-Correlated Load on ICAP event day

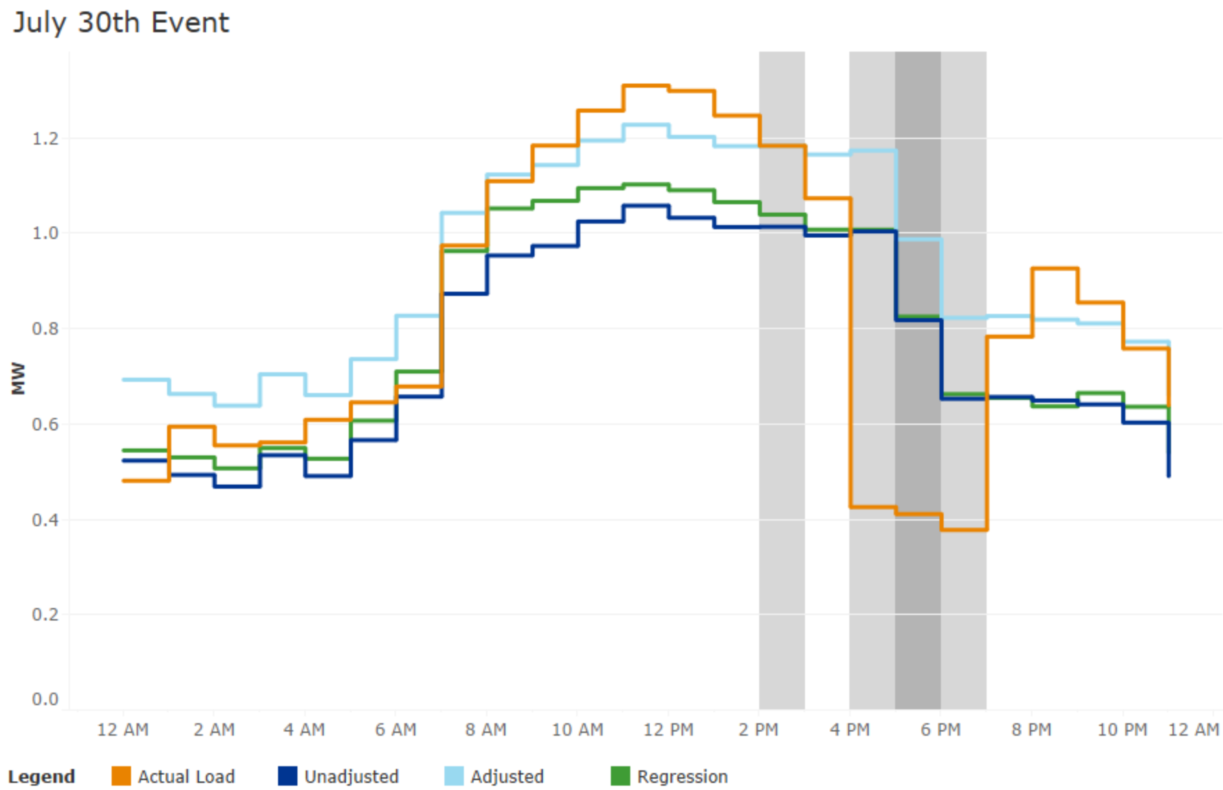


Figure 5-19. shows the event day surrounded by the two other hot days. The relationship between the regression estimates and actual load are not appreciably different across the three days. This is evidence that the regression is simply failing to provide good estimates on these particularly hot days. The adjust 10 of 10 baseline also provides evidence that no gaming is occurring in the actual for this customer. The adjustment hour loads (where actual and adjusted overlap) for the first two days are at exactly the same level. This indicates that this customer's pre-event load was identical on the event day as the day before when there was no event scheduled.

Figure 5-19. Example Customer 2 - Weather-Correlated Load on Days around ICAP Day

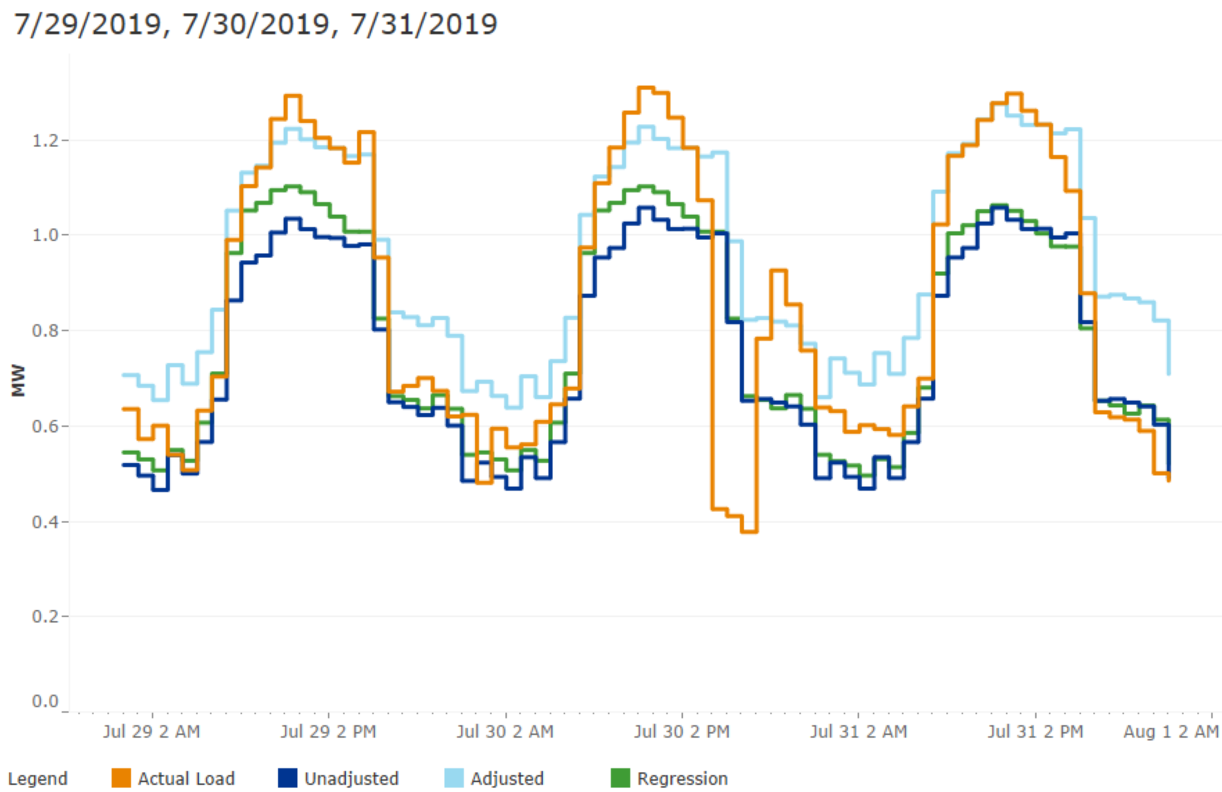
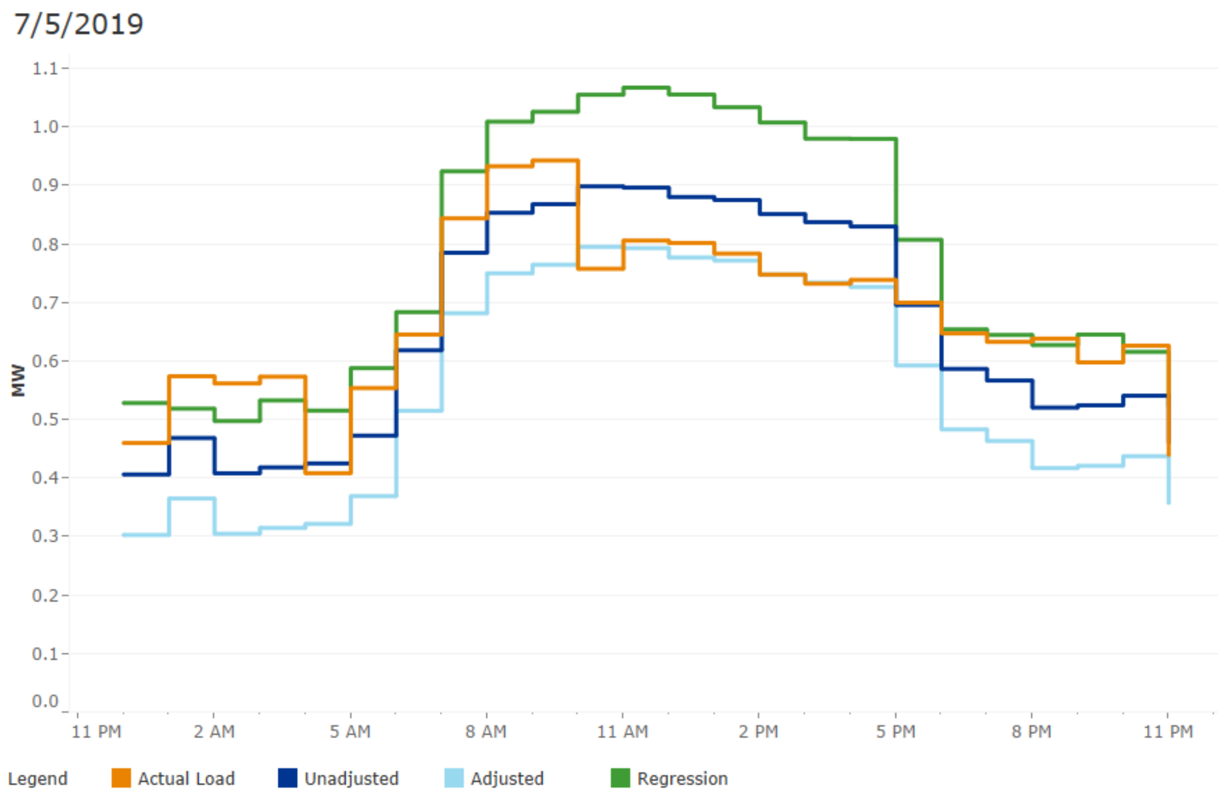


Figure 5-20. demonstrates what happens during a shutdown day. This customer did not shut down completely on the Friday after July 4th, but load was well below expected in the afternoon. In this instance, the regression baseline is well above the unadjusted baseline and both are well above actual load in the afternoon. The symmetrically adjusted baseline is adjusted down to the level of actual load. The asymmetric baseline for this customer and day would be the unadjusted baseline because the adjustment to actual load is downward. This provides an illustration of how the upward bias associated with the asymmetric baseline occurs. The substantial upward adjustment required for the days in the prior figure should be approximately equaled out by the downward adjustments on days like July 5th. However, the asymmetric baseline avoids that negative adjustment.

Figure 5-20. Example Customer 2 - Weather-Correlated Load on July 5th, 2019



5.1 Figure 5-21. provides an aggregate plot of the ADR initiative in Massachusetts on July 30th. This plot illustrates that the customer level choice of the higher of the two baselines – unadjusted and adjusted – leads to an asymmetrical baseline that is higher than the symmetrically-adjusted baseline. In this plot, the asymmetrical baseline is higher than actual load in every hour. The symmetrically adjusted baseline has the potential to be biased up or down based on the net pre-event load adjusting that might be occurring. The extent and direction of the potential bias of the symmetrically-adjusted baseline bias is an important consideration addressed at length in Section 4.7.3, Process Evaluation Methodology

For the process evaluation, the evaluation team employed the following data collection and analysis activities.

5.1.1 PA Staff Interviews and Documentation Review

The evaluation team reviewed the initiative documents and data as well as any pertinent information obtained from the PA's websites to inform the development of data collection instruments and interpretation of findings. Following this review, the team conducted over-the-phone in-depth interviews in December and January with the following:

- One Eversource initiative implementation staff member familiar with initiative administration in all three states
- One National Grid initiative implementation staff member familiar with initiative administration in Massachusetts
- One Unitil initiative implementation staff member familiar with initiative administration in Massachusetts and New Hampshire
- One ISO-NE staff member involved with the ISO's Price Responsive Demand initiative
- All four Curtailment Service Providers (CSPs) who are approved to execute customer participation in the PA's DR initiatives

The team interviewed these stakeholders to investigate the following topics:

- Overall goals of the DR initiative and lessons learned
- Barriers to implementation and potential areas for improvement
- Overlap between PA ADR events and ISO-NE Forward Capacity Market (FCM) and how the overlap can impact PA ADR initiative performance
- DR behaviors or actions taken
- Satisfaction with the CSPs
- Other opportunities for peak demand management

5.1.2 Participant Survey

The team conducted a mixed-mode (online-phone) participant survey in November and December of 2019. Tables 3-1 and 3-2 provide an overview of each participant stratum. The goal was to achieve 90% confidence and 10% relative precision overall and 85% confidence and 15% precision in all PA strata except the "Unitil" stratum, in which there were only seven participants. Expected precision is based on a 0.5 coefficient of variance.

Our goal was to achieve as many survey completions by state as possible and as such contacted all participants. Note that participant populations in Connecticut and New Hampshire were very small and, thus, the survey samples for those two states are equally small. To optimize survey response among groups with small populations, the team contacted participants multiple times (making up to five attempts to reach non-responding participants) through two modes (e-mail and phone). The response rates ranged from 22% to 43% by PA territory and 24% to 38% by state.

As shown in Table 3-1 and Table 3-2, participant response rates, overall and by group, were less than 50%, indicating a possibility of nonresponse bias. Nonresponse bias is introduced when respondents differ in a significant way from non-respondents. Although the team could not test for this bias due to lack of non-respondent data, the bias is still a concern considering response rates were quite low for certain PAs and state-level groups. Thus, the reported survey findings should be interpreted with caution.

Table 3-1. Participant Survey Response Counts by PA

PA	Population/ Sample Frame (Organizations)[a]	Survey Completes	Response Rates	Confidence/ Precision
Eversource – non-battery	72	21	29%	90/15
Eversource – battery	2	2	100%	N/A
National Grid	147[b]	32	22%	90/14
Unitil	7	3	38%	N/A
Total	**228**	**58**	**N/A**	**90/10**

[a] Some organizations had multiple participating locations in a PA territory. To manage survey length and respondent survey fatigue, the team did not ask those overseeing multiple locations to report on satisfaction, typical curtailment actions, and other aspects of the DR initiative by location. Thus, the responses from those overseeing multiple locations represent overall participation experience rather than location-specific participation experience.

[b] The evaluation team received participant lists from all but one National Grid CSP. The combined list of 147 likely includes most of National Grid's participants but not all. We refer to this list as the sample frame. A sample frame denotes a list of those in the population who can be sampled.

Note, the "Survey Completes" in Table 3-2, below, have higher totals than "Survey Completes" in the prior table because several respondents had participating sites in multiple states and reported that their responses in the survey reflect their experience across multiple states. For those who said that their experience was the same across their sites in multiple states, to ensure that their survey responses reflected their experience in all the states in which they had a participating site, the evaluation team duplicated their survey record. For example, one retail respondent had participating sites in all three states and reported have the same experience across all states (that is, they stated that their survey responses reflect their experience in all three states). The team then made copies of that respondent's survey record and attributed one record to Massachusetts, another to Connecticut and the final record to New Hampshire. This ensured that the data set (used to generate results by state) reflected that participant's responses in each state.

Table 3-2. Participant Survey Response Counts by State[1]

State	Population / Sample Frame			Survey Completes by State	Response Rates by State
	Eversource	National Grid	Unitil		
Massachusetts	56	147	3	49[a]	24%
Connecticut	20	–	–	7[b]	35%
New Hampshire	9	–	4	5[c]	38%

[1] Includes customers with sites in multiple states.

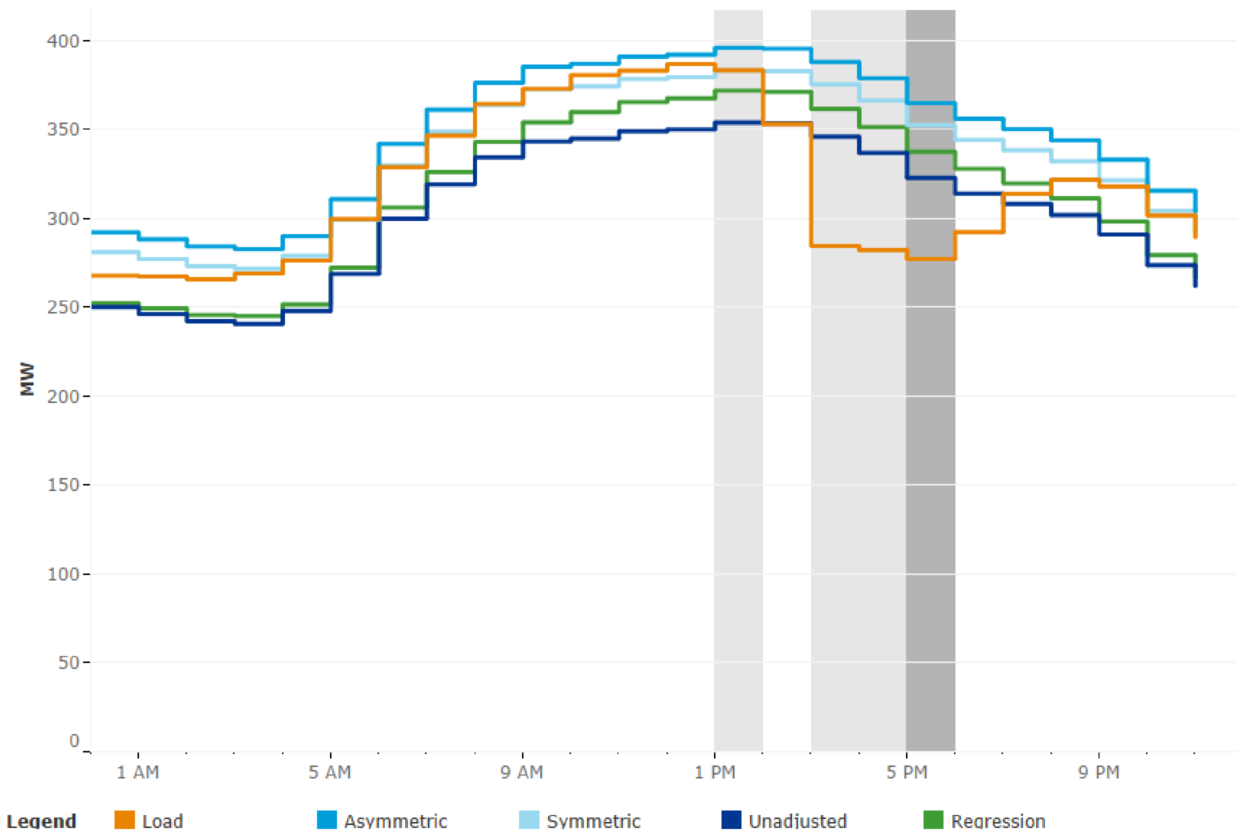
[a] Among 49 Massachusetts survey respondents, 29 were National Grid, 15 were Eversource, three were both National Grid and Eversource customers, and two were Unitil customers.

[b] All seven Connecticut respondents were Eversource customers.

[c] Among five New Hampshire respondents, four were Eversource and one was a Unitil customer.

Impact Evaluation Methodology. This is different that the systematic upward bias caused by the asymmetrical adjustment.

Figure 5-21. ADR Initiative – Massachusetts - July 30th ICAP Day Event



Appendix F: Eversource Extraordinary Weekend Event

Demand response events for the ADR initiative are only called on weekdays, however, the ICAP hour was forecasted on Saturday, July 20th, 2019 due to peak load conditions. In response, Eversource called an extraordinary, voluntary event. Event performance for the weekend event does not factor into the average load reduction that settlement is based upon.

The rolling baselines for Saturdays are constructed in the same way as for weekdays, as a rolling average of load during each hour during the most recent similar days. The characteristics that make these rolling baselines different for Saturdays is that only non-holiday Saturdays are considered to be similar days and only the 5 most recent eligible days are used to construct these 5-of-5 rolling baselines.

Table 5-6 summarizes load reduction for the weekend event across each state with Eversource customers. Note, no reported results were provided for the weekend event. Without reported results, there is no reason to show validated results. Similarly, the enrolled capacities are not applicable for weekend events. All other evaluated results are shown.

The evaluated weekend results maintain the general pattern seen for the weekday results. The asymmetric baseline estimates the most amount of load reduction. The symmetrically adjusted baseline estimates load reduction less than the asymmetric baseline, due to the application of downward baseline adjustments. The unadjusted baseline estimates the least amount of load reduction due to no day-of-event adjustment. The regression estimates load reduction between the magnitude of the unadjusted and symmetrically adjusted baselines, likely due to a mixture of weather-correlated loads and non-weather-correlated loads.

Table 5-6. Eversource - Weekend Event Impact Summary

Result	Weekend Event Average Reduction (kW)	Weekend Event Average Reduction (kW)	Weekend Event Average Reduction (kW)
State	**MA**	**NH**	**CT**
Evaluated - Asymmetric	6,610	2,083	5,130
Evaluated - Unadjusted	(16,123)	(207)	(323)
Evaluated - Symmetric	4,454	1,213	1,261
Evaluated - Forecast	N/A	N/A	N/A
Evaluated - Regression	(5,416)	1,743	1,760
Accounts	**153**	**40**	**95**

Figure 5-22., Figure 5-23., and Figure 5-24. show visual representations of the July 20th weekend event for Massachusetts, New Hampshire, and Connecticut, respectively. Each of these figures

show actual load (orange line), the unadjusted 5-of-5 baseline (dark blue line), the symmetrically adjusted baseline (light blue line), the asymmetrically adjusted baseline (cyan line), and the regression baseline (green line). The adjustment window and event period are shaded in gray.

As expected, actual load and baseline load is significantly less for each group than during the weekdays. Reduced load within these groups likely means that even if the event were not voluntary, the achievable load reduction is significantly less.

Figure 5-22. Eversource - Massachusetts - July 20th Weekend Event

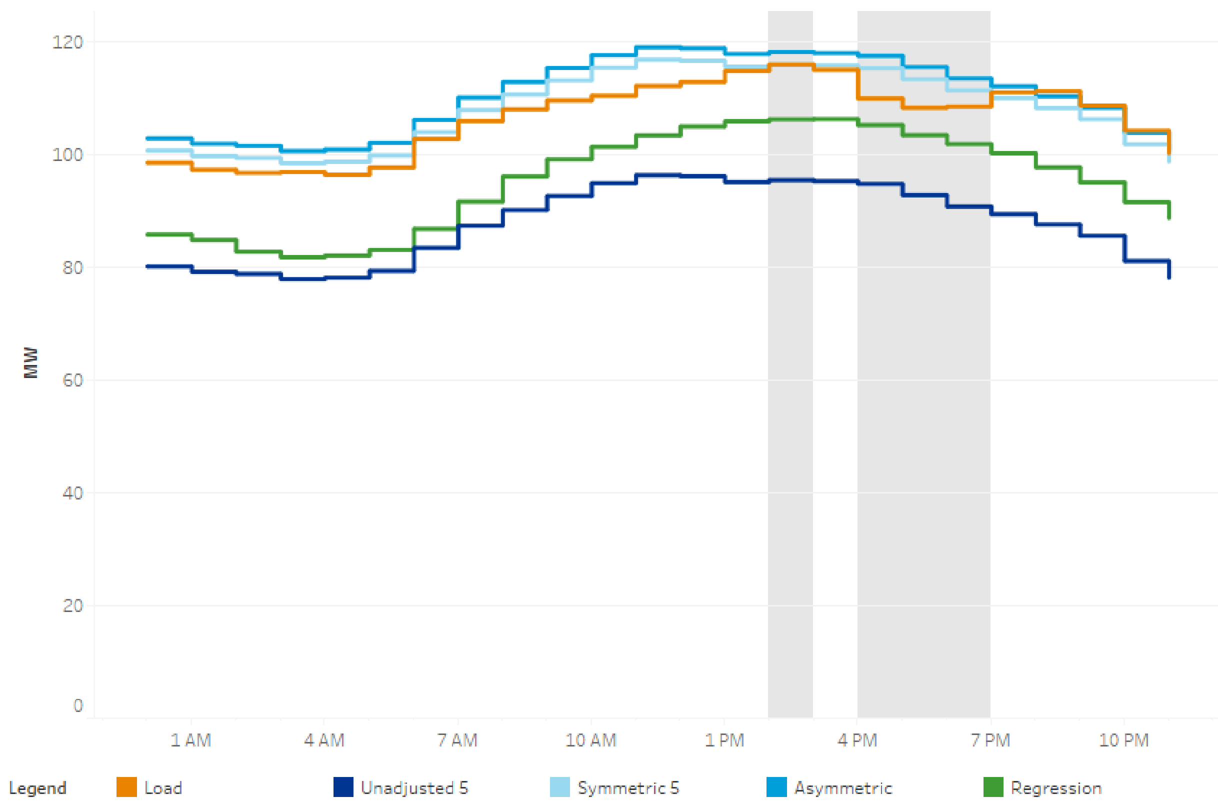


Figure 5-23. Eversource - New Hampshire - July 20th Weekend Event

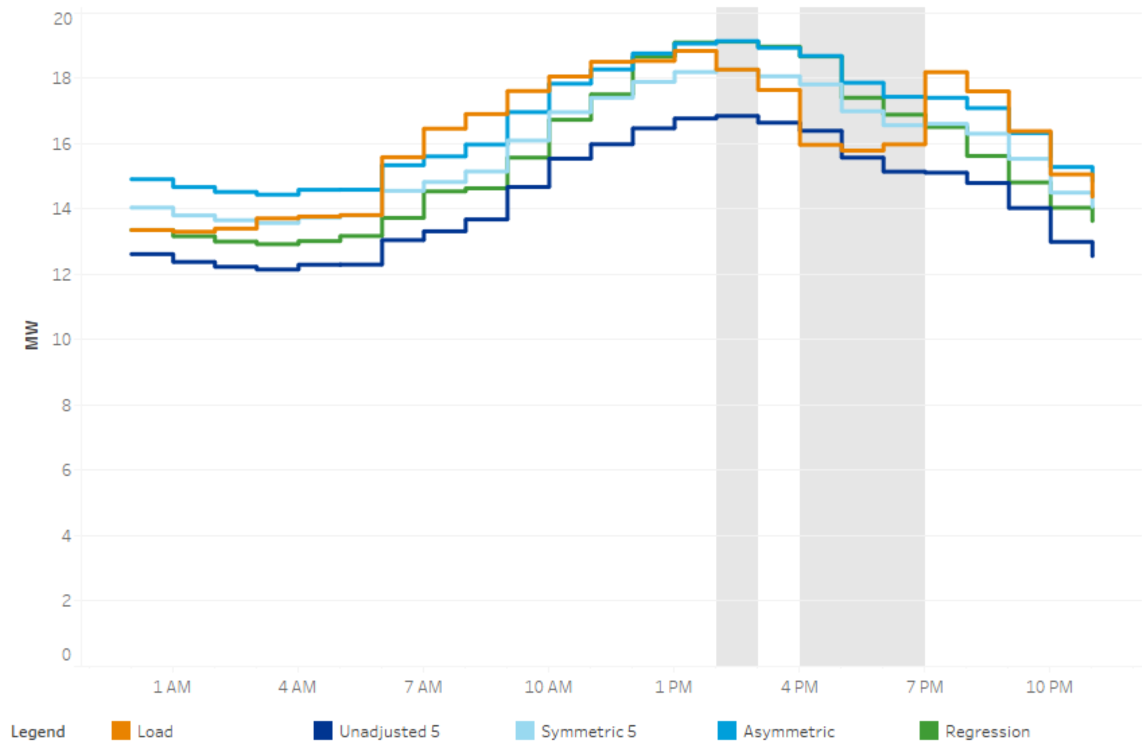


Figure 5-24. Eversource - Connecticut - July 20th Weekend Event