PSE PUGET SOUND ENERGY

Evaluation Report Response

Program: Demand Response Pilot

Program Manager: Thomas Smith

Study Report Name: Final Report 2022-2023 Demand Response Pilot Program Evaluation

Draft Report Date: October 9, 2023

Evaluation Analyst: Kasey Curtis

Date of Final Report Provided to Program Manager: December 15, 2023

Date of Program Manager Response: January 20, 2024

Overview

Puget Sound Energy (PSE) has projected shortfalls in infrastructure capabilities for transporting electricity to Bainbridge Island and gas to Duvall during peak winter hours. PSE started demand response (DR) programs in these geographies in 2022 as their first demand management effort to reduce peak hour load and delay or reduce need for infrastructure upgrades. The controlled technologies for this first year were smart thermostat (space heating) and water heating. Participants who enrolled in the programs, which were called 'Peak Energy Rewards,' receive a $75 annual incentive.

In winter 2022-2023, PSE called six events for Duvall and seven events for Bainbridge. By the end of the season, 74 customers were enrolled in the smart thermostat program in Duvall, and 65 in Bainbridge. Thermostat brands included in the program were Ecobee, Nest, and Mysa. Only one customer enrolled in water heater DR. DNV evaluated this program for winter 2022-2023 and plans to evaluate the program again after winter 2023-2024.

Evaluation

The evaluation objectives across two years of research are:

1. Provide PSE with an independent estimate of achieved demand reduction during called DR events during the 2022-23 and 2023-24 winter peaking seasons.
2. Provide PSE with a recommended settlement method to use for future winter peaking seasons based on a comparison of potential settlement methods and DNV's estimate of achieved demand reduction.
3. Provide PSE with recommendations to achieve DR goals in future peaking seasons.
4. Provide PSE with insights on demand impacts outside of called event periods, such as snapback or pre-heating impacts.

5. Provide PSE with an independent assessment of program operations and identify opportunities for program improvement based on material review, staff interviews, and customer surveys.
6. Assess the integration of the DR program with PSE grid operations.

Objectives 1 and 4 are achieved as the result of our impact evaluation and Objectives 5 and 6 are achieved through the process evaluation, which includes customer surveys, staff interviews, and program document review. Objective 3 is based on the combination of all evaluation research. Objective 2 will be addressed in the 2024 evaluation study.

The impact evaluation relied on a matched comparison group and a difference-in-difference methodology. The details of these methodologies are described in the body of the report. This approach essentially creates a pseudo-control group (as in a randomized control trial) that allows the best evaluation of event days that are often colder, and therefore not comparable, to non-event days. This analysis includes impact results on an hourly basis and explores several potential determinants of variation in impact magnitude. Impacts are also compared to Autogrid's measurement and verification (M&V) estimates. Water heaters were excluded from this year's analysis because only one customer enrolled a water heater.

The process evaluation included four interviews with PSE staff from four different departments, a survey of customers participating in the DR program, and a review of program materials. DNV sent surveys to all 101 2022-2023 participants and received completed surveys from 35 participants.

Key Findings, Recommendations, and PSE Response

This section provides key findings and recommendations resulting from DNV's evaluation, along with PSE's response to the recommendations. For the sake of organization, each finding, its associated recommendation, and PSE response are numbered below.

Finding 1:

Key Finding – Curtailment varied by smart thermostat manufacturer, with Nest having the highest estimated curtailment, Ecobee having much less curtailment, and no significant event impact for Mysa thermostats.

Recommendation -- PSE may want to consider both working with the thermostat manufacturers to see if they are able to increase curtailment. If thermostat manufacturers are not able to increase curtailment, PSE could restrict enrollment or reduce incentives for those thermostats that provide less or no curtailment. Next year's evaluation should provide additional evidence for the difference in thermostat brand effectiveness in curtailment.

PSE Response

> *Thermostat manufacturers (OEMs) standardize DR program parameters nationally and do not allow utilities to override their internal policies. OEMs also have a wide variance of curtailment capabilities compounded on by the fact that most OEMs allow customers to set their own curtailment levels, the exception being Nest. This explains the more consistent curtailment results from Nest customers, as other OEMs allow customers to set a wide range of pre-conditioning and set-point offset options that utilities cannot influence.*

PSE intends to maintain a single smart thermostat incentive to ensure customers have an equitable opportunity for the incentive without creating the perception that the utility is favoring one brand over another, and to avoid unequally rewarding participants.

Finding 2:

Key Finding – DNV found that most program participants already had smart thermostats when they enrolled.

Recommendation – PSE may want to consider actively recruiting customers who do not yet have smart thermostats and educating them on the benefits of these devices. Additionally, PSE could offer an alternative program for customers who do not want to have a smart thermostat. This program, like DR programs before smart device technology, could rely on switches installed on heating equipment that would reduce the equipment's runtime during called events.

PSE Response

PSE's marketing team takes an active approach to customer segmentation, providing customers targeted marketing depending on if they have already received a smart thermostat or other qualified equipment rebates. If customers have not received a smart thermostat rebate from PSE, they will be engaged by a different marketing approach for enrollment and participation than those who have already purchased a smart thermostat.

To ensure a wide range of participation options, PSE has launched two new Behavioral Demand Response programs as of Q4 of 2023 to offer customers multiple ways to participate in DR in the event they do not want to enroll their device in an automated control program.

Finding 3:

Key Finding – Most program participants also reported having a variety of other devices that could be controlled as part of DR programs to reduce electric and gas load for Bainbridge and Duvall, respectively.

Recommendation – PSE could consider more active recruitment for the water heater device part of the program, and expanding its programs so that other controllable devices could enroll.

PSE Response

PSE is actively investigating additional control points for water heaters and ductless heat pumps, and are working with industry leaders to identify the potential for connected control expansion in 2024.

Finding 4:

Key Finding – Many program participants suggested that pre-heating was not working well for them, causing increased overrides.

Recommendation – PSE may want to consider longer pre-heat periods that would work better for heat pumps and pre-heating that works for customers with a nighttime setback. DNV can also ask more targeted pre-heating questions in the customer survey next year to provide more detailed recommendations on pre-heating.

PSE Response

PSE is working in consultation with its DR aggregation vendor to identify adequate pre-heating configurations that strike a balance between desired load flexibility goals and customer comfort, with an emphasis on a positive customer experience. As PSE produces more event results, a more consistent and beneficial pre-conditioning policy for Summer and Winter DR seasons may be established.

Finding 5:

Key Finding – PSE staff and customer surveys suggested that not all gas DR program participants have gas heating systems.

Recommendation – Program managers could consider a simple validation algorithm of heating equipment, using AMI data.

PSE Response

PSE has begun development on a platform that will establish a series of checks and balances, leveraging data insights to better ensure customers attempting to enroll in a DR program are placed in the appropriate customer treatment. PSE plans to have this functionality executed and integrated into its automated enrollment and validation system in 2024.

Finding 6:

Key Finding – Most participants reported finding out about the program in an email, while program staff reported putting a lot of effort into community outreach and a high level of community engagement.

Recommendation – While the customer survey after the winter 2023-2024 program season will shed some light on a greater number of participants over more time, staff should ensure that their efforts are translating into program sign-ups.

PSE Response

PSE Program and Marketing teams work closely together to monitor enrollment activity during pre-and-post marketing pushes to assess marketing effectiveness. To date, PSE notes a 1:1 relationship with marketing efforts and increased enrollments within a 3-5 day window of the email being sent. With the launch of PSE's system-wide DR programs, PSE has gained the ability to market more widely, and with additional channels not previously available, and looks forward to seeing increased enrollment numbers with more sustained and broad-spectrum marketing.

Finding 7:

Key Finding – Most program participants report high satisfaction and program staff describe putting a lot of effort into enrollment materials to make enrollment and participation as smooth as simple as possible for all types of participants.

Recommendation – PSE should continue its excellent customer service.

PSE Response

Noted and accepted.

Finding 8:

Key Finding – Some customers reported that they would prefer earlier program notifications, with all notifications received at least the day before the event. Some even said they would prefer more false notifications if notifications could be earlier.

Recommendation – PSE should continue to study this preference in the customer survey after the next program season and send notifications as early as possible.

PSE Response

Similarly to pre-conditioning and curtailment options, thermostat OEMs do not allow utilities to set advance notification times except in very rare instances. PSE has the ability to provide day-ahead (or more) notifications for Behavioral Demand Response programs and notes this recommendation, but Automated Demand Response program event notifications are set at the purview of the OEM and cannot be altered.

Finding 9:

Key Finding – In staff interviews, DNV heard that VPP staffing organization is rapidly developing and has many stakeholders, leading to decentralized leadership and lack of staffing continuity. This affects all VPP stakeholders, including grid operations, and problematic communication is most apparent with the gas operations group, who seemed unaware of VPP operational plans.

Recommendation – PSE may want to centralize VPP leadership, reduce staff turnover, and set up regular communications with grid and gas operations staff.

PSE Response

Noted and accepted.

Finding 10:

Key Finding – PSE grid operations staff are generally up to date with program understanding and are working towards a standard practice for determining when events should be called.

Recommendation – PSE grid operations staff should continue their engagement with VPP efforts.

PSE Response

Noted and accepted.

Finding 11:

Key Finding – M&V is a consistent area of frustration for VPP platform staff, program managers, grid operators, and the gas operations group. M&V results are hard to find, not integrated in Autogrid's Flex system, and the best methods for M&V are unclear. Neither the grid nor gas operations team trusts Autogrid's M&V results so have spent considerable time and effort producing their own results.

Recommendation – PSE may want to consider a system for rapid AMI transfer to a third-party evaluator to ensure accurate, rapid evaluation of event impacts.

PSE Response

> *We acknowledge the recommendation and are currently investigating approaches that might provide the kind of analysis the evaluator recommends. For the time being every demand response season will be undergo evaluation by an independent third party to continue to provide PSE information needed to assess the accuracy of M&V platforms and determine ways to improve.*

DEMAND RESPONSE

Final Report
2022-2023 Demand Response Pilot Program Evaluation

Puget Sound Energy

Date: October 9, 2023

Primary Contact:
Chelsea Liddell, Data Scientist/Project Manager, Markets and Risk
DNV Energy
chelsea.liddell@dnv.com

Project Sponsor
Geoff Barker, Senior Consultant, Markets and Risk
DNV Energy
Geoff.Barker@dnv.com

PSE Contract Manager
Kasey Curtis, Senior Market Analyst, Evaluation and Research
Puget Sound Energy
Kasey.Curtis@pse.com

Table of contents

List of tables

List of figures

1 EXECUTIVE SUMMARY

1.1 Background and approach

Puget Sound Energy (PSE) has projected shortfalls in infrastructure capabilities for transporting electricity to Bainbridge Island and gas to Duvall during peak winter hours. PSE started demand response (DR) programs in these geographies in 2022 as their first demand management effort to reduce peak hour load and delay or reduce need for infrastructure upgrades. The controlled technologies for this first year were smart thermostat (space heating) and water heating. Participants who enrolled in the programs, which were called 'Peak Energy Rewards,' receive a $75 annual incentive.

In winter 2022-2023, PSE called six events for Duvall and seven events for Bainbridge. By the end of the season, 74 customers were enrolled in the smart thermostat program in Duvall, and 65 in Bainbridge. Thermostat brands included in the program were Ecobee, Nest, and Mysa. Only one customer enrolled in water heater DR. DNV evaluated this program for winter 2022-2023 and plans to evaluate the program again after winter 2023-2024.

The evaluation objectives across two years of research are:

1. Provide PSE with an independent estimate of achieved demand reduction during called DR events during the 2022-23 and 2023-24 winter peaking seasons.
2. Provide PSE with a recommended settlement method to use for future winter peaking seasons based on a comparison of potential settlement methods and DNV's estimate of achieved demand reduction.
3. Provide PSE with recommendations to achieve DR goals in future peaking seasons.
4. Provide PSE with insights on demand impacts outside of called event periods, such as snapback or pre-heating impacts.
5. Provide PSE with an independent assessment of program operations and identify opportunities for program improvement based on material review, staff interviews, and customer surveys.
6. Assess the integration of the DR program with PSE grid operations.

Objectives 1 and 4 are achieved as the result of our impact evaluation and Objectives 5 and 6 are achieved through the process evaluation, which includes customer surveys, staff interviews, and program document review. Objective 3 is based on the combination of all evaluation research. Objective 2 will be addressed in the 2024 evaluation study.

The impact evaluation relied on a matched comparison group and a difference-in-difference methodology. The details of these methodologies are described in the body of the report. This approach essentially creates a pseudo-control group (as in a randomized control trial) that allows the best evaluation of event days that are often colder, and therefore not comparable, to non-event days. This analysis includes impact results on an hourly basis and explores several potential determinants of variation in impact magnitude. Impacts are also compared to Autogrid's measurement and verification (M&V) estimates. Water heaters were excluded from this year's analysis because only one customer enrolled a water heater.

The process evaluation included four interviews with PSE staff from four different departments, a survey of customers participating in the DR program, and a review of program materials. DNV sent surveys to all 101 2022-2023 participants and received completed surveys from 35 participants.

1.2 Evaluation results

1.2.1 Impact evaluation

We estimate that the average participant load curtailment achieved by the program to date is 1.59 kW for electric participants and 0.106 ccf/hr for gas participants across all events. We also observed significant amounts of pre-heating in the hour before an event was called, as well as significant "snapback" heating in the two hours after an event.

- Pre-heating accounted for an average load increase of 0.73 kW per customer for electric participants and 0.079 ccf/hr per customer for gas participants in the hour before the event.
- Snapback heating accounted for an average load increase of 0.84 kW per customer for electric participants and 0.089 ccf/hr per customer for gas participants in the two hours after the event.

Based on the per-customer average curtailment from each event and the number of customers enrolled, we estimate that the winter season events were responsible for an event average total curtailment of 96 kW per event for electric events and 7.3 ccf/hr per event for gas events.

Figure 1-1 and Figure 1-2 show the average per-customer load curtailment by hour into an event, averaged across all events (except the 2/24 event). The x-axis in both figures shows relative hours to an event. Negative hours are the hours before an event. Hours 1-3, shown as green bars, are the event hours. The remaining positive hours are hours following an event. We use these data visualizations to compare the nth hour of an event between events that have different start times and/or different lengths. Positive values in these plots correspond to load curtailment, while negative values correspond to increased load. As described above, there is increased load in the hour preceding an event (pre-heating) and the increased load in the hours following an event (snapback). We see that the timing of the event relative to the system peak is particularly important, because calling an event just an hour after peak could cause an increase of load during that peak hour.

Figure 1-1. Load curtailment by hour into event, electric participants

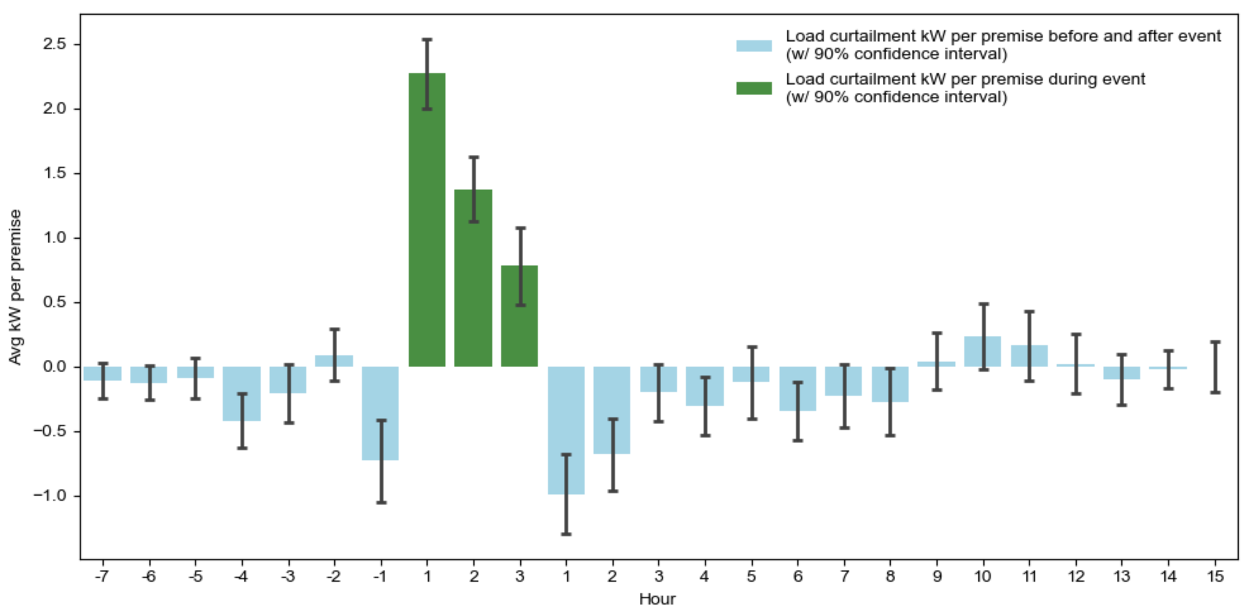
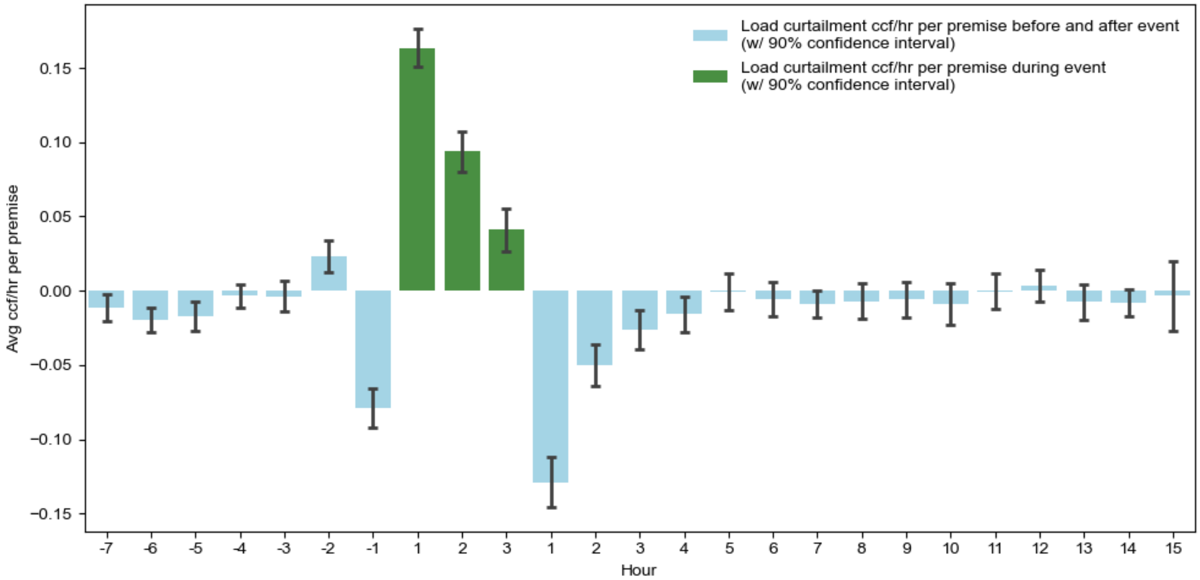
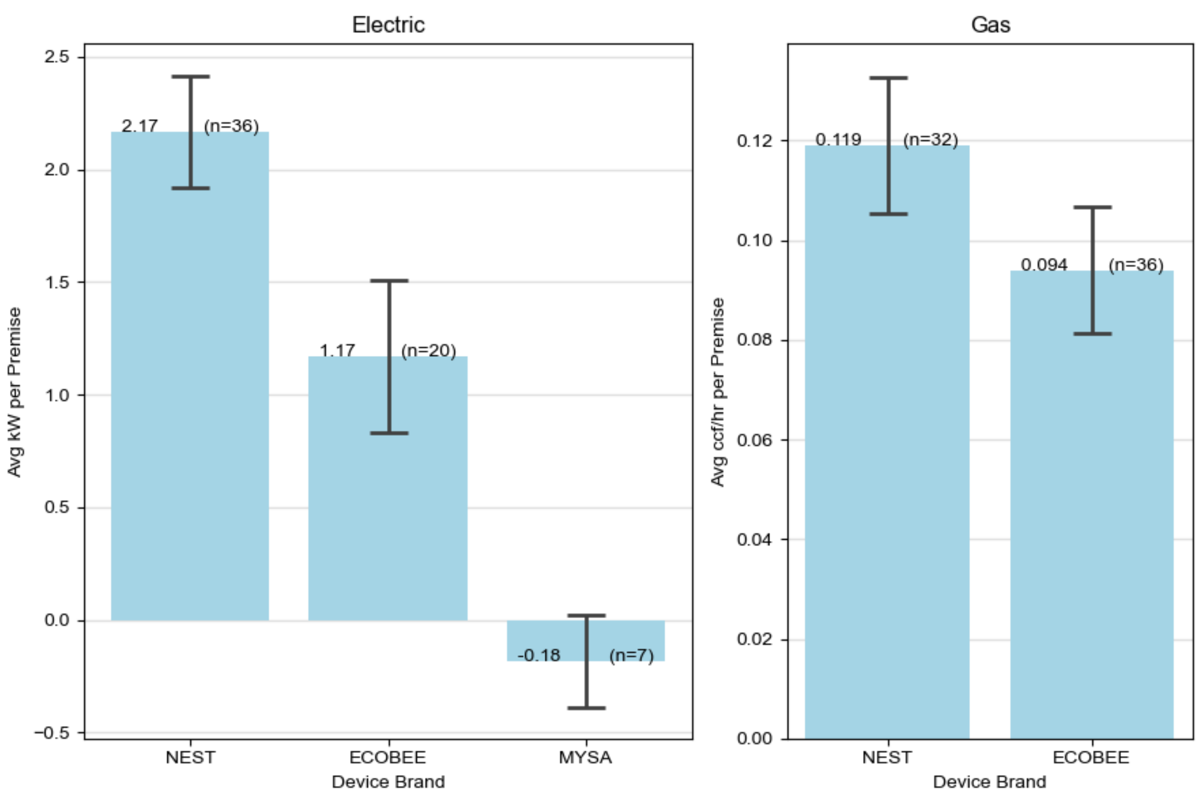


Figure 1-2. Load curtailment by hour into event, gas participants



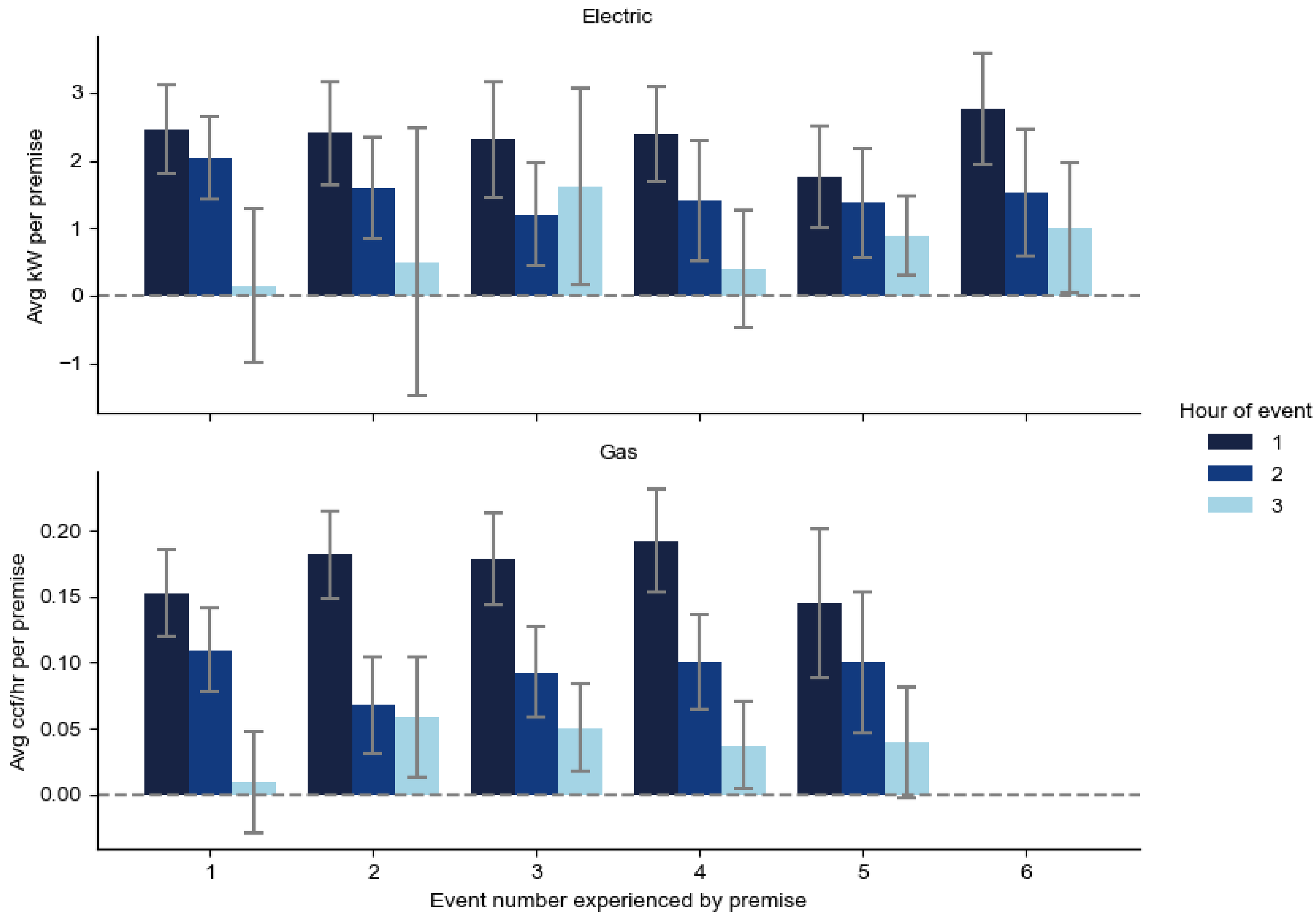
Our analysis also revealed that curtailment differed by device brand. Nest thermostats are associated with the most curtailment, with Ecobee thermostats demonstrating less curtailment, and Mysa thermostats having no significant curtailment.

Figure 1-3. Curtailment estimates by device brand



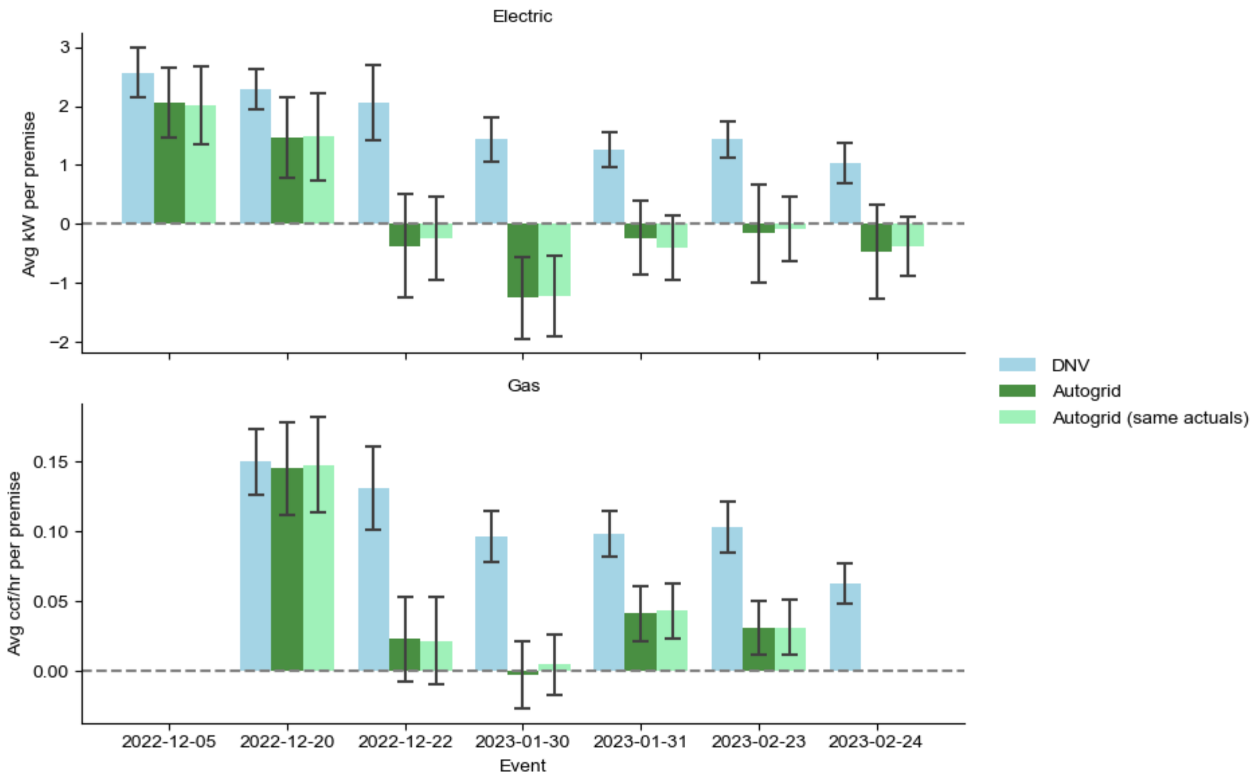
Impacts do not seem to have obvious trend over the course of the winter season, though three-hour events do have lower curtailment than two-hour events.

Figure 1-4. Curtailment by hour and event number experienced



DNV also compared our results to the M&V estimates provided by Autogrid. Autogrid's estimates are consistently lower than DNV's validated curtailment values, and are often negative, implying load growth during events. The data DNV received from Autogrid also indicated inconsistencies between actual usage from PSE's advanced metering infrastructure (AMI) data and the values Autogrid presented for actual usage. In Figure 1-5, the dark bars show Autogrid's M&V estimates, and the light bars show estimates from Autogrid once participants with data issues were removed.

Figure 1-5. Comparison of DNV and Autogrid curtailment estimates by event date



1.2.2 Process evaluation

DNV's process evaluation included a customer survey, PSE staff interviews, and document review.

1.2.2.1 Customer survey

The customer survey provided insights on which PSE customers are participating in the program, how to grow the program and its impact, how to learn more about program impacts, and how to increase customer satisfaction. The overall number of customers who replied to the survey was high at 35 out of a program population of 101 (a 35% response rate). Respondents differed from the overall participant population in the brands of thermostats, so there may be some respondent bias in the results. However, to the degree that survey respondents represent the population of program participants, we can make some conclusions on how program participants differ from the general (US Census based) populations in Bainbridge and Duvall. In general, program participants appear to be more educated, live in newer homes, and include a higher proportion of Asian customers than the overall population. The survey provided a variety of insights on how the program could be improved. Ways that the program could grow include:

- Recruiting customers who do not have smart thermostats, as most participants do before participating.
- Expanding DR programs for other controllable technologies, which many participants own.
- Expanding recruiting channels, as 82% of customers heard about the program through email.
- Emphasizing incentives and environmental benefits in recruiting materials, as these are customers' top participation motivations.
- Ensuring no Duvall participants primarily heat their homes through electricity, as one in 12 Duvall participants indicated having electric heat only.

Overall, customers were very satisfied with the program, indicating high willingness to continue participating. However, respondents did have some recommendations for improvements. Specifically, respondents wanted event notifications no later than the day before and preferred to be notified via email and text. Respondents also would like pre-heating that is more customized to their heating systems and nightly heating setbacks.

1.2.2.2 Staff interviews

Staff interviews were extremely informative in understanding how members of different departments see the DR program and opportunities to increase the smooth deployment of the program within PSE going forward. From staff interviews, DNV collected insights on program organization and goals, the virtual power plant (VPP) platform[1], utility of Autogrid M&V, and how events are called. With regard to program organization, staff reported decentralized leadership and high staff turnover making building the VPP challenging, so clarifying roles and centralized leadership would likely both improve the VPP and all demand flexibility programs that rely on it. It may also lead to better communication with all PSE stakeholders; DNV found that some stakeholders lacked essential knowledge about program plans for next year. This was especially true for the gas operations team, who have important curtailment needs but lack coordinated planning with the team. More specific program goals, including MW and ccf/hr goals by year and by program, may help the team understand if it is on track to meet its overarching 2030 goal.

PSE's VPP platform effort is impressive, as are materials and systems established around the smart thermostat program. These are contributing to high customer satisfaction. However, Autogrid M&V has been frustrating for all groups at PSE that we spoke with. Results are challenging to find, are not integrated in Autogrid's Flex system, and are not reliable (see Figure 1-5 and discussion above for a comparison of Autogrid's curtailment estimates and DNV's estimates). Both electric and gas operations groups have created their own evaluation estimates, investing significant time and effort. PSE may want to create a dashboard to report rapid-turnaround impact results from a trusted third-party evaluator.

The PSE grid operations team is actively working toward protocols for calling events and should continue its planning with an eye toward customer stratification and the effects of pre-heating and snapback. Validation of customer fuel types would increase the overall curtailment for each event. While staff expressed concern about a time-of-use rate from an evaluation perspective, an opt-out time-of-use rate could create additional morning curtailment while addressing some equity concerns.

1.2.2.3 Documentation review

DNV reviewed a program process flow chart, program newsletters, and email outreach materials. Materials appeared to be well put-together, which contributed to the program's success.

1.3 Key evaluation findings and recommendations

This section provides key findings and recommendations resulting from DNV's evaluation. Additional findings are presented within each program-specific section.

1.3.1 Impact findings and recommendations

In this section, we describe findings from the quantitative evaluated curtailment from the program.

1) Evaluation Objective 1: Independent estimate of demand reduction

Finding: The average participant load curtailment achieved by the program to date is 1.59 kW for electric participants and 0.106 ccf/hr for gas participants across all events. Curtailment was highest in the first hour of events and declined in the second and third hours. The evaluated curtailment was quite different from that calculated by Autogrid using the 10-of-10

[1] The virtual power plant platform is set up to manage current and future PSE demand response programs. This program is the first one it has been used to manage.

settlement baseline. DNV's impact estimates were consistently positive, and always greater than 1 kW/participant for electric and 0.06 ccf/hr for gas. Autogrid's estimates were consistently smaller than DNVs and many were negative, indicating a load increase when the event was called.

2) Evaluation Objective 4: Demand impacts outside of called event periods

Finding: Participant load tends to increase in the hour before events and in the two hours after events. Pre-heating in the hour before the event accounted for an average load increase of 0.73 kW per customer for electric participants and 0.079 ccf/hr per customer for gas participants. Snapback heating in the two hours after the event accounted for an average load increase of 0.84 kW per customer for electric participants and 0.089 ccf/hr per customer for gas.

3) Evaluation Objective 3: Recommendations to achieve future DR goals

Finding: Curtailment varied by smart thermostat manufacturer, with Nest having the highest estimated curtailment, Ecobee having much less curtailment, and no significant event impact for Mysa thermostats.

- **Recommendation** – PSE may want to consider both working with the thermostat manufacturers to see if they are able to increase curtailment. If thermostat manufacturers are not able to increase curtailment, PSE could restrict enrollment or reduce incentives for those thermostats that provide less or no curtailment. Next year's evaluation should provide additional evidence for the difference in thermostat brand effectiveness in curtailment.

1.3.2 Process findings and recommendations

In this section, we discuss findings from staff interviews, customer surveys, and material reviews.

1) Evaluation Objective 3: Recommendations to achieve future DR goals

We have four key recommendations for PSE in this section.

Finding: DNV found that most program participants already had smart thermostats when they enrolled.

- **Recommendation** – PSE may want to consider actively recruiting customers who do not yet have smart thermostats and educating them on the benefits of these devices. Additionally, PSE could offer an alternative program for customers who do not want to have a smart thermostat. This program, like DR programs before smart device technology, could rely on switches installed on heating equipment that would reduce the equipment's runtime during called events.

Finding: Most program participants also reported having a variety of other devices that could be controlled as part of DR programs to reduce electric and gas load for Bainbridge and Duvall, respectively.

- **Recommendation** – PSE could consider more active recruitment for the water heater device part of the program, and expanding its programs so that other controllable devices could enroll.

Finding: Many program participants suggested that pre-heating was not working well for them, causing increased overrides.

- **Recommendation** – PSE may want to consider longer pre-heat periods that would work better for heat pumps and pre-heating that works for customers with a nighttime setback. DNV can also ask more targeted pre-heating questions in the customer survey next year to provide more detailed recommendations on pre-heating.

Finding: PSE staff and customer surveys suggested that not all gas DR program participants have gas heating systems.

- **Recommendation** –Program managers could consider a simple validation algorithm of heating equipment, using AMI data.

2) Evaluation Objective 5: Opportunities for program improvement

We have three key recommendations in this section.

Finding: Most participants reported finding out about the program in an email, while program staff reported putting a lot of effort into community outreach and a high level of community engagement.

- **Recommendation** – While the customer survey after the winter 2023-2024 program season will shed some light on a greater number of participants over more time, staff should ensure that their efforts are translating into program sign-ups.

Finding: Most program participants report high satisfaction and program staff describe putting a lot of effort into enrollment materials to make enrollment and participation as smooth as simple as possible for all types of participants.

- **Recommendation** – PSE should continue its excellent customer service.

Finding: Some customers reported that they would prefer earlier program notifications, with all notifications received at least the day before the event. Some even said they would prefer more false notifications if notifications could be earlier.

- **Recommendation** – PSE should continue to study this preference in the customer survey after the next program season and send notifications as early as possible.

3) Evaluation Objective 6: Assess integration of program with PSE grid operations

We have three final recommendations in this section.

Finding: In staff interviews, DNV heard that VPP staffing organization is rapidly developing and has many stakeholders, leading to decentralized leadership and lack of staffing continuity. This affects all VPP stakeholders, including grid operations, and problematic communication is most apparent with the gas operations group, who seemed unaware of VPP operational plans.

- **Recommendation** – PSE may want to centralize VPP leadership, reduce staff turnover, and set up regular communications with grid and gas operations staff.

Finding: PSE grid operations staff are generally up to date with program understanding and are working towards a standard practice for determining when events should be called.

- **Recommendation** – PSE grid operations staff should continue their engagement with VPP efforts.

Finding: M&V is a consistent area of frustration for VPP platform staff, program managers, grid operators, and the gas operations group. M&V results are hard to find, not integrated in Autogrid's Flex system, and the best methods for M&V are unclear. Neither the grid nor gas operations team trusts Autogrid's M&V results so have spent considerable time and effort producing their own results.

- **Recommendation** – PSE may want to consider a system for rapid AMI transfer to a third-party evaluator to ensure accurate, rapid evaluation of event impacts.

2 INTRODUCTION

2.1 Program overview

PSE has projected shortfalls in infrastructure capabilities for transporting electricity to Bainbridge Island and gas to Duvall during peak winter hours. Of the 6.6 MW projected shortfall on Bainbridge, PSE plans to lower peak demand about 3.3 MW with energy efficiency and demand management programs, with demand management programs responsible for about half of the 3.3 MW reduction. On Duvall, PSE plans to use gas curtailment during peak times to avoid installing additional gas pipelines. PSE started a DR program in these geographies in 2022 as their first demand management effort. The controlled technologies for this first year were smart thermostat (space heating) and water heating. PSE built its own internal VPP platform to dispatch these programs and future load shifting programs. The VPP platform handles enrollment, event calls, and incentive payments. The VPP includes an automated system to interface with the program implementer and thermostat manufacturers for the enrollment and event calls.

In winter 2022–2023, PSE called six events for Duvall and seven events for Bainbridge. By the end of the season, 74 customers were enrolled in the smart thermostat program in Duvall, and 65 in Bainbridge. Participants who enrolled in the programs, which were called 'Peak Energy Rewards,' receive a $75 annual incentive. See Table 2-1 for event specific enrollment and other details. Thermostat brands included were Ecobee, Nest, and Mysa. Only one customer enrolled in water heater DR. DNV evaluated this program for winter 2022-2023 and plans to evaluate the program again after winter 2023-2024.

Table 2-1. Event details

Fuel Type	Event Date	Event Length (Hours)	Event Start	Event End	# of Premises	Daily Min Temperature by Geography (F)
Gas	12/20/2022	2	7:00:00	8:59:00	54	26.1
	12/22/2022	2	7:00:00	8:59:00	54	16.0
	1/30/2023	3	7:00:00	9:59:00	71	26.1
	1/31/2023	3	7:00:00	9:59:00	71	32.0
	2/23/2023	3	7:00:00	9:59:00	74	27.0
	2/24/2023	3.5	6:30:00	9:59:00	74	21.9
Electric	12/5/2022	2	7:00:00	8:59:00	39	34.0
	12/20/2022	2	7:00:00	8:59:00	50	30.0
	12/22/2022	2	7:00:00	8:59:00	50	17.1
	1/30/2023	3	7:00:00	9:59:00	60	21.9
	1/31/2023	3	7:00:00	9:59:00	60	30.9
	2/23/2023	3	7:00:00	9:59:00	65	28.0
	2/24/2023	3.5	6:30:00	9:59:00	65	25.0

2.2 Evaluation objectives and researchable issues

The evaluation objectives across two years of research are:

1. Provide PSE with an independent estimate of achieved demand reduction during called DR events during the 2022-23 and 2023-24 winter peaking seasons.
2. Provide PSE with a recommended settlement method to use for future winter peaking seasons based on a comparison of potential settlement methods and DNV's estimate of achieved demand reduction.

3. Provide PSE with recommendations to achieve DR goals in future peaking seasons.
4. Provide PSE with insights on demand impacts outside of called event periods, such as snapback or pre-heating impacts.
5. Provide PSE with an independent assessment of program operations and identify opportunities for program improvement based on material review, staff interviews, and customer surveys.
6. Assess the integration of the DR program with PSE grid operations.

Objectives 1 and 4 are achieved as the result of our impact evaluation, which, in addition to providing during-event, pre-heating, and snapback (post-event) load impact estimates, also assesses impact per device type and persistence of load reduction (within and across events). We also compare evaluated impacts to Autogrid's impact estimates and explore reasons for differences. These insights are provided in *Section 4 Impact evaluation results*. Because there was only one water heater participant and only one consecutive-day instance that we could evaluate, we plan to include insights on those topics in next year's evaluation.

Objectives 5 and 6 are achieved through our process evaluation, which includes an assessment of program operations, opportunities for program review, and an examination of program integration with grid operations. Other insights gained from staff interviews, participant surveys, and material review are also included in the report. These results are found in *Section 5 Process evaluation results*.

Objective 3, recommendations for achieving goals in the future, is a result of the combination of impact and process evaluation activities. These results are found in *Section 6 Findings and Recommendations*.

Objective 2, recommending a settlement for future seasons, we plan to address in the 2024 report. In this report, we establish that the current settlement baseline underestimates program impacts and explore reasons for that issue.

3 EVALUATION APPROACH AND METHODOLOGY

In this section, we describe the methods used for the impact and process evaluations.

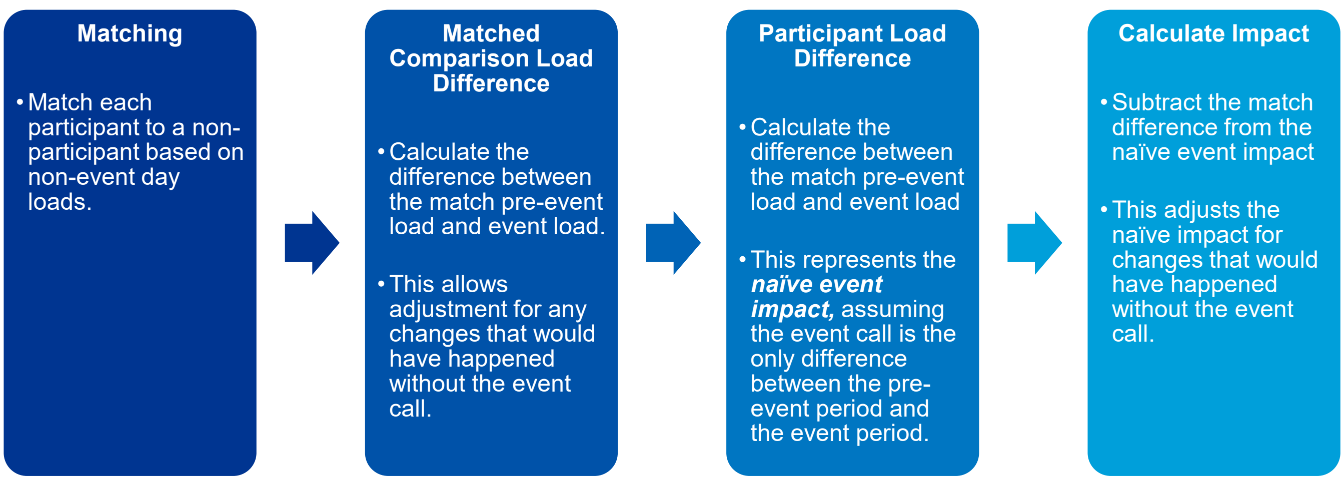
3.1 Impact evaluation methods

Program impact evaluation was composed of three primary steps, each discussed in detail below:

1. Data ingestion and cleaning
2. Matched comparison group selection
3. Difference-in-difference load impact estimation

Before going into detail about each of these steps, we give an overview of the matched comparison group selection and difference-in-difference load estimation, as shown in Figure 3-1. The matched comparison group selection step (shown in the first box, “Matching”) allows us to find nonparticipants who have similar hour-by-hour load and are from the same geographical area as participants. This can be thought of as approximating a randomized control trial, where a subset of participants is randomly held out from the event. In both cases, customers who are not in the event provide evidence of what the event participants would have done in the absence of the program. This is particularly important in the DR context, because event days are generally more extreme than surrounding days, limiting the effectiveness of settlement-type baselines[2] without a comparison group. Our matched comparison group is a pseudo–control group and helps us to calculate the baseline for the impact estimate calculations.

Figure 3-1. Impact evaluation analysis matching and difference-in-difference process



The difference-in-difference calculation (boxes 2-4 in Figure 3-1) allows us to correct for both (a) differences between non-event days and event days, and (b) any residual differences between program participants and the matched comparison group. The first difference is the difference for the matched comparison group between a pre-period and the event day, which adjusts for load difference between the two days. The second difference calculates this pre-period vs. event day difference for the treatment group, which is the naïve event impact, assuming the pre-period and the event day are the same except for the event call. The difference in difference is a subtraction of the match group difference from the naïve event impact, thus correcting for the difference between the pre-period and event day load characteristics. This process is explained in detail in Section 3.1.1 through Section 3.1.3.

2 Settlement baselines are simple and quick estimates of the baseline usage for a participant on event days. They are compared to actual usage for a rough impact calculation, traditionally used to ‘settle up’ with the customer, who was paid based on how much they curtailed during the event period.

3.1.1 Data ingestion and cleaning

DNV requested and received data for all program participants and all eligible matched comparison group customers. For both groups, we requested AMI data from the pre-program period through the 2022-2023 winter. DNV loaded and cleaned the data as discussed below. DNV standardized, formatted, aggregated, and combined these data, as necessary, to create a dataset of one hourly read per customer and premise for each hour of the day.

We also requested that participant data include the date of entry into the program and the timing of events. In combination, these data make it possible to identify non-event days for the rolling pre-event periods used for matching for each event.

For this program, customers were removed from analysis for three primary reasons: participation in other PSE energy efficiency programs during the study period, missing data, or not being present in the datasets. Table 3-1 shows the number of participating customers provided by PSE, the number of customers removed at each step, and the final number of customers included in the analysis.

Table 3-1. Participant exclusion

		Event Date							
Fuel Type	**Attrition Reason**	**Unknown Event**	**12/5/22**	**12/20/22**	**12/22/22**	**1/30/23**	**1/31/23**	**2/23/23**	**2/24/23**
Electric	Starting participant count	3	39	51	51	60	61	65	65
	Not in AMI data	-	1	1	1	1	1	1	1
	Not in event data	3	-	-	-	-	-	-	-
	Participated in other PSE EE program	-	1	1	1	1	1	1	1
	Missing >10% of data during matching period	-	-	-	1	1	1	1	1
	Final participant count	**-**	**37**	**49**	**48**	**57**	**58**	**62**	**62**
Gas	Starting participant count	2	-	54	54	71	71	74	74
	Not in AMI data	-	-	3	3	5	5	5	5
	Not in event data	2	-	-	-	-	-	-	-
	Participated in other PSE EE program	-	-	-	-	-	-	-	-
	Missing >10% of data during matching period	-	-	8	9	2	1	3	3
	Final participant count	**-**	**-**	**43**	**42**	**64**	**65**	**66**	**66**

3.1.2 Matched comparison group analysis

The second step of the impact evaluation process, as described above, is finding the matched comparison group. In this step, each participant in the DR program is matched with one or more similar control customers, allowing the impact calculation to control for differences between pre-event day usage and event day usage, in the absence of an event.

DNV tested multiple matching options for each fuel type in order to find the matched comparison group that provided the most accurate impact estimates. We assessed each for quality, and then chose the best-performing option across events for each fuel type. Matching was performed on a per-event basis to account for differing participant groups in each event, based on when customers signed up for the program and whether they participated in a given event.

This approach uses a model validation method that sets up a relevant and quantitative basis for distinguishing among options and then chooses the option that performs best among those tested. In this case, the test is for the option that best predicts load on event-like days that fall immediately after the event.

3.1.2.1 Matching options

All matched groups were created by finding nonparticipants with the most similar hourly energy usage values to participants on previous days that are both closest to the event day and have the most similar temperatures. To do this, we found the nonparticipants with the minimum Euclidean distance from participants, for hourly energy usage values. We then selected either (a) the *k* nearest neighbors for some number *k* > 1, or (b) the one nearest neighbor for each participant. For the first approach, which allowed multiple matches, eligible match customers could be used for multiple participants. For the second approach, each nonparticipating match could only be used for one participant.

Below, we provide a numbered list of the 10 different matching strategies we tested to find the one that would produce the most accurate impact results. These methods vary with regards to the hours of the day that were used for matching, the number of matched nonparticipants per participant, and in whether they matched on actual hourly usage or modeled hourly usage. Most methods (unless otherwise specified below) matched on hourly energy usage in the five most similar minimum temperature days to a given event in the 20 non-holiday, non-event, non-weekend days preceding an event.[3] These are referred to as pre-event matching days.

- "1:1 subset hours" matching:
  1. **All hours:** Participants were matched to their single most similar control customer based on hourly loads for all hours of the five pre-event matching days.
  2. **5 am-10 am:** Participants were matched to their single most similar control customer based on hourly loads from 5 am to 10 am during the five pre-event matching days.
  3. **3 am-10 am:** Participants were matched to their single most similar control customer based on hourly loads from 3 am to 10 am during the five pre-event matching days.
  4. **Event hours:** Participants were matched to their single most similar control customer based on hourly loads during the hours of a given event during the five pre-event matching days.
- "1:10 subset hours" matching:
  5. **All hours:** Each participant was matched to 10 controls (with replacement) based on all hours of the five pre-event matching days.

[3] Before matching, any customer that was missing more than 10% of hourly reads during the baseline period was removed from consideration, and any customer missing between 0 and 10% of reads had those reads imputed with the hour-specific mean load of that customer.

6. **5 am-10 am:** Participants were matched to their 10 most similar control customers based on hourly loads from 5 am to 10 am during the five pre-event matching days.

7. **3 am-10 am:** Participants were matched to their 10 most similar control customers based on hourly loads from 3 am to 10 am during the five pre-event matching days.

8. **Event hours:** Participants were matched to their 10 most similar control customers based on hourly loads during the hours of a given event during the five pre-event matching days.

- **"1:1 predicted event load" matching:**

9. **All hours:** Event-day loads were predicted for each participant and eligible comparison group site using a site-specific piecewise linear model where energy usage is a function of the time of week and the temperature.[4] These models were built using data from the two weeks prior to each event (excluding event days) and were then used to predict event day hourly loads given actual event-day temperatures. Matches were then made on all 24 hours of predicted event-day hourly loads.

10. **Event hours:** Event-day loads were predicted for each site using the modeling approach described above. Matches were then made on predicted event day loads during only the hours of a given event.

3.1.2.2 Matching assessment

Match quality was assessed using absolute mean difference (AMD) for each event against three "pseudo-events." Pseudo-events are defined as the first three non-holiday, non-event, non-weekend most similar minimum temperature days following a given event. This methodology allows for the assessment of match quality on event-like days, without tuning match selection towards a specific event day outcome.

AMD was calculated as the difference between the actual treatment load on these pseudo-event days and the calculated baseline for the pseudo-event days using the difference-in-difference methodology. On pseudo-event days, the actual load should be the same as the calculated baseline, because there is no event. Therefore, the difference is a measure of the error in each method.

The difference-in-difference baseline, which is describe above in Section 3.1.3, comes from taking the match group load on the pseudo-event day and subtracting the matching day difference between the match and treatment groups. This corrects for both weather on the pseudo-event day (by looking at match group load on that day) and any differences between the match and treatment groups (by looking at how they are different in the pre-period).

$$\text{AMD} = \frac{1}{n}\sum_{i=1}^{n}\left|\overline{t_{i1}} - \left[\overline{c_{i1}} - \left(\bar{c}_0 - \bar{t}_0\right)\right]\right|$$

where

$\bar{t}_0$ = average participant load during the pre-event matching period
$\bar{c}_0$ = average match group load during the pre-event matching period
$\overline{c_{i1}}$ = average match group load during specified hours on pseudo-event day i
$\overline{t_{i1}}$ = average participant load during specified hours on pseudo-event day i
n = number of pseudo-events

[4] More information on the details of this model can be found in Section 3b at http://docs.caltrack.org/en/latest/methods.html

The AMD metric was computed for all matching methods outlined above, and the method with the lowest AMD was chosen as the best method for each fuel type. AMD was calculated only on the hours of the event that each pseudo-event was approximating. For electric customers, 1:10 all hours matching resulted in the smallest AMD, and for gas customers, 1:1 all hours matching resulted in the smallest AMD. For each event, we selected a match group using these fuel-specific methods. See Appendix A for AMD results for each matching method.

Figure 3-2 and Figure 3-3 show pre-event matching day load shapes for the February 23, 2023, event for the best-performing matching approaches. The overlapping confidence intervals (shown as transparent bands in the plots), demonstrate the similarity of the pre-period load shapes for all 24 hours in the case of both electric and gas participants.

Figure 3-2. Pre-event matching load shapes, electric participants

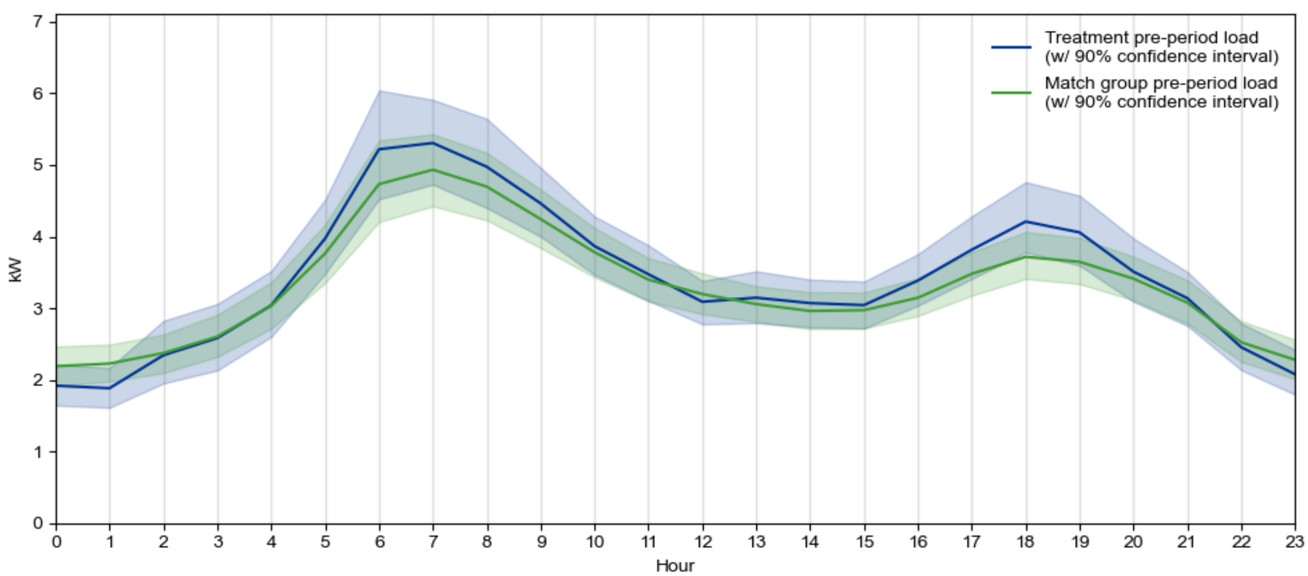
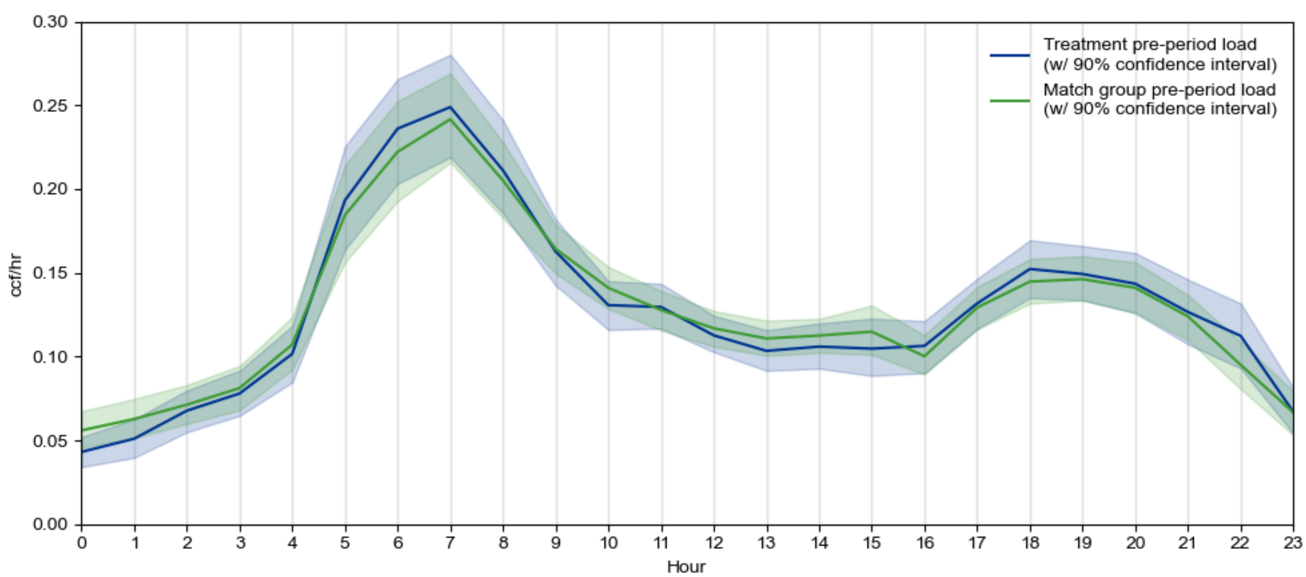


Figure 3-3. Pre-event matching load shapes, gas participants



3.1.3 Event curtailment estimation

In the last step of the impact evaluation analysis, the difference-in-difference calculation, the matches found for every program participant are used to correct for differences between the pre-period and the event day. This is especially important because events tend to be called on days that are the coldest of the season and therefore not well represented by any other days. For example, if the matched customer load is higher during the event period than calculated for the pre-event matching days, then the difference between the participant pre-event matching load and the participant's event load is adjusted upward.

To make this adjustment DNV takes the pre-vs-event usage difference for program participants and subtracts that difference observed in the match group. Load curtailment is therefore calculated (for each hour) as

$$curtailment = (t_0 - t_1) - (c_0 - c_1)$$

where

$t_0 =$ hour-specific participant load, averaged over pre-event matching days
$t_1 =$ hour-specific participant load during event
$c_1 =$ hour-specific match group load, averaged over pre-event matching days
$c_1 =$ hour-specific match group load during event

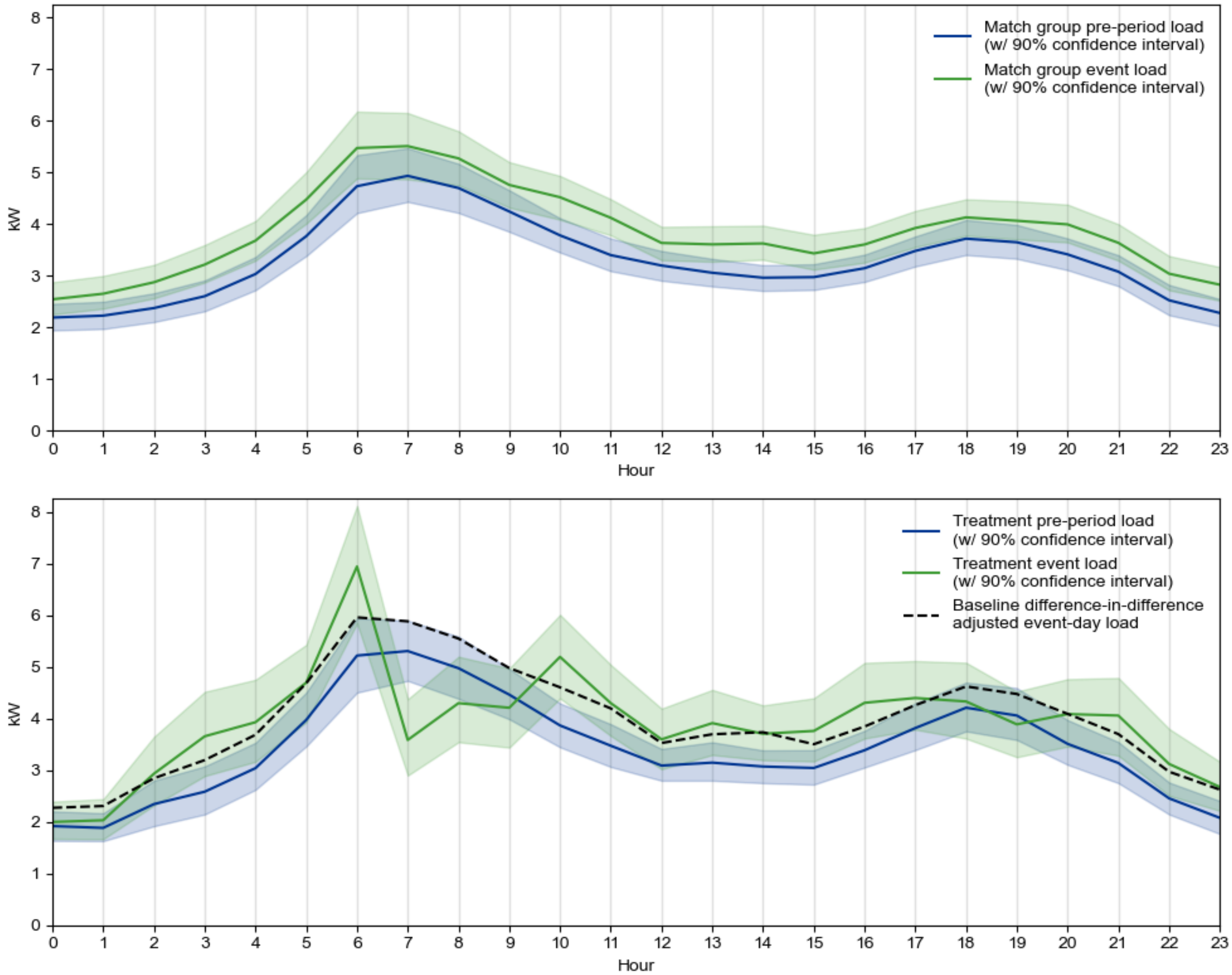
One way to conceptualize this methodology is as a "most-similar" five of 20 with an adjustment for weather. In other words, we averaged load over the five most similar weather days to an event date in the 20 eligible days preceding an event, and then used a matched comparison group to adjust for weather related differences between the event day and the five similar weather days. This differs from Autogrid's methodology, a 10 of 10 methodology without adjustment. That is, Autogrid used the average load on all 10 days out of the 10 days preceding an event and did not have any adjustment used to account for differences in weather between the event day and the 10 preceding days.

Curtailment estimates were calculated for each hour. We aggregated curtailment results across several different dimensions to understand program performance and variations. The priority aggregations are relative hour of event, event overall, and season overall.

Figure 3-4 shows a visualization of the average per customer curtailment experienced during the February 23, 2023, event by electric participants. These visuals align with the example provided above.

- The top frame shows the match group's pre-event matching consumption (blue) and event day consumption (green). The event consumption is higher than the pre-event matching consumption, so the difference results in a higher adjusted participant baseline load.
- The second plot shows the treatment group's pre-event matching consumption (blue), the adjusted participant baseline load (dashed black line), and the actual event-day consumption (green). Event curtailment was calculated as the hourly delta between the treatment event-day load (green line) and the adjusted baseline load (dashed black line).

Figure 3-4. Event-day loadshapes compared to pre-period baseline, electric participants



3.2 Process evaluation methods

The process evaluation was designed to identify program successes and opportunities for program improvement, and to contextualize the impact results using information gathered from participating customer surveys, interviews with program staff, and reviews of program documents. This information allows us to fully understand how the program is run and what customers experienced as participants. Combined results from these research activities provide insights on elements of the program that are going well and improvements that can lead to a smoother and more effective program that is easier to deliver.

The process research objectives we focused on for this study and the activities we undertook for each objective are shown in Table 3-2.

Table 3-2. Evaluation objectives and activities

Research objective	Evaluation activity
Characterize program participants: geography, equity, house size, etc.	Program participant web survey Comparison of survey respondents with larger population
Document program participation trends over time	Program participant web survey and Program staff interviews
Document program activities and spending over two-year program	Program staff interviews Documentation review
Assess customer reaction to program through customer surveys and interviews	Program participant web survey
Assess integration of DR into PSE operations	Program staff interviews

3.2.1 Customer survey

As the primary data source for the process evaluation, DNV sent surveys via email to 101 customers enrolled in the DR program in Duvall and on Bainbridge Island (Table 3-3). Of the customers who received a survey, 35 customers completed the survey, including 23 Bainbridge Island participants and 12 Duvall participants.

Table 3-3. Survey response rate by geography

Geography	Survey population	Number of completes	Response rate
Bainbridge Island	70	23	33%
Duvall	31	12	39%
Overall	**101**	**35**	**35%**

Data collected from the survey included questions on:

- Participant characteristics, including demographics and current electrification household equipment
- Reasons for participation
- Satisfaction with participation
- Event experiences

3.2.2 Program staff interviews

DNV conducted four in-depth interviews with PSE staff involved in the DR program. These interviews were designed to improve our understanding of program organization, program goals, the VPP platform, M&V estimates, DR events and results, program challenges, and program successes. We worked with PSE to identify teams and individuals to speak with regarding the processes involved in implementing the program, including staff in Product Development, Information Technology, and Distribution System Operations. Topics covered during the interviews included:

- Background, roles, and involvement with the DR program
- Program assessment including marketing approaches, outreach methods, and perceived successes and barriers to program participation
- Key performance indicators and program goals
- Opportunities for growth or process improvements

3.2.3 Documentation review

To better understand the internal processes regarding program implementation, we reviewed existing program logic models. Additionally, DNV reviewed newsletters and direct email messaging to assess how customers are hearing about the program and the information they are given regarding enrollment and benefits to participation.

4 IMPACT EVALUATION RESULTS

DNV provides DR program impact results in kW per participant for the electric program and ccf/hr per participant for gas. Curtailment (load reduction) is represented as a positive value, and negative values indicate increased load. We use the standard error of the mean (SEM) and a *z*-coefficient of 1.645 to construct 90% confidence intervals for our estimates.

The aggregate results do not include load impacts from the February 24, 2023, event because its hourly impacts were not comparable to other events. This event lasted three and half hours, from 6:30 am PST until 10:00 am PST, but our analysts only had access to hourly AMI data. This meant that impact estimates for the 6:00 am hour included a half hour of precondition and a half hour of event data. Similarly, the impact estimates for the 7:00 and 8:00 am hours were not real first-hour or second-hour impacts, because they missed the first half hour of the event. We therefore present impacts by hour for this event but do not include its impacts in the aggregated results. In future years, we will work with PSE to evaluate impacts based on 15-minute AMI data, so as to align evaluation hours with event-call hours.

4.1 Overall impacts

We estimate that the average participant load curtailment achieved by the program to date is 1.59 kW for electric participants and 0.106 ccf/hr for gas participants across all events (see Table 4-1). Also see Figure 4-6 and Table 4-4 for event-specific average impacts. We also observed significant amounts of pre-heating in the hour before an event was called, as well as significant "snapback" heating in the two hours after an event.

- Pre-heating accounted for an average load increase of 0.73 kW per customer for electric participants and 0.079 ccf/hr per customer for gas participants in the hour before the event.
- Snapback heating accounted for an average load increase of 0.84 kW per customer for electric participants and 0.089 ccf/hr per customer for gas participants in the two hours after the event.

Table 4-1. Summary of curtailment, pre-heating, and snapback impacts

Fuel Type	Impact Period	Curtailment	Units	# of Hours	# of Premises	90% CI
Electric	**Participant avg. curtailment across all events**	**1.59**	**kW**	**2 or 3**	**63**	**0.18**
	Participant avg. pre-heating increase across all events	-0.73	kW	1	63	0.32
	Participant avg. snapback heating across all events	-0.84	kW	2	63	0.21
Gas	**Participant avg. curtailment across all events**	**0.106**	**ccf/hr**	**2 or 3**	**68**	**0.009**
	Participant avg. pre-heating increase across all events	-0.079	ccf/hr	1	68	0.013
	Participant avg. snapback heating across all events	-0.089	ccf/hr	2	68	0.012

We calculated event total curtailment as average per-premise curtailment during event hours multiplied by the number of participants enrolled in that event, regardless of whether they were removed during data cleaning. As a result, total curtailment for a given event is a function of the number of participants in that event and the hourly average for that event. Based on the per-customer average curtailment from each event and the number of customers enrolled, we estimate that the winter season events were responsible for an event average total curtailment of 95 kW per event for electric events and 7.3 ccf/hr per event for gas events. Table 4-2 shows event-specific totals. See Appendix D for more detail on cumulative event impacts.

Table 4-2. Total curtailment by event

Fuel Type	Event Date	Curtailment	Units	# of Premises	90% CI
Electric	12/5/2022	100	kW	39	21
	12/20/2022	114	kW	50	28
	12/22/2022	103	kW	50	32
	1/30/2023	86	kW	60	28
	1/31/2023	76	kW	60	24
	2/23/2023	93	kW	65	24
Gas	12/20/2022	8.1	ccf/hr	54	1.6
	12/22/2022	7.1	ccf/hr	54	1.6
	1/30/2023	6.8	ccf/hr	71	1.4
	1/31/2023	7.0	ccf/hr	71	1.3
	2/23/2023	7.6	ccf/hr	74	1.5

4.2 Impacts by hour

We found that for both fuel types on average, curtailment dropped as time went on, with average impacts during each hour significantly lower than the previous hour. This trend can be seen in Figure 4-1 and Figure 4-2 with associated values in Table 4-3, which show the average per-customer load curtailment by hour into an event, averaged across all events (except the above-mentioned 2/24 event). The x-axis in both figures shows relative hours to an event. Negative hours are the hours before an event. Hours 1-3, shown as green bars, are the event hours. The remaining positive hours are hours following an event. We use these data visualizations to compare the nth hour of an event between events that have different start times and/or different lengths. Positive values in these plots correspond to load curtailment, while negative values correspond to increased load. It is important to note that there is increased load in the hour preceding an event (pre-heating) and increased load in the hours following an event (snapback). We see that the timing of the event relative to the system peak is particularly important, because calling an event just an hour after peak could cause increased load during that peak hour.

Figure 4-1. Load curtailment by hour into event, electric participants

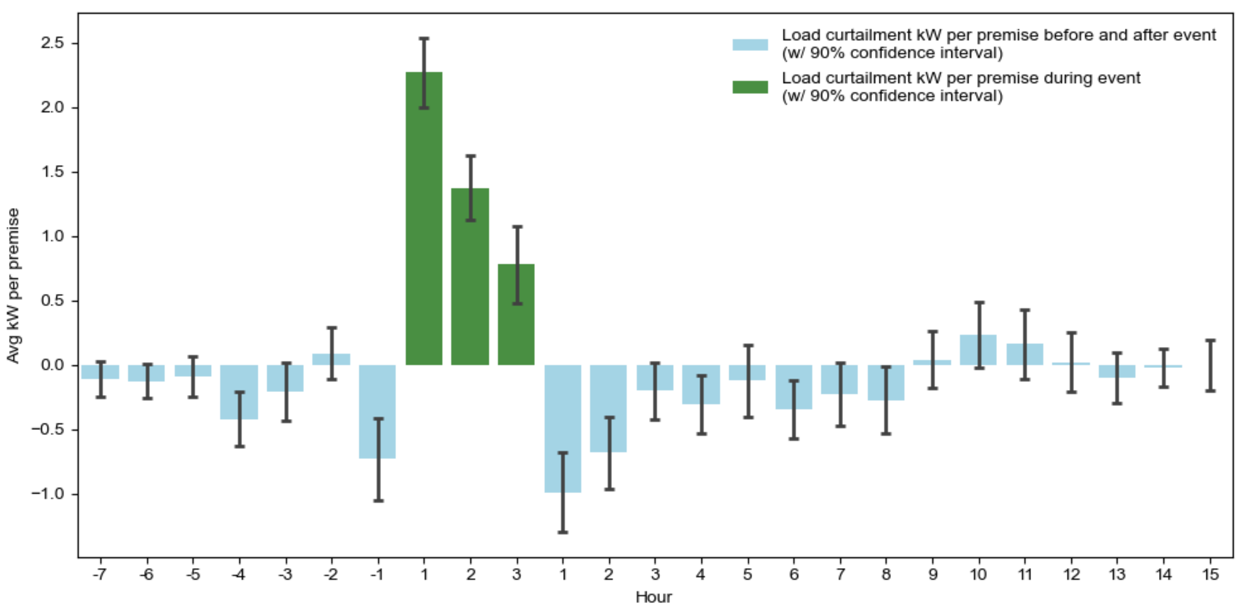


Figure 4-2. Load curtailment by hour into event, gas participants

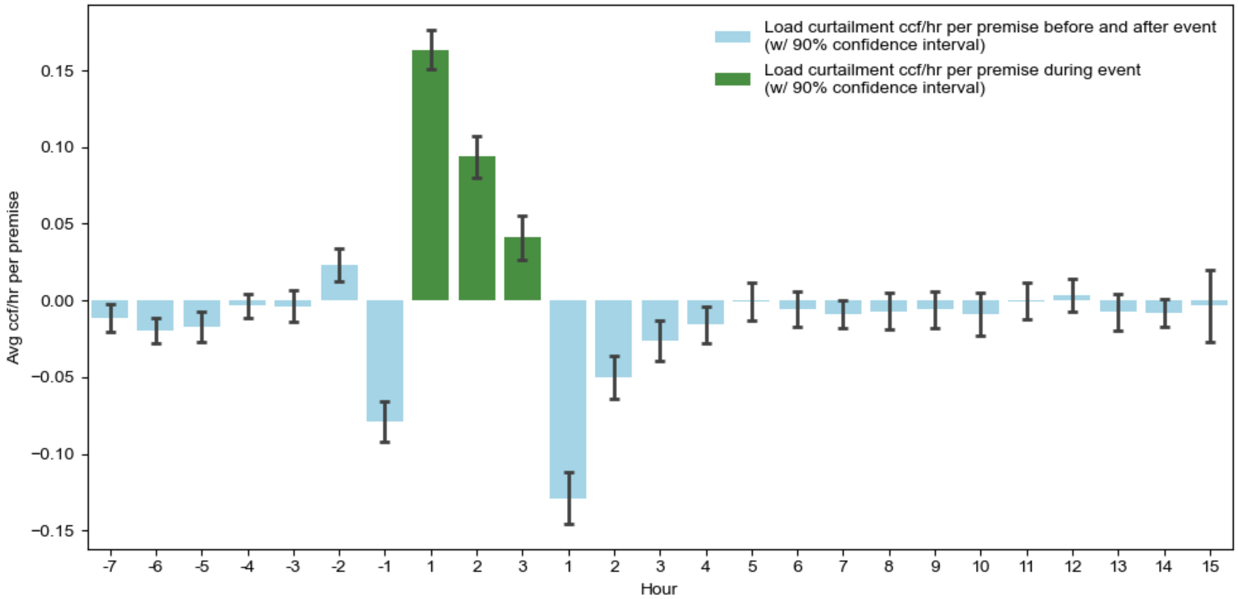


Table 4-3. Curtailment estimates by hour of event

Fuel Type	Hour of Event	Curtailment	Units	# of Premises	90% CI
Electric	1	2.27	kW	63	0.27
	2	1.38	kW	63	0.25
	3	0.78	kW	62	0.30
Gas	1	0.164	ccf/hr	68	0.013
	2	0.094	ccf/hr	68	0.014
	3	0.041	ccf/hr	68	0.015

Hourly impacts varied by customer. Figure 4-3 and Figure 4-4 show the distribution of average load impacts per customer by hour into an event. The distributions are mostly symmetric and centered to the right of zero. For all event hours, there were some number of customers that appear to have experienced increased load. The variation in per-person average impacts shown in these plots is primarily due to natural variation in impact among customers, and the associated error. The matched comparison group will generally substantially reduce this kind of error, relative to a settlement-type baseline. When creating a matched comparison group, not every individual participant will have an exact match, leading to some comparison group-related error for their specific estimate. However, compared to the error from a participant, non-event day baseline alone, this error is small. Additionally, the standard errors for estimates using a comparison group take into consideration this additional error that comes with the substantial bias correction.

Figure 4-3. Distribution of hourly per customer impacts, electric events

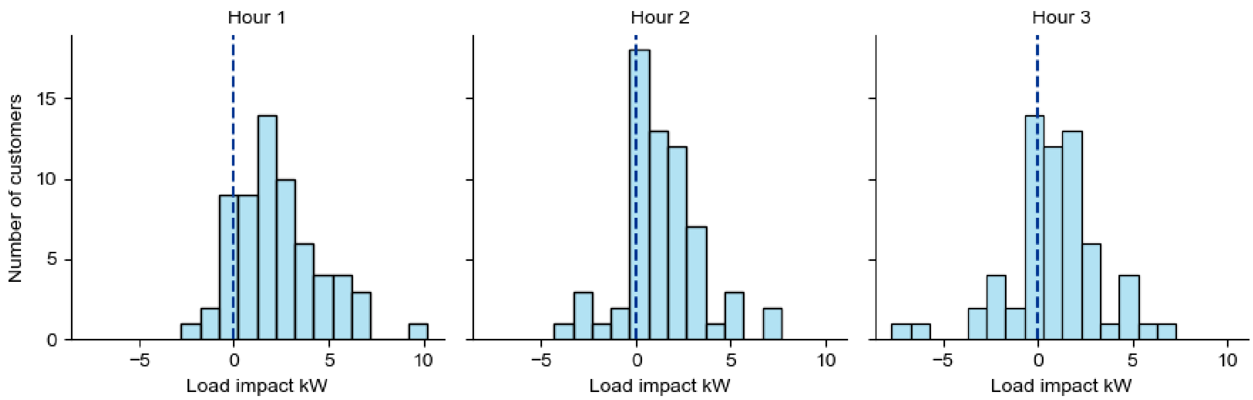
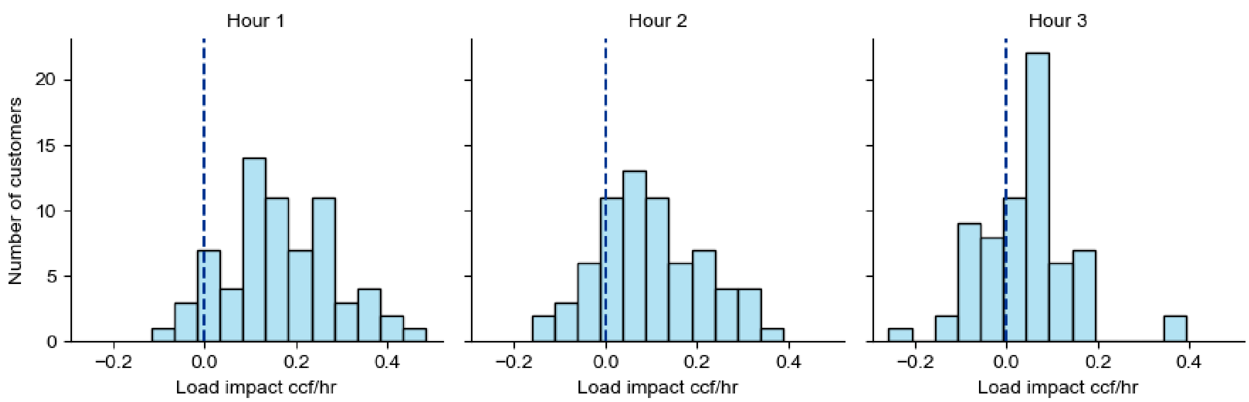


Figure 4-4. Distribution of hourly per customer impacts, gas events

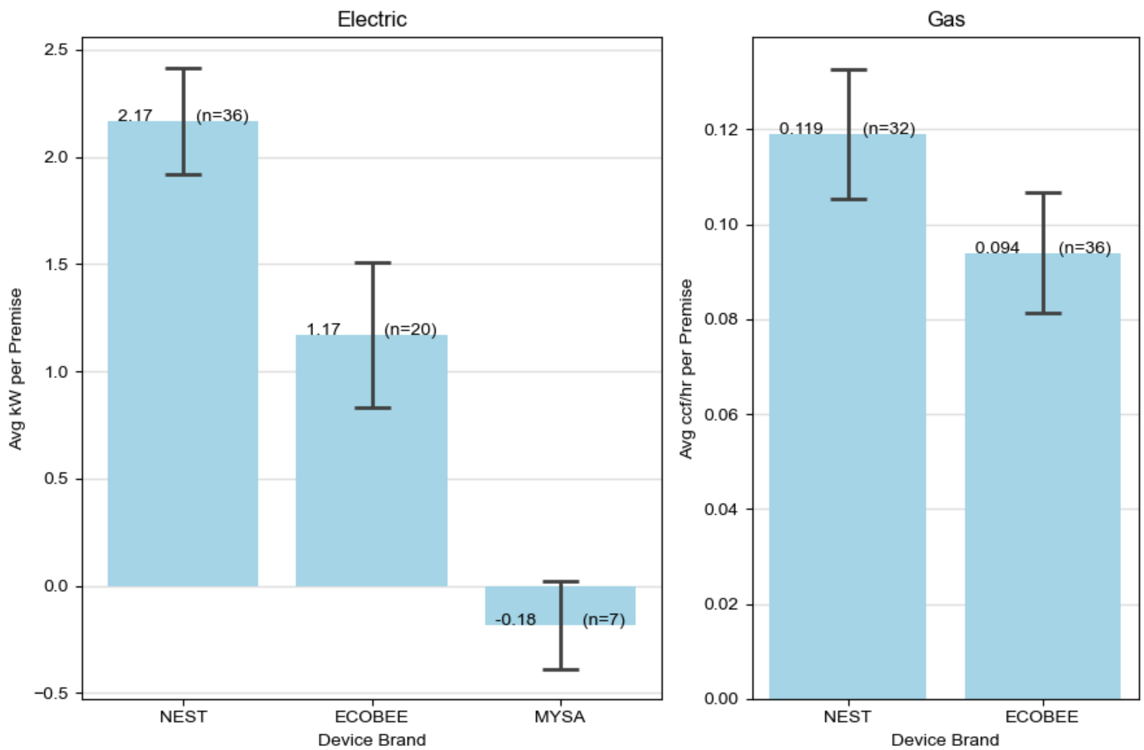


4.3 Impacts by device

Our analysis revealed that curtailment differed by device brand. For the electric program, premises with Mysa thermostats experienced minimal load impacts during events, and these impacts were not statistically different from zero. For both fuel types, Nest devices outperformed Ecobee devices. Premises with a Nest device experienced 26% more gas curtailment and 85% more electric curtailment during event hours on average compared to premises that had an Ecobee device installed. Less curtailment from Ecobee thermostats compared to Nest thermostats is consistent with other DR evaluations we have done and has been at least partially attributed to more confusing messaging in Ecobee thermostats. However, we also expect the populations purchasing the different brands are different (we saw much different survey response rates for the different thermostat types (see Figure 4-5) and we know that the different brands have different algorithms for trying to reduce energy usage while still keeping customers comfortable.

Figure 4-5 displays the curtailment by thermostat brand graphically. Additional exploration suggests that gas customers with a Nest device engaged in more pre-heating than those with an Ecobee. Additionally, for both fuels, Ecobee devices reduced load in the two-hours-before the hour preceding an event before increasing load during the one-hour-before pre-heating period. This suggests that the hour before pre-heating and the hour of pre-heating for Ecobee devices could combine to bring household temperatures to normal levels rather than achieving true pre-heating. This could, potentially, cause more overrides because of customer discomfort. Meanwhile, pre-heating for Nest devices would raise household temperature, allowing customers to go longer without overriding the thermostat. See Appendix G for hourly impacts by device brand, including pre-heating and snapback heating effects.

Figure 4-5. Curtailment estimates by device brand



4.4 Water heater impacts

According to the data provided to DNV, there was one customer with a water heater enrolled in the DR program. The analysis of this device was limited by large uncertainty in estimates due to both the small, expected impact from water heaters and the small sample size of one person. To assess curtailment from water heaters, a much larger sample size is needed to provide the necessary precision with small magnitudes of curtailment.

4.5 Seasonal persistence

This section provides the results of our curtailment persistence analysis. Persistence in this case refers to long term persistence, or how achieved curtailment changed over the course of the winter. We first look at persistence from event-to-event across the season. We then examine whether curtailment changed as a participant experienced more events.

Looking at curtailment persistence over the course of the winter, we found that hourly impacts appeared consistent across all non-consecutive events. There appears to be a spike in hour one of the 12/22 event, potentially due to extremely low temperatures on this day, and a dip during hour one of the 1/31 electric event, potentially because this was part of a consecutive day event. The second-hour electric curtailment amount does appear to decline through the season, but it is difficult to determine if that is because the last three events were longer (and thermostat algorithms may be different for longer events) or for other reasons, like temperature differences. Other than this, there are no clear trends. Figure 4-6 visualizes these patterns and Table 4-4 displays the details.

Figure 4-6. Average per-customer impacts by event and hour

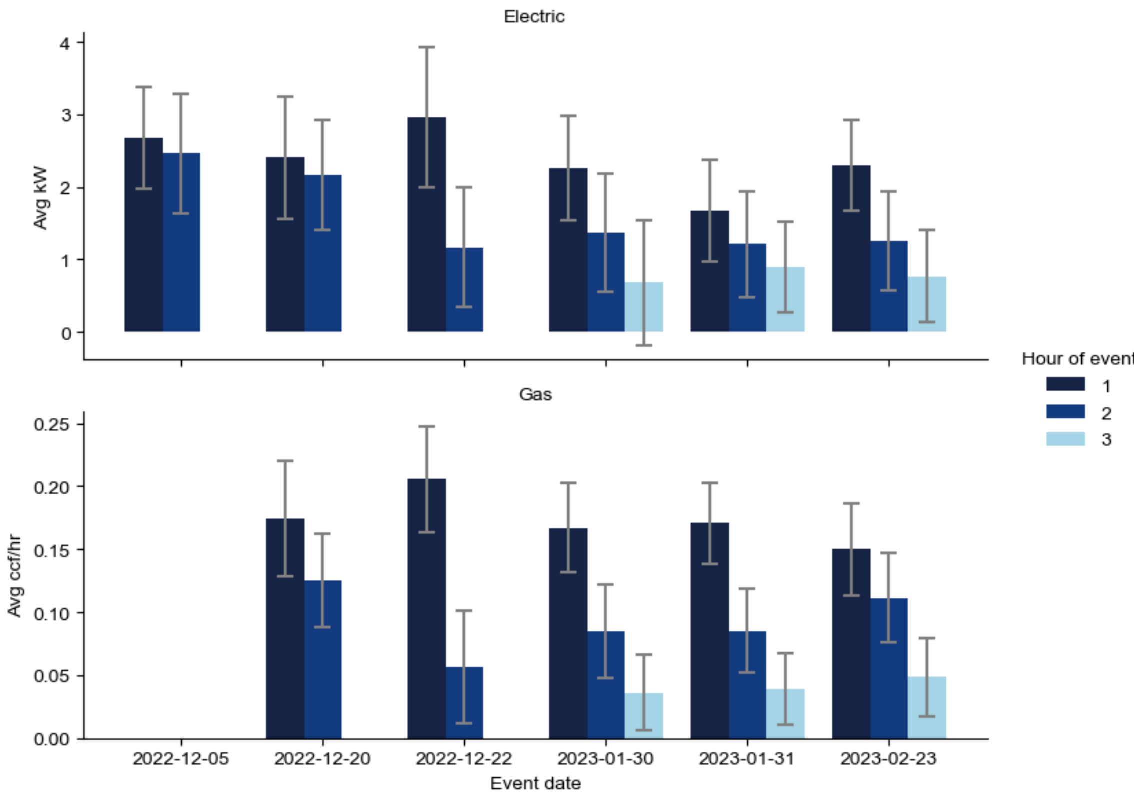


Table 4-4. Average per-customer impacts by event and hour

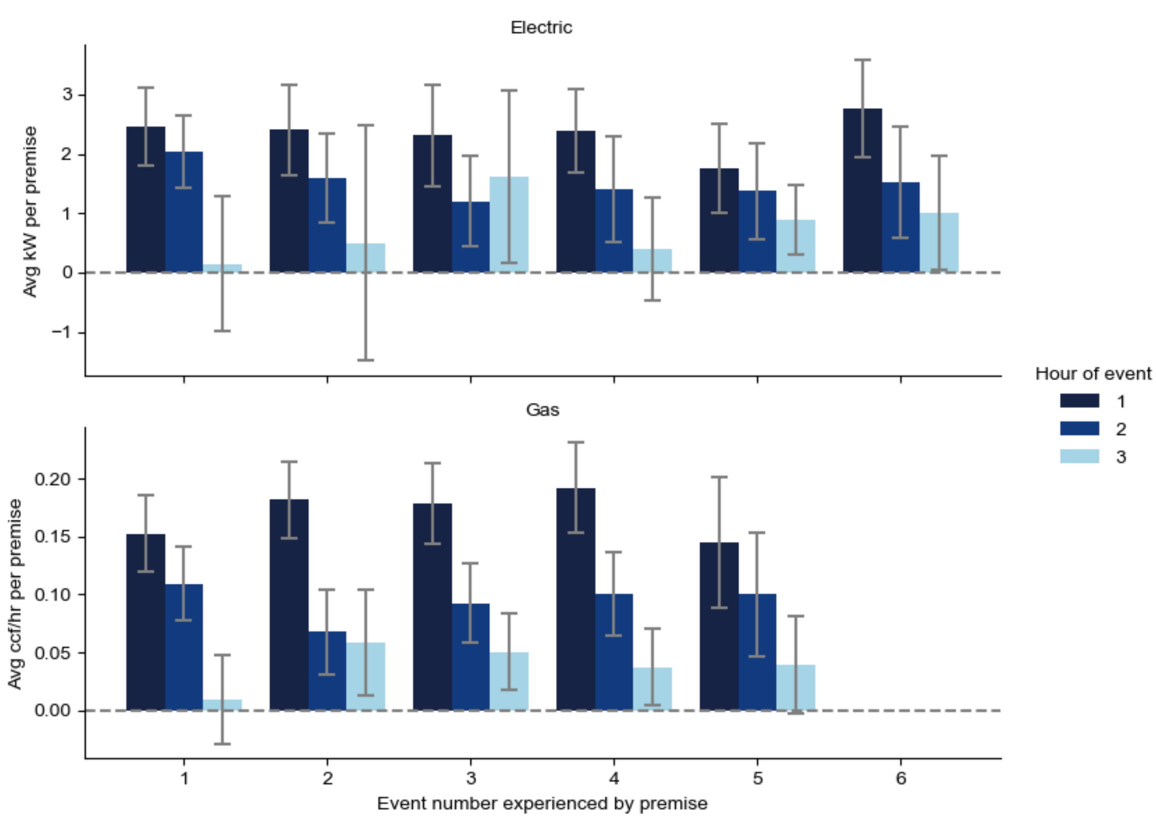
Fuel Type	Event Date	Hour of Event	Curtailment	Units	# of Premises	90% CI
Electric	12/5/2022	1	2.67	kW	37	0.70
		2	2.46	kW		0.83
	12/20/2022	1	2.40	kW	48	0.85
		2	2.16	kW		0.76
	12/22/2022	1	2.96	kW	47	0.97
		2	1.16	kW		0.82
	1/30/2023	1	2.26	kW	57	0.72
		2	1.37	kW		0.82
		3	0.67	kW		0.86
	1/31/2023	1	1.67	kW	57	0.70
		2	1.20	kW		0.73
		3	0.89	kW		0.62
	2/23/2023	1	2.30	kW	62	0.63
		2	1.25	kW		0.68
		3	0.77	kW		0.64
Gas	12/20/2022	1	0.174	ccf/hr	43	0.046
		2	0.125	ccf/hr		0.037
	12/22/2022	1	0.205	ccf/hr	42	0.042
		2	0.056	ccf/hr		0.044

Fuel Type	Event Date	Hour of Event	Curtailment	Units	# of Premises	90% CI
	1/30/2023	1	0.167	ccf/hr	64	0.036
		2	0.085	ccf/hr		0.037
		3	0.036	ccf/hr		0.030
	1/31/2023	1	0.170	ccf/hr	63	0.032
		2	0.085	ccf/hr		0.033
		3	0.039	ccf/hr		0.028
	2/23/2023	1	0.150	ccf/hr	66	0.037
		2	0.111	ccf/hr		0.035
		3	0.048	ccf/hr		0.031

The second persistence analysis explores whether participants are likely to change their curtailment as they experience more events. For example, participants could be more likely to override their thermostats after experiencing how cold their homes get during the first few events they participate in.

According to Figure 4-7, curtailment by hour appears to be relatively stable, regardless of how many events a participant experienced. In the figure, each hour in a given group may have a different number of premises contributing to the average because not all sites participated in all events. For example, some sites only participated in the last event, meaning the first event they experienced lasted three hours, whereas many other sites may have participated in all events, meaning their first event lasted only two hours. This differing sample size is responsible for the wide variation of error bounds for a specific value of "event number experienced" and for the especially large error bounds for hour three.

Figure 4-7. Curtailment by hour and event number experienced



The last way we sought to examine persistence was how curtailment changed during consecutive-day events, but we were not able to draw any generalizable conclusion here. There were two instances of consecutively called events during the winter season: the pair of events on 1/30/23 and 1/31/23 and the pair of events on 2/23/23 and 2/24/23. As was mentioned previously, the 2/24 event was removed from consideration due to data issues, meaning that the 2/23 event (as part of this pair) had to be removed as well. A visual assessment of results from the one instance of consecutive-day events with good data suggests that curtailment is reduced on the second day. But the observable difference is not statistically significant. It is important to gather more consecutive-day event data before drawing conclusions, as the difference between 1/30 and 1/31 could be due to the days being consecutive, but could also be a result of other factors, such as differing weather or different days of the week.

4.6 Comparison with aggregator estimated impacts

DNV observed large discrepancies between our estimates of achieved demand reduction and Autogrid's estimates, primarily because of differing baselines and partially because of different actual event-day usage in our datasets. DNV's impact estimates varied from about 1.0 to 2.5 kW/participant for electric and 0.06 to 0.15 ccf/hr for gas. In contrast, Autogrid's estimates varied from about -1.3 to 2.1 kW/participant for electric and 0 to 0.14 ccf/hr for gas. Autogrid's highest estimates are similar to DNV's, but their lowest actually indicate load growth (for electric) or no effect (for gas). DNV, in contrast, estimates consistent reduction of load during events for both fuels. One of the drivers of differing impact estimates between DNV and Autogrid was apparent data issues and inconsistencies from Autogrid (discussed in detail below). In addition to an overall comparison, we compared our estimates of impacts to the subset of customers that had the same event day loads (same actuals) between our data and Autogrid for a more apples-to-apples comparison. See Figure 4-8 and Table 4-5.

Figure 4-8. Comparison of DNV and Autogrid curtailment estimates by event date

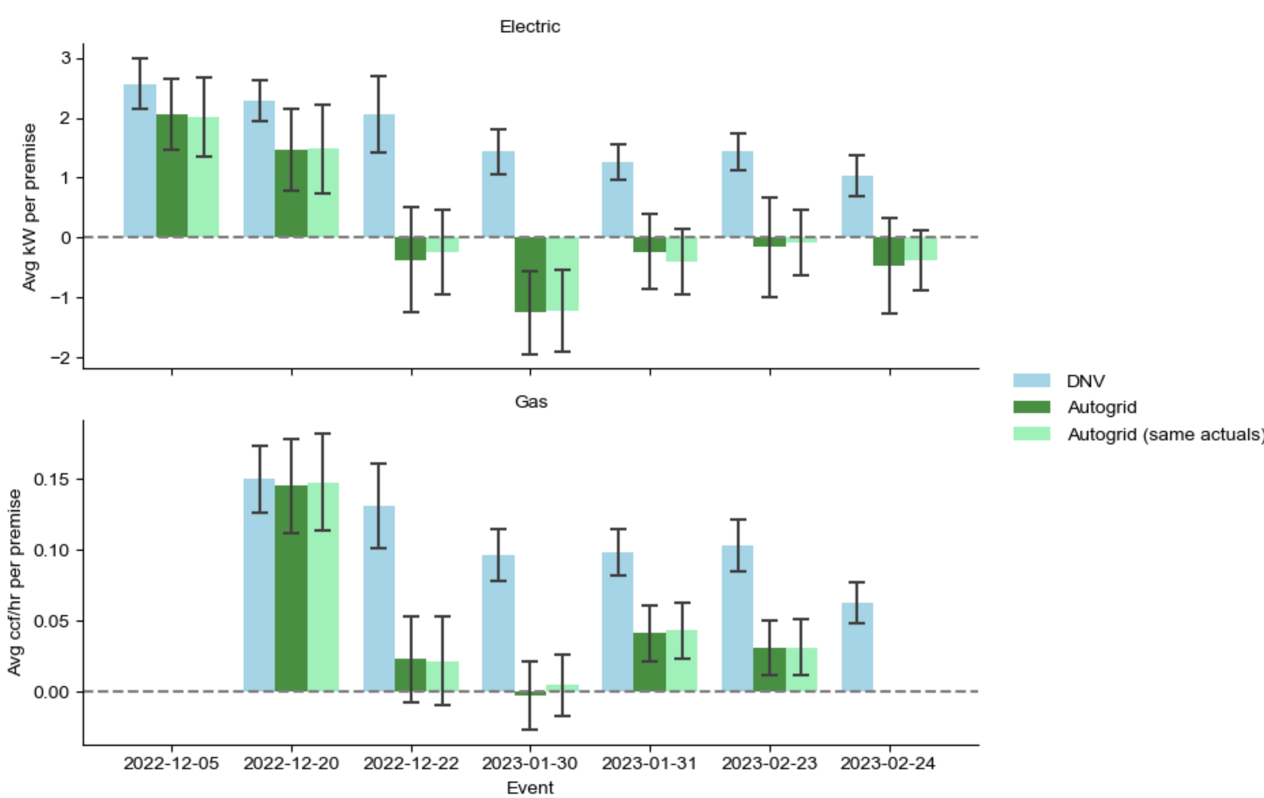


Table 4-5. Comparison of DNV and Autogrid curtailment estimates by event date

			Curtailment		# of Premises		90% CI	
Fuel Type	**Event Date**	**Units**	**Autogrid**	**DNV**	**Autogrid**	**DNV**	**Autogrid**	**DNV**
Electric	12/5/2022	kW	2.06	2.56	35	37	0.59	0.42
	12/20/2022	kW	1.47	2.28	49	48	0.69	0.34
	12/22/2022	kW	-0.37	2.06	48	47	0.87	0.64
	1/30/2023	kW	-1.25	1.43	58	57	0.69	0.37
	1/31/2023	kW	-0.24	1.25	58	57	0.63	0.29
	2/23/2023	kW	-0.16	1.44	64	62	0.84	0.31
	2/24/2023	kW	-0.47	1.04	64	62	0.80	0.34
Gas	12/20/2022	ccf/hr	0.145	0.150	44	43	0.033	0.024
	12/22/2022	ccf/hr	0.023	0.131	46	42	0.030	0.030
	1/30/2023	ccf/hr	-0.003	0.096	64	64	0.024	0.018
	1/31/2023	ccf/hr	0.041	0.098	65	63	0.020	0.017
	2/23/2023	ccf/hr	0.031	0.103	68	66	0.019	0.018
	2/24/2023	ccf/hr		0.063		63		0.015

An investigation into the discrepancies revealed two primary drivers of lower impact estimates from Autogrid: differing baseline estimates and differing actual event day usage values.

Autogrid's estimates of baseline usage were consistently lower than DNV estimates of baseline usage. Autogrid used the 10 of 10 methodology for impacts, which used a specific participants' usage on the 10 days before the event as the baseline. Because events are called on the coldest days, these 10 days are highly likely to have lower heating load, resulting in an underestimation of impact estimates. DNV's matched control difference-in-difference method adjusts for these cold days by looking at how the matched customer's usage varies between a pre-period and the event day. The pre-period is also selected to be more similar to the event day than the 10 days before the event.

Additionally, we identified a handful of customers for which Autogrid's data on actual observed event day usage were integer multiples larger than the values in data provided from PSE. For example, if one of these customers' actual average event hour demand was 1.2 kW in the data we were provided, event data from Autogrid indicated that this same value was 2.4, 3.6, or 4.8 kW, depending on the case. That is, two, three, or four times larger than the usage data from PSE. Autogrid also did not make a distinction between fuel types in their event data. We identified a handful of Duvall customers whose impacts were calculated based on reads from electric smart meters rather than gas smart meters. These customers were removed when aggregating Autogrid's estimates.

See Appendix C for a detailed comparison of baselines, event day actuals, and impacts under different scenarios.

4.7 Energy savings

DNV also assessed whether there were overall energy (kWh or ccf) savings over the course of each event day. However, we see that the energy impacts are inconsistent by day, so the program event days are not consistently saving energy for either electricity or gas and may sometimes be associated with increased energy use (see Figure 4-9 and Figure 4-10). This is not unexpected, given that the goal of the program is demand reduction during specific hours. These results also represent the relatively small number of enrollees in this year's program, and results will be more certain in 2024 reporting.

Figure 4-9. Average kWh saved per premise, by event day

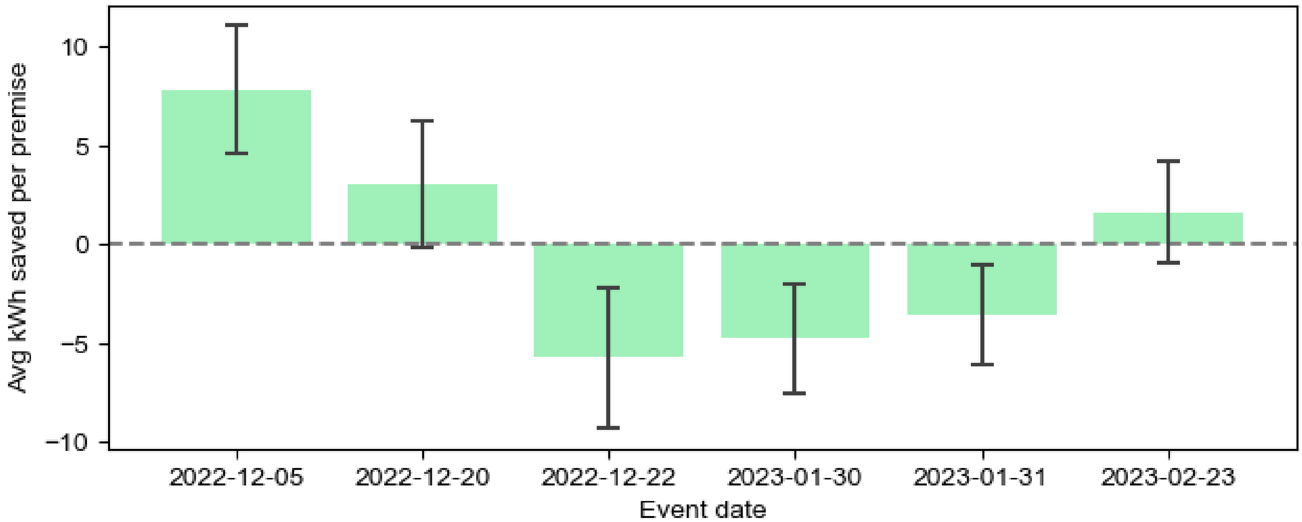
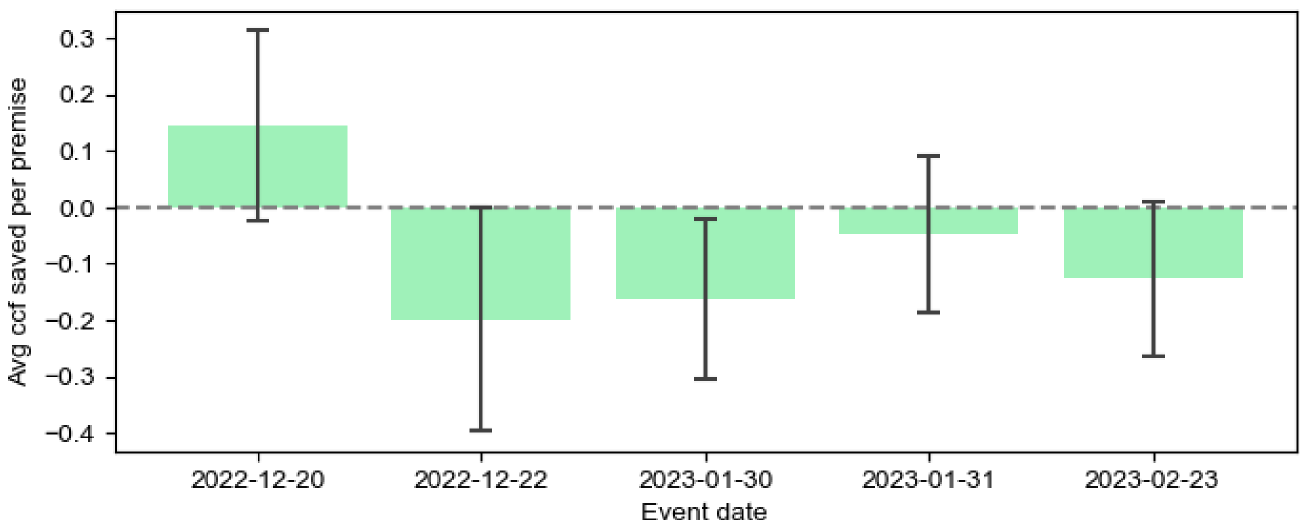


Figure 4-10. Average ccf saved per premise, by event day



5 PROCESS EVALUATION RESULTS

In this section, we provide the results of the process evaluation, which include findings from the customer survey, program staff interviews, and documentation review. For a description of these research activities, process evaluation methods, and customer survey disposition, please see *Section 3.2 Process evaluation methods*.

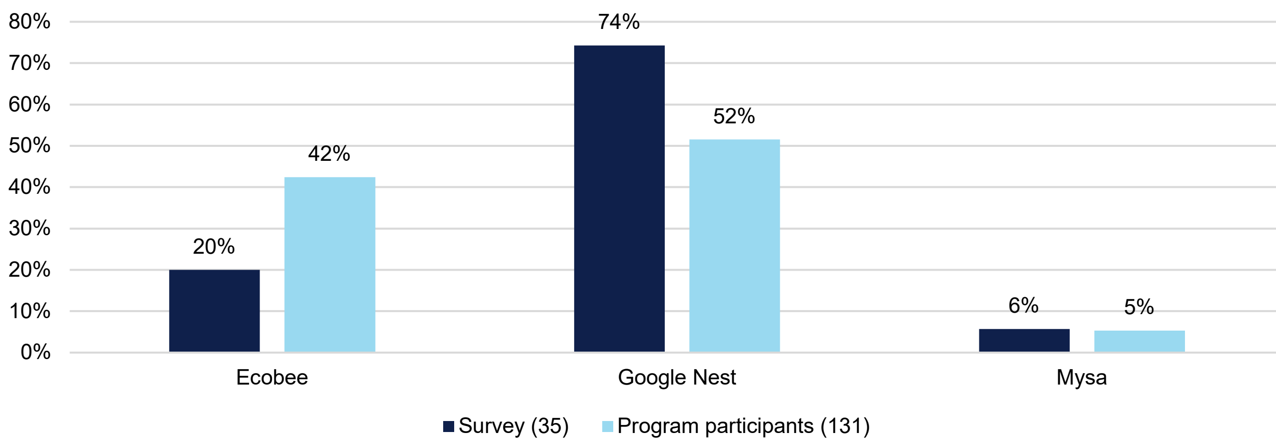
5.1 Customer survey

In cases with notable differences in responses between Bainbridge and Duval respondents, we break out the results by geography. In most cases, we combine the responses because there was little to no difference between the groups.

5.1.1 Survey respondents relative to program participants and overall populations

The goal of collecting demographic information in this survey is to determine how program participants compare to the overall population. This type of analysis can help PSE find out which population segments are easiest to recruit to the program (they should be overrepresented among participants), and which segments are more challenging to enroll (they will be underrepresented). However, these conclusions are dependent on survey respondents being representative of program participants. It is always challenging to know if the bias in who responds to the survey is large enough that survey respondents are not representative of program participants. In this case, we are somewhat less certain because of the small size of the first-year pilot (and corresponding small survey numbers). Additionally, we know that survey respondents did not represent the proportions of brands of thermostat in the program (Figure 5-1) but is unclear how that relates to demographic representativeness.

Figure 5-1. Brand of smart thermostat owned by participants (n=35)



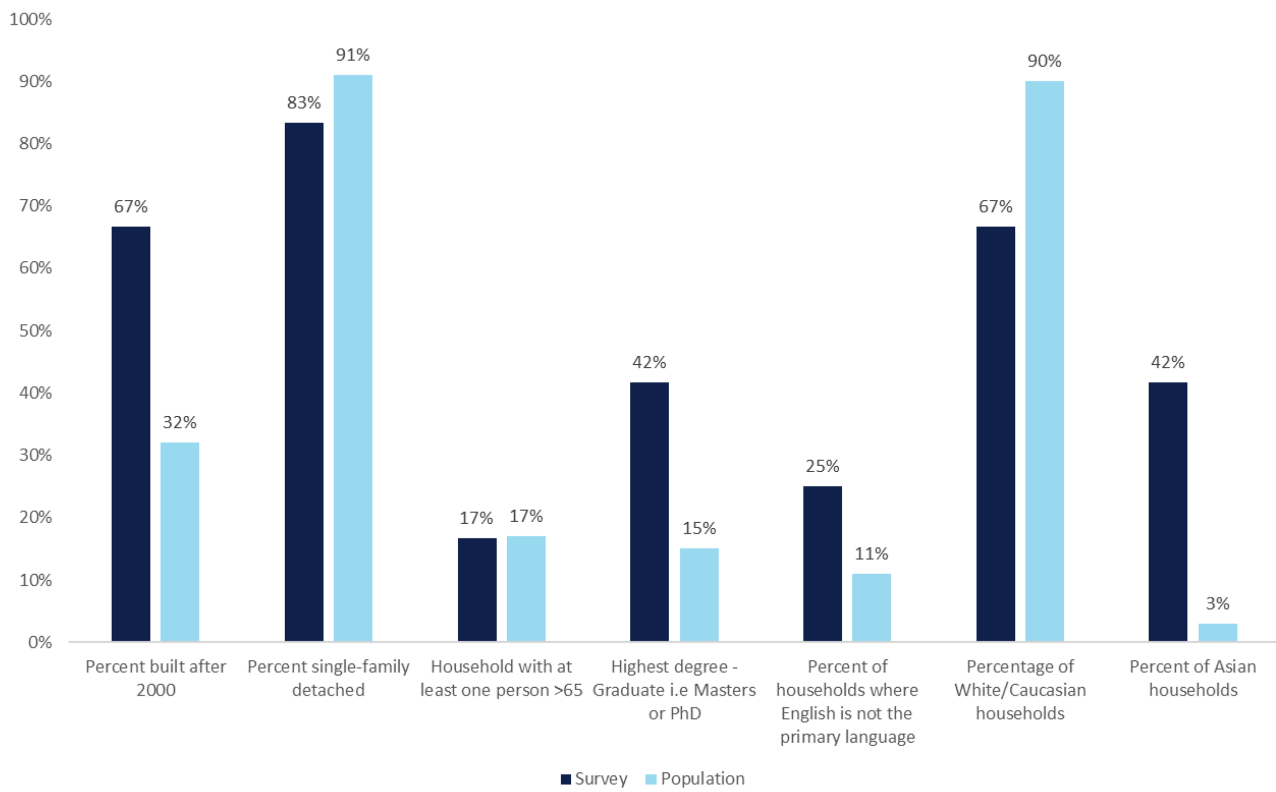
In the below graphs, we show how survey respondents compare to the overall populations of Bainbridge and Duvall and describe insights dependent on the survey respondents being representative of program participants. Therefore, these should be interpreted with caution. Our report on the 2023-2024 program year will include more survey responses and should reduce the level of uncertainty in the results.

The below graphs, which show how participants are different from the rest of the populations in these geographies, give insights into what demographic groups are most likely to continue adoption and what groups have less participation and can be targeted to expand the program's reach.[5]

[5] Population-level data is from the 2019 American Community Survey, published by the United States Census Bureau. These are from the 5-year datasets for zip codes 98109 (Duvall and surrounding areas) and 98110 (Bainbridge Island). https://www.census.gov/programs-surveys/acs/news/updates/2019.html

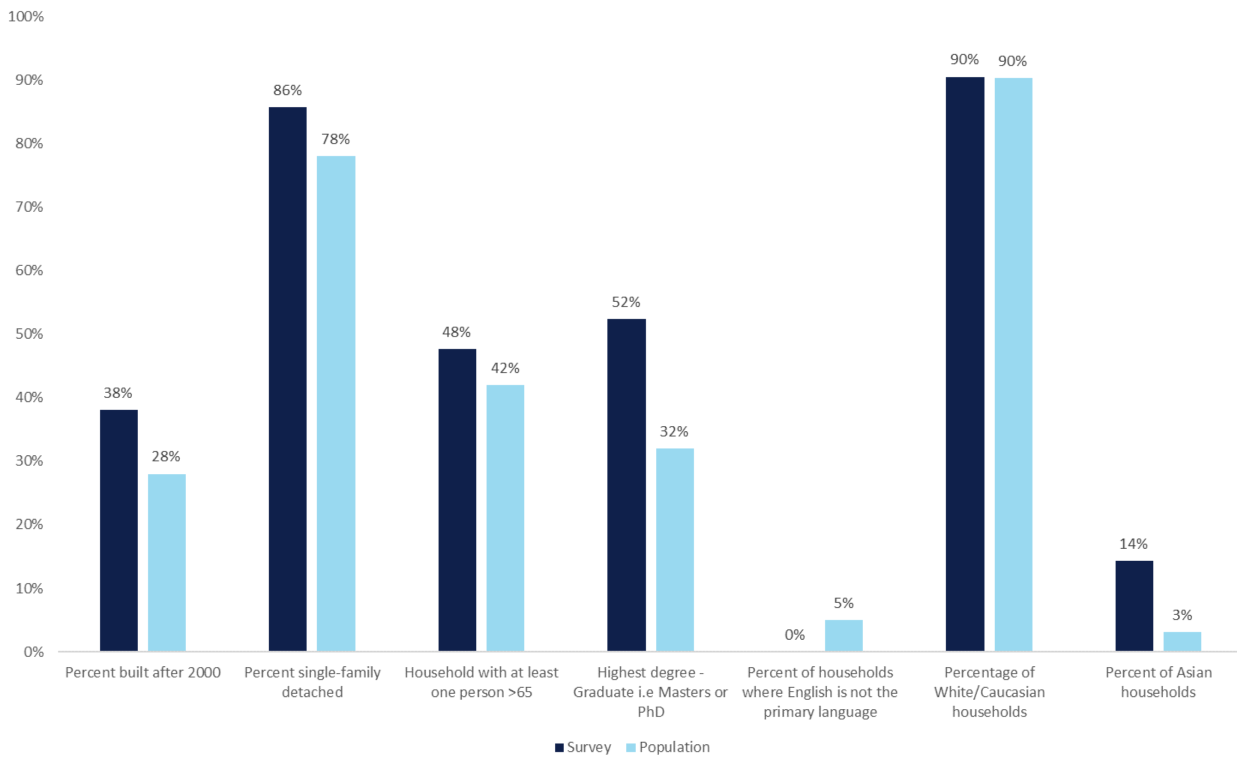
Figure 5-2 shows how survey respondents differed from the overall population in Duvall. This indicates that participants tend to be more educated, less likely to speak English at home, and more likely to be Asian than the overall Duvall population. Participants also tend to have newer homes. This indicates that residents who are highly educated, Asian, and/or live in newly constructed homes are the most likely to enroll in the program and are likely to require relatively little outreach effort to recruit. On the other hand, residents who are less educated, white, and/or live in older homes are less likely to already be enrolled and represent an opportunity for increased participation.

Figure 5-2. Duvall participant demographics



In Figure 5-3, we see a similar comparison of survey respondents to the overall population of Bainbridge Island. Overall, Bainbridge program participants appear to be more likely to reflect the overall demographics than we see in Duvall. However, participants are more likely to be more educated, Asian, and/or live in newer houses. Again, these demographic groups are likely to be easiest to recruit into the program, while less educated, white, and/or those who live in older houses represent groups that are less likely to enroll in the program but provide an opportunity for increasing enrollment.

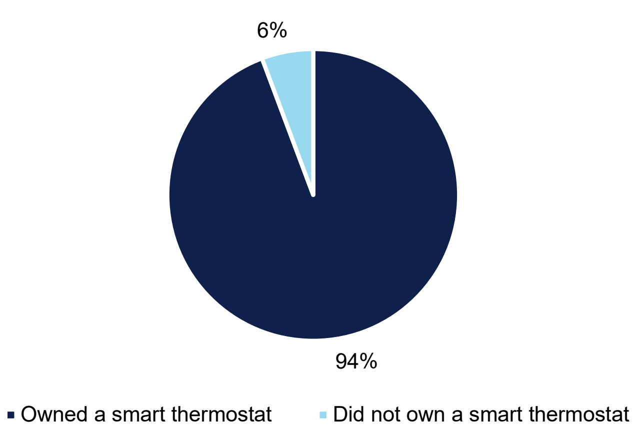
Figure 5-3. Bainbridge participant demographics



5.1.2 Opportunities to grow program

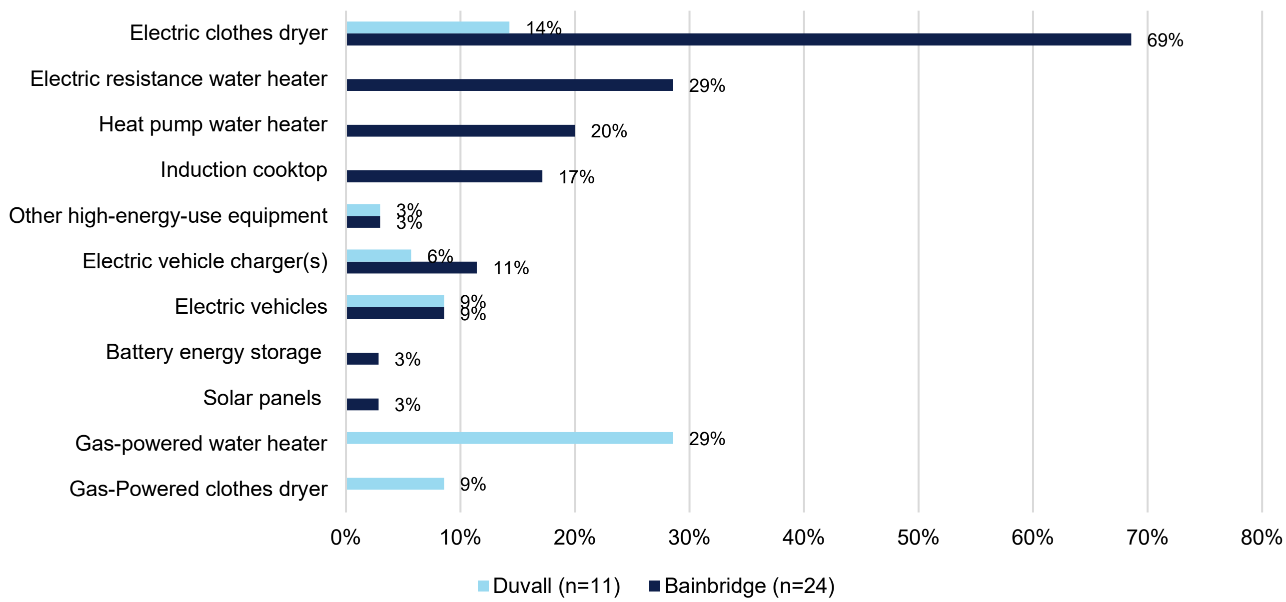
The customer survey shed light on ways to increase program participation. Our first finding from the survey was that most participants had a smart thermostat installed in their home before they learned about the program, as shown in Figure 5-4. In the future, PSE may want to target customers who do not currently have a smart thermostat, as they represent an untapped population of potential enrollees. Alternatively, or additionally, PSE could add a switch program as a DR program offering, which attaches a switch to a heating device, causing that device to reduce its runtime during event calls. This type of program would work as a participation option for customers who cannot or do not want to install a smart thermostat.

Figure 5-4. Percent of respondents who owned a smart thermostat before learning about the program (n=35)



In addition to increasing thermostat DR impacts, other devices can be curtailed through either expanded DR programs or through behavioral programs like time-of-use rates or behavioral event alerts. Many participants in this program have other technologies that would be good candidates for DR or behavioral programs. PSE is already working to expand the inclusion of water heaters in its Targeted Demand Side Management (TDSM) DR programs, and we assume that close to 100% of customers have water heaters. In Duvall, we see (Figure 5-5) that all participants who chose a water heater type said they had a gas water heater, so this is a promising gas curtailment opportunity. In Bainbridge, managed electric vehicle charging represents another opportunity to shift electric load away from morning hours. Equipment that could be influenced by a time-of-use (TOU) rate in Bainbridge includes electric clothes dryers and induction cooktops; in Duvall, this includes gas-powered clothes dryers.

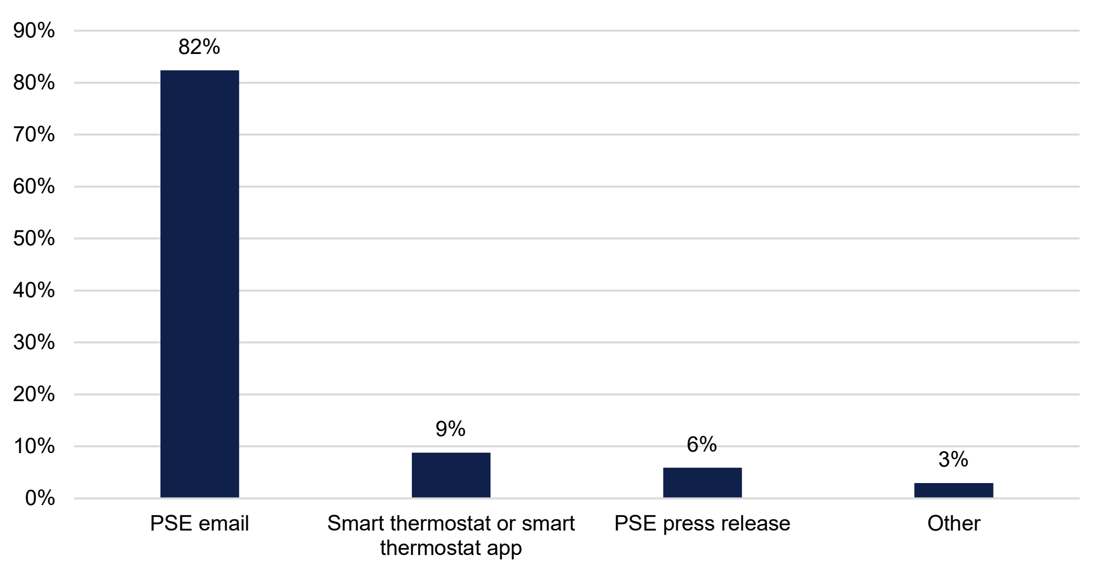
Figure 5-5. Respondent-reported existing high-energy-use equipment in the home (n=35)



"Other high-energy-use equipment" reported by two respondents include a hot tub and a gas range.

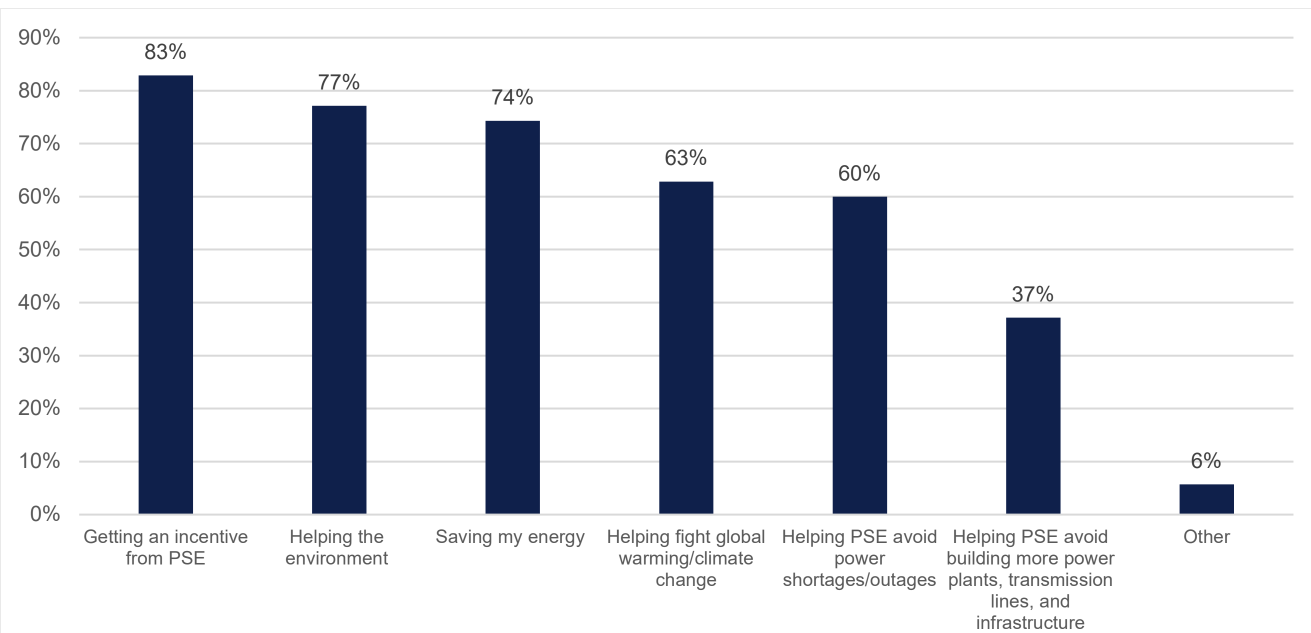
Effective program marketing is crucial to program adoption. In interviews, PSE staff suggested that the community in Bainbridge engages in spreading the program message. However, most respondents learned about the program via email and one participant said that they had never heard of the program before talking to a PSE representative. Additional messaging via community channels, smart thermostat apps, social media groups, customer bills, and home energy report messaging is likely to help expand program awareness.

Figure 5-6. Ways in which respondents initially heard about the program (n=35)



Understanding customers' reasons for enrolling will also help PSE craft program information (and incentives) that encourage enrollment. Respondents indicated that incentives and environmental reasons are their two primary reasons for enrolling in the program. Emphasizing these points in program advertising material could increase the number of participants recruited. Participants, especially in Duvall, seemed not to be motivated by the ability to reduce infrastructure costs. This could either indicate that messaging should be clearer about why the program is needed or that this should be de-emphasized because it is not motivating for potential participants.

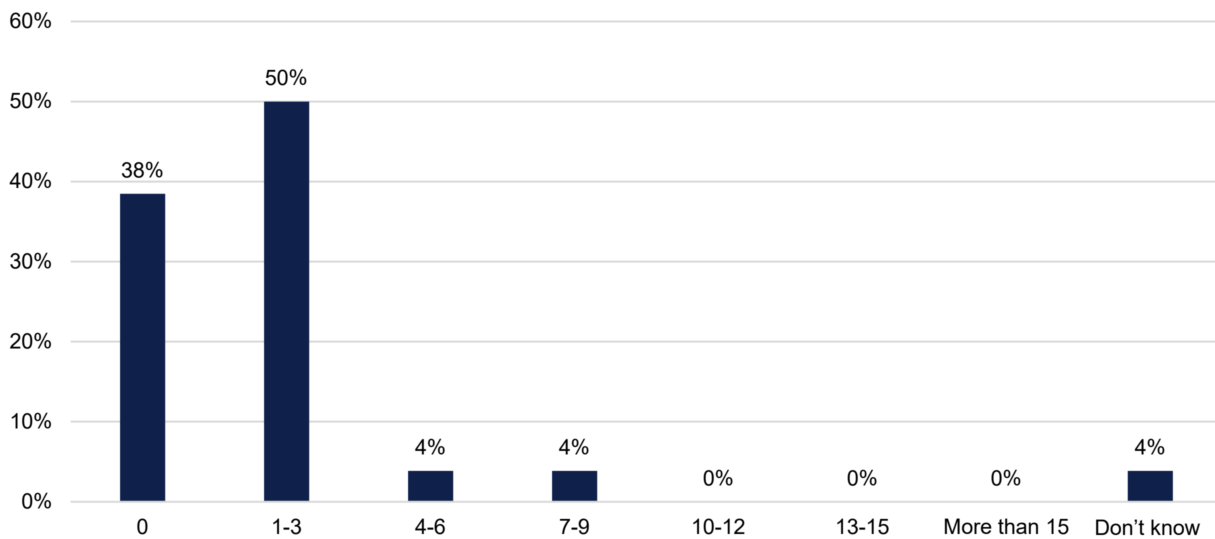
Figure 5-7. Primary motivations for participating in the program (n=35)



5.1.3 Opportunities to increase the impact per event

Having high per-participant event impact is also key to program success. Having few overrides can contribute to much higher event impacts. As shown in Figure 5-8., participants self-reported overriding their thermostats on relatively few event days, contributing to higher impact. However, some participants reported overriding every event. As the program matures, PSE should consider reaching out to these high-override participants.

Figure 5-8. Number of times someone in the household overrode the remote adjustment during DR events (n=26)



Another way to maximize event impact is to ensure no one enrolled in the Duvall program has electric heat, as participants with electric heat do not contribute to the goal of gas curtailment in Duvall. One of 12 respondents from Duvall reported having electric-only heat. This trend should continue to be tracked. We suggest options for reducing this incidence in Section 5.2.5.

Finally, as part of program outreach, PSE could consider including messaging to educate participants about strategies to stay warm. About 45% of Duvall participants had no strategy to stay warm during events. Duvall participants also had a higher reported override rate. Education on strategies to stay warm could reduce this override rate and increase the program per event.

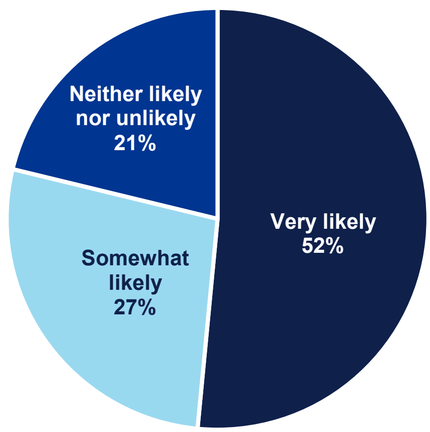
5.1.4 Opportunity to learn about long-term program load impacts

About half of the survey respondents reported that they adjusted their thermostats for all days (including non-event days) since they enrolled in the program. It may be informative for PSE to examine the long-term effect of the program (pre-enrollment vs post-enrollment non-event days) to understand how participation is affecting load on non-event days. Participants may have realized they need less morning heat and have turned down their morning thermostats or could have turned up morning heat to prepare for heat shut offs during events.

5.1.5 Opportunities to increase satisfaction

Overall, customers are largely satisfied with the program. This means that they are not only happy with PSE (reducing staff challenges), but more likely to stay in the program and continue contributing to curtailment in coming years. For example, when asked how likely they are to recommend the program to someone they know, most respondents (52%) said they were highly likely to recommend the program (Figure 5-9). None of the respondents said they were unlikely to recommend the program.

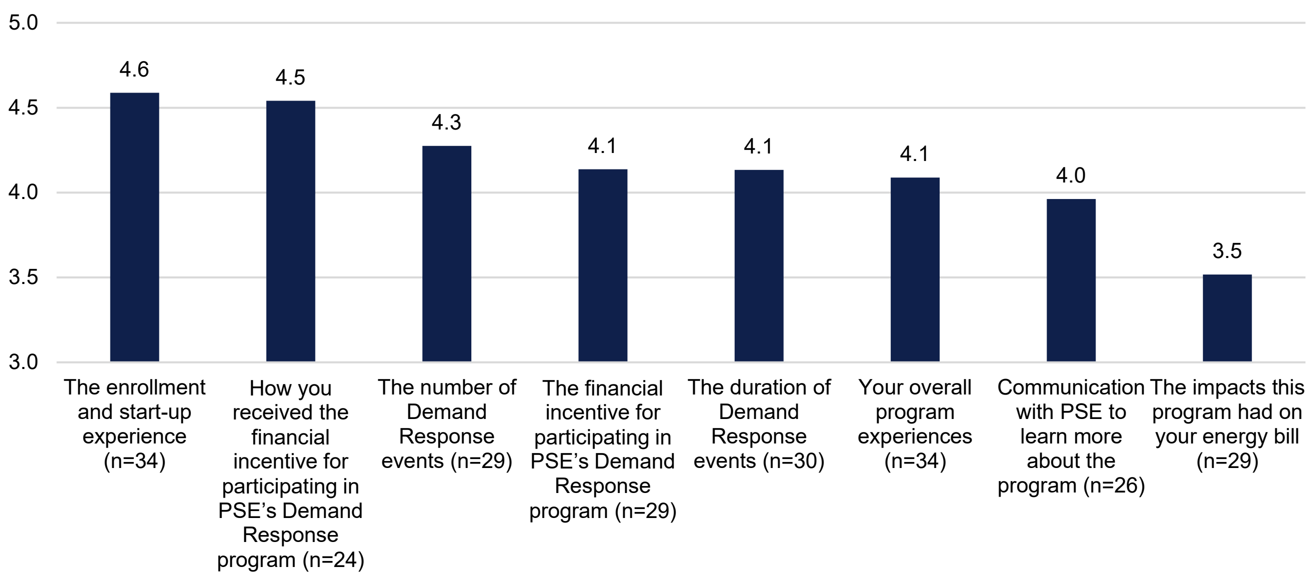
Figure 5-9. Likelihood to recommend the program to someone the respondent knows (n=33)



Respondents were asked to rate their satisfaction with various components of the DR program on a scale of 1-5, with 5 being very satisfied, 4 being somewhat satisfied, 3 as neutral or too soon to tell, 2 being somewhat dissatisfied, and 1 being very dissatisfied.

Figure 5-10 shows the average rating for each component. The top-rated categories were the enrollment and start-up experience and how they received their financial incentive. The lowest-rated category was the impact the program had on their energy bills. This is likely due to the fact that some participants enrolled in the program without being informed that the program has minimal impact on their bills.

Figure 5-10. Average satisfaction rating for various program components



While customers were highly satisfied overall, many did include recommendations for improving the program, and some themes emerged from those recommendations:

- Some respondents appeared to think that the program would save energy and money on their monthly bills. However, the program is not likely to change customer bills. It may be worth explaining in sign-up material that the program is not expected to change bills one way or another. One customer suggested that they would prefer a time of use rate, as this would help shift load and save money.
- Many respondents noted that they would prefer earlier event notifications that were sent at least a day ahead of events. One customer noted that she/he would prefer to have notifications the day before even if some were false alarms. Nearly all respondents who commented said they would prefer email and text notifications.
- Many respondents noted issues with pre-heating. One noted that their heat pump is designed to work with an auxiliary system, so rapid pre-heating in the hour before the event could trigger the auxiliary system, whereas slower and longer pre-heating should be employed for heat pumps. Another noted that they would also prefer a longer pre-heating period because they have an in-floor heating system that needs more than six hours to change the home temperature. A third respondent noted that their household has a nighttime temperature setback, so the pre-heating is based on this nighttime temperature, not the temperature they need to be comfortable once they wake up in the morning. This means that they are too cold during the event and often need to override. PSE may want to consider helping customers customize their pre-heating duration and temperature, which would help these customers participate fully and comfortably in events.

5.2 Program staff interviews

DNV conducted four in-depth interviews with PSE staff involved in the DR program. These interviews were designed to improve our understanding of program organization, program goals, the VPP platform, M&V estimates, program DR events and results, challenges, and successes. We present findings and recommendations from the interviews in the sections below.

5.2.1 Program organization and staff understanding

Through our discussions with PSE staff, we found that staff across departments have good knowledge of the program design and its goals. However, PSE may want to consider clarifying and solidifying the staff organization around the VPP program, as interviewees reported high levels of internal staff turnover and group dissolution, making it challenging to work efficiently.

Additionally, PSE may want to consider working toward centralized leadership for the suite of demand flexibility programs that will be called through its VPP platform. Currently, upward of five departments are providing input on what the VPP platform should do. Centralized VPP leadership could provide clear VPP team priorities and determine the order in which programs should be incorporated into the platform.

Central VPP leadership would also allow easier collaboration with external groups to help Washington and the region account for DR resources as capacity in the market, which could help make the VPP more successful from a big-picture perspective. A leadership group could also advocate to industries that manufacture products with demand flexibility potential, helping design technologies that have DR control capability. One example is mini-split ductless heat pumps, for which there is currently no easy control mechanism for DR programs. Original equipment manufacturers (OEMs) could also be encouraged to align enrollment practices and make program qualification more logical and flexible (qualification is currently based on thermostat wiring and does not always make sense for winter or year-round programs). A third external effort the PSE team could pursue is collaborating and communicating with similar utilities about best practices in DR programs and using information from other established DR programs to understand the potential (both in per-person curtailment and in percent adoption) of programs that PSE is launching or planning to launch.

The ad-hoc program organization as the VPP rapidly scales up into a functional program may also be leading to some communication breakdowns and divergent goals between groups at PSE. For example, the gas distribution team reports that significant DR is needed in Duvall to avoid needing additional equipment and hires, a loss of gas pressure, or an additional pipeline. In contrast, some members of the DR team seemed to think the program was designed more experimentally, just to see what they could get, and did not have ccf goals. They had modest participant number goals. The gas team, in contrast, wanted as many participants as possible to avoid the system issues cited above.

In addition to different goals, the gas team seemed to have less knowledge of the program status and plans. The gas team reported not knowing the number of current program enrollees or the number of events planned for the coming winter. They also expressed a need for accurate M&V delivered to them instead of having to perform calculations themselves and the ability to control events. However, in other interviews, DNV learned that the gas team will be expected to wait 1-2 years before these needs are met. It is unclear if this has been communicated to the gas team. While gas DR is not a long-term system-wide need, given the importance of gas DR to Duvall and similar gas-constrained areas, systems should be established to make it an effective program. Because of these issues, PSE should consider having a clear plan and goals for gas DR. Relatedly, PSE should consider having a clear mechanism for communications and information transfer about the gas DR program to the gas distribution team, a clear and simple way for that team to call DR events, and rapid visualization of curtailment estimates.

5.2.2 Program goals

Through our interviews, we found that the program has clear goals in participant numbers for the next four years and for MW curtailed by 2029. However, it is not clear whether those goals are specific to automatic smart thermostat DR or all DR. If the latter, PSE may want to consider creating goals specific to each DR program. Additionally, PSE may consider attaching MW curtailment goals to each year and program to make sure it is on track to hit its total 2029 goals, which are ultimately MW/ccf needs. These program-specific goals are especially important because the different DR programs are known to be associated with quite different per-customer curtailment (for example, automated smart thermostats, water heaters, and behavioral DR) and because PSE's participation goals for both Bainbridge Island and Duvall are ambitious.

Additionally, PSE should consider setting a gas curtailment goal, which could be based on the need for gas curtailment in Duvall in a modeled design-day scenario.

5.2.3 Virtual power plant platform

PSE's VPP platform is impressive. Its continuous improvements, strong leadership, schedule of dispatch, and ease of changing event timing are all necessary features for the VPP to be successful. PSE has also automated the processing around enrollments, a key piece to creating a smooth VPP with diverse OEMs (and, in the near future, technologies). Significant effort was dedicated to ensuring customer sign-ups are as easy as possible. As illustrated in Section 5.1.5, materials providing guidance, the program website, phone support, and other program support all seem to have been well done and well-received overall.

5.2.4 Measurement and verification

Regarding measurement and verification, all groups that DNV interviewed expressed some level of frustration with Autogrid's estimation of program impacts. Staff seemed to have a difficult time finding the existing Autogrid estimates of program impact. PSE staff would like the Autogrid impact estimates to be included in a dashboard in the Flex system, but that is not yet the case. Both the electric and gas planning teams manually created their own estimates of event baselines, a time- and effort-intensive process, because they are skeptical of Autogrid's methodology. They cited, among other issues, a lack of consideration of temperature differences on event days. Some PSE staff reported that Autogrid had provided gas evaluation reporting, but others said that Autogrid was unable to handle gas meter data, so processed it incorrectly in combination with electric meter data, not considering differences in gas and electric unit measurements. DNV also observed this in the M&V data we received from Autogrid.

As noted above, PSE appears to plan no automated M&V for gas DR for another one to two years. The long-term plan is to have a separate Autogrid Flex system to collect data used for Autogrid's M&V. Neither of these plans align with the desires of the gas team, who are not satisfied with doing their own M&V and do not feel Autogrid's methodologies for calculating savings are reliable. PSE may want to bring the VPP, DR, and gas teams together to find a solution for M&V that works for all parties and balances the value of gas DR with the effort needed for gas DR M&V. Additionally, PSE may want to consider a central, easily accessible location for reporting Autogrid's M&V estimates. This could also include clear documentation of Autogrid's methods and the strengths and weaknesses of these methods. This could take the form of an internal dashboard or simply an internal location for files that are accessible to all invested parties at PSE.

As an alternative to Autogrid's estimates, PSE may want to consider creating a system for rapid third-party trusted M&V. Staff specifically mentioned the desire for post-event or even quick-turnaround post-season accurate third-party M&V. A system is already set up for rapid AMI transfer to Autogrid after each event, so with continued effort and communication, AMI transfers to a third-party evaluator should be possible.

5.2.5 Event calls and results

There are several actions that PSE may want to consider as it works to call events at the right time and have the most impactful programs possible. PSE is actively working toward establishing criteria for calling electric DR events, which is an important step as the program matures. It could also be important to do the same for gas events, which may start with determining if gas events should have different criteria than electric events, or if it makes sense to continue calling them at the same time.

Additionally, as PSE continues to grow its DR enrollment, it will be able to (because of Nest rules[6]) increase stratification of customers for event calls (by zip code, etc.). PSE should consider beginning to plan for its clearest needs for stratification and documenting those needs to prepare for event calls in different scenarios.

There are also several ways that PSE can work to ensure the highest benefit for the grid and grid infrastructure. First, PSE may want to plan for the spikes associated with pre-heating and snapback (before and after the event). The rapid ramp from pre-heating to curtailment may be especially abrupt for the grid. This could require the staggering of different customer dispatch within each geography.

Second, before determining if a customer is eligible for enrollment, PSE may want to consider validating each customer's fuel type. For Bainbridge, which is electric only, this is not needed, but for Duvall and all other areas with dual fuel service, it could vastly improve the cost-effectiveness of the program. Homes with gas heat are not going to contribute to electric curtailment, and vice-versa. A simple algorithm could be designed to pull either billing or AMI data for a customer and determine the likely heat source for that home. Alternatively, though it would be more time-intensive, staff could manually identify heat sources from photos taken by customers. In either case, if PSE lacks the staff to implement either measure, as suggested in interviews, a third party could perform these tasks.

[6] Nest only allows certain stratifications if a large enough number of customers are enrolled.

Finally, given concerns about the optics of lack of equity around these programs, PSE could consider an opt-out time-of-use rate for these geographies, which would incentivize load shift and potentially reduce the incentive amounts needed to reach its curtailment goals. PSE has expressed concern that allowing time of use rate enrollment for Bainbridge and Duvall customers could cloud its understanding of the other TDSM effort impacts. However, if PSE eventually rolls out this rate for all PSE customers, it may be more important to understand the impact of TDSM efforts for customers on time of use rates. PSE is already rolling out a suite of programs for the TDSM effort, and enrolling a customer in time of use rates in addition to other programs may not be much different than enrolling a customer in multiple TDSM programs that are currently offered.

5.3 Documentation review

To better understand internal processes, DNV reviewed documentation used as reference by PSE staff. We also reviewed outreach materials (newsletters and direct email outreach) to gain insight into how customers are hearing about the program and what information they are given regarding enrollment and benefits to participating.

5.3.1 Internal program documentation

PSE provided a process flow chart illustrating the SysOps forecast aggregation dispatch process flow. The chart is detailed with teams involved, the tools used at each step, and decision points that affect process flow. Using the existing flow chart, PSE staff should continue to prioritize their effort to document how the programs work, their goals, etc. for all teams involved and as the Peak Time Rebate program evolves. This effort is in-progress and will be important as the number of PSE staff affected by the programs grows and as there is turnover in program staff.

5.3.2 Newsletters

The newsletters to Bainbridge Island customers regarding the Peak Energy Rewards program are engaging and informative. The initial newsletter announcing the program in November 2022 featured Peak Energy Rewards at the top of the publication, which likely caught the eyes of customers who routinely read the newsletter. In the subsequent newsletters, the sections specific to the program are different every month (location in the newsletter, announcing events specific to the program), which helps to keep the program information interesting without becoming repetitive.

5.3.3 Email outreach

The initial email launch for both Bainbridge and Duvall were well designed and provided detailed yet succinct information that customers need to know before enrolling. Because most customers found out about the program through the PSE email and likely enrolled through the link in the email, PSE should continue this mode of outreach to customers as the program grows and expands.

6 FINDINGS AND RECOMMENDATIONS

This concluding section of the report documents DNV's findings, recommendations, and considerations associated with this program. Overall, we found a well-run DR pilot program that functioned as intended in its first year. Below, we first summarize our results on peak demand impacts and drivers of differential impacts. We then move to key findings from the process evaluation, which was based on customer surveys, PSE staff interviews, and document review. We organize our results around the evaluation goals from the scope of work. We will address Evaluation Objective 2, providing a recommended settlement method, in our 2024 evaluation of the 2023-2024 winter season.

6.1.1 Impact findings and recommendations

In this section we describe findings from the quantitative evaluated curtailment from the program.

1) Evaluation Objective 1: Independent estimate of demand reduction

Finding: The average participant load curtailment achieved by the program to date is 1.59 kW for electric participants and 0.106 ccf/hr for gas participants across all events. Curtailment was highest in the first hour of events and declined in the second and third hours. The evaluated curtailment was quite different from that calculated by Autogrid using the 10-of-10 settlement baseline. DNV's impact estimates were consistently positive, and always greater than 1 kW/participant for electric and 0.06 ccf/hr for gas. Autogrid's estimates were consistently smaller than DNVs and many were negative, indicating a load increase when the event was called. We discuss PSE's challenges with Autogrid's M&V estimates in more detail below and provide recommendations for future years.

2) Evaluation Objective 4: Demand impacts outside of called event periods

Finding: Participant load tends to increase in the hour before events and in the two hours after events. Pre-heating in the hour before the event accounted for an average load increase of 0.73 kW per customer for electric participants and 0.079 ccf/hr per customer for gas participants. Snapback heating in the two hours after the event accounted for an average load increase of 0.84 kW per customer for electric participants and 0.089 ccf/hr per customer for gas.

3) Evaluation Objective 3: Recommendations to achieve future DR goals

Finding: Curtailment varied by smart thermostat manufacturer, with Nest having the highest estimated curtailment, Ecobee having much less curtailment, and no significant event impact for Mysa thermostats.

- **Recommendation** – PSE may want to consider both working with the thermostat manufacturers to see if they are able to increase curtailment. If thermostat manufacturers are not able to increase curtailment, PSE could restrict enrollment or reduce incentives for those thermostats that provide less or no curtailment. Next year's evaluation should provide additional evidence for the difference in thermostat brand effectiveness in curtailment.

6.1.2 Process findings and recommendations

In this section, we discuss key findings and recommendations from the staff interviews, customer surveys, and material reviews.

1) Evaluation Objective 3: Recommendations to achieve future DR goals

Finding: DNV found that most program participants already had smart thermostats when they enrolled.

- **Recommendation** – PSE may want to consider actively recruiting customers who do not yet have smart thermostats and educating them on the benefits of these devices. Additionally, PSE could offer an alternative program for customers who do not want to have a smart thermostat. This program, like DR programs before the advent of smart device technologies, could rely on switches installed on heating equipment that would reduce the equipment's runtime during called events.

Finding: Most program participants also reported having a variety of other devices that could be controlled as part of DR programs to reduce electric and gas load for Bainbridge and Duvall, respectively.

- **Recommendation** – PSE could consider more active recruitment for the water heater device part of the program, and expanding its programs so that other controllable devices could enroll.

Finding: Many program participants suggested that pre-heating was not working well for them, causing increased overrides.

- **Recommendation** – PSE may want to consider longer pre-heat periods that would work better for heat pumps and pre-heating that works for customers with a nighttime setback. DNV can also ask more targeted pre-heating questions in the customer survey next year to provide more detailed recommendations on pre-heating.

Finding: PSE staff and customer surveys suggested that not all gas DR program participants have gas heating systems.

- **Recommendation** – Program managers could consider a simple validation algorithm for heating equipment, using AMI data.

2) Evaluation Objective 5: Opportunities for program improvement

Finding: Most participants reported finding out about the program in an email, while program staff reported putting a lot of effort into community outreach and a high level of community engagement.

- **Recommendation** – While the customer survey after the winter 2023-2024 program season will shed some light on a greater number of participants over more time, staff should ensure that their efforts are translating into program sign-ups.

Finding: Most program participants report high satisfaction and program staff describe putting a lot of effort into enrollment materials to make enrollment and participation as smooth and simple as possible for all types of participants.

- **Recommendation** – PSE should continue its excellent customer service in the delivery of the program.

Finding: Some customers reported that they would prefer earlier program notifications, with all notifications received at least the day before the event. Some even said they would prefer more false notifications for called events that are later cancelled if notifications could be earlier.

- **Recommendation** – PSE should continue to study this preference in the customer survey after the next program season and send notifications as early as possible.

3) Evaluation Objective 6: Assess integration of program with PSE grid operations

Finding: In staff interviews, DNV heard that VPP staffing organization is rapidly developing and has many stakeholders, leading to decentralized leadership and lack of staffing continuity. This finding affects all VPP stakeholders, including grid operations, and problematic communication is most apparent with the gas operations group, who seemed unaware of VPP operational plans.

- **Recommendation** – PSE should consider centralizing VPP leadership, reducing staff turnover, and setting up regular communications with grid and gas operations staff.

Finding: PSE grid operations staff are generally up to date with program understanding and are working towards a standard practice for determining when events should be called.

- **Recommendation** – PSE grid operations staff should continue their engagement with VPP efforts.

Finding: M&V is a consistent area of frustration for VPP platform staff, program managers, grid operators, and the gas operations group. Autogrid's M&V results are hard to find, not integrated in their Flex system, and the M&V methods they use are unclear. PSE’s grid and gas operations teams do not feel that Autogrid’s M&V results are accurate and reliable and have spent considerable time and effort producing their own results.

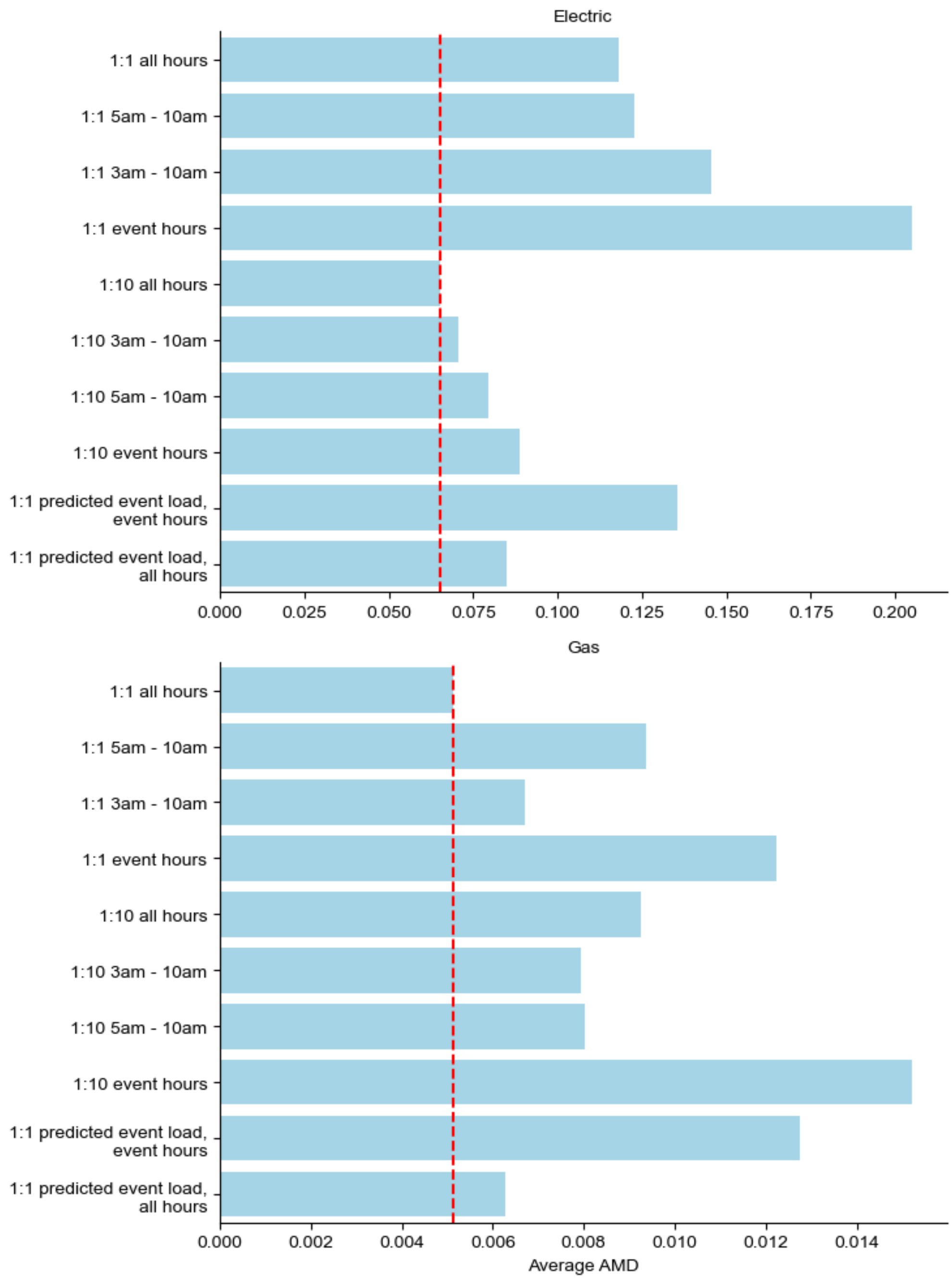
- **Recommendation** – PSE may want to consider a system for rapid AMI transfer to a third-party evaluator to ensure accurate, rapid evaluation of event impacts.

7 APPENDICES

7.1 Appendix A: Matching methods

The figure below shows the average AMD of each matching method, averaged across all event days. The red line shows the value of the best performing matching method.

Figure 7-1. Matching scores (AMD) of different match approaches



7.2 Appendix B: Demand Response Program participants web survey guide

Memo to: Kasey Curtis, Puget Sound Energy **From:** Chelsea Liddell, Katie Ryder, Andrew Wood, DNV

Date: May 30, 2023

PSE: Demand Response Program Participants Web SURVEY GUIDE

7.2.1 Interview guide overview

Objective: The Evaluation Team will conduct web surveys with PSE Demand Response program participants to gain insight into their participation, behavior, decision factors, and satisfaction.

Anticipated timing (survey length): Approximately 15 minutes

Method of data collection: Web Survey

Table 1: Research Objectives Mapped to Questions in this Instrument

Question	Instrument Goal
Q1	Screening
Q2 - Q6	Background
Q7 - Q9	Participation/Decision Factors
Q10 - Q24	Outcomes/Satisfaction
Q25 - Q35	Event Experiences
Q36 - Q45	Household Characteristics
Q46 - Q48	Closing

When the term “Demand Response Event” shows up in the survey, we will include the following hover text:

HOVER TEXT: Demand Response Events may occur when the weather is very cold and energy demand may spike. On those days, we work with your smart thermostat to heat your home before the predicted energy spike, then reduce your use a little during the Demand Response Events themselves, flattening the energy spike and keeping energy prices down

7.2.2 Survey Guide

Table 2: Overview of Data Collection Approach

Data Collection	Description
Population Description	PSE Demand Response program participants
Sampling Method	Census
Instrument Type	Web Survey
Survey/Interview Length	Approximately 15 minutes
Description of Contact Sought	Those who are currently participating in the PSE program.

Table 3: Database Information Piped into Survey Instrument

Database Information Piped into the Survey Instrument	Variable Description
Name	Contact's first name

Email Invitation Template

[FROM] Puget Sound Energy

[SUBJECT]: PSE Demand Response Program End of Season Survey

Hello **[PIPE IN FROM DATA: Name]**,

You are invited to participate in Puget Sound Energy's Demand Response program end of season survey!

Puget Sound Energy is committed to providing its customers with products and services designed to service customers like you. As part of this effort, we are conducting an evaluation of the Demand Response program. As a participant in PSE's program, your opinions are important. PSE would like your input and perspectives to understand how to best structure this program in the future for customers like you.

To get started, click on this link: [INSERT LINK]

It will take approximately 15 minutes to answer our questions. Your individual responses will be kept confidential and anonymous. Any analyses will not identify individuals.

To thank you for completing this survey! If you complete the survey before [DATE], you will be entered into a drawing for a $200 Amazon gift card. If you complete the survey after [DATE], you will be entered into a drawing for a $100 Amazon gift card.

Thank you for your participation.

If you have any questions about this research effort, please contact the PSE Evaluation and Research Group at EESEvaluations@PSE.com.

Thank you for participating in PSE's program evaluation. We really appreciate your input!

Puget Sound Energy
355 110th Ave NE
Bellevue, WA 98004

7.2.2.1 Introduction

[DISPLAY FOR ALL RESPONDENTS]

Welcome to the PSE Demand Response program survey!

You have been selected to participate in this important survey because our records indicate that you participate in the Demand Response program.

This survey should take approximately fifteen minutes to complete. Your participation will help PSE better understand this program and improve all our programs.

The analysis will only use summary level data and will not identify individual respondents. Your responses will be kept confidential and anonymous.

To begin the survey, click on the arrow below.

7.2.2.2 Screening

[ASK ALL RESPONDENTS]

Q1. According to our records, you are participating in PSE's Demand Response program, which adjusts your thermostat(s) during Demand Response Events. Demand Response Events occur when the weather is cold and energy demand is high. On these days, we work with your smart thermostat to heat your home before the predicted energy spike, then reduce your use during Demand Response Events, flattening the energy spike and keeping energy prices down. Are you aware of your participation in this program?

[SINGLE RESPONSE, FORCE RESPONSE]

1. Yes **[PASSED SCREENING]**
2. No **[TERMINATION SCRIPT 1]**

[PASSED SCREENING]: Great! You are eligible to take our survey. Let's get started.

[TERMINATION SCRIPT 1]: Thank you for answering our questions. However, we are looking for respondents who are participating in the Demand Response program. Your response has been recorded. Have a great day.

7.2.2.3 Background

[ASK ALL RESPONDENTS]

Q2. What brand(s) of smart thermostat(s) is installed in your home?

[MULTIPLE RESPONSE, RANDOMIZE 1-4]

1. Ecobee
2. Google Nest

3. Mysa
4. Skycentrics
5. Other (please specify)

[ASK ALL RESPONDENTS]

Q3. Did you have a smart thermostat(s) installed in your home before learning about the Demand Response program?

[SINGLE RESPONSE]

1. Yes
2. No

[ASK ALL RESPONDENTS]

Q4. What, if any, of the following heating/cooling technologies do you have in your home? Please select all that apply.

[MULTIPLE RESPONSE, RANDOMIZE 1-7]

1. Gas furnace
2. Electric resistance heating
3. Heat pump for heating and/or cooling
4. Gas fireplace insert(s)
5. Wood-burning fireplace
6. Wood-burning stove
7. Electric space heating

97. Other home heating equipment, please specify: __________ **[INSERT OPEN-ENDED RESPONSE]**

8. None of the above

[ASK ALL RESPONDENTS]

Q5. What, if any, of the following technologies do you have? Please select all that apply

[MULTIPLE RESPONSE, RANDOMIZE 1-10]

1. Electric vehicles
2. Electric vehicle charger(s)
3. Induction cooktop
4. Gas-powered clothes dryer
5. Electric clothes dryer
6. Gas-powered water heater

7. Electric resistance water heater
8. Heat pump water heater
9. Solar panels
10. Battery energy storage
97. Other high-energy-use equipment, please specify: __________ **[INSERT OPEN-ENDED RESPONSE]**
11. None of the above

[ASK IF YES TO Q5, ANY OF 1-10 and 97]

Q6. Approximately when did you begin using the following technologies

[ASK FOR EACH TECHNOLOGY SELECTED IN Q5]

1. Before November 2021
2. November 2021 – October 2022
3. November 2022 –April 2023
4. Don't remember
98. Don't know

7.2.2.4 Participation/Decision Factors

[ASK ALL RESPONDENTS]

Q7. Thinking back, how did you find out about the Demand Response program?

[SINGLE RESPONSE, RANDOMIZE 1-5]

1. PSE press release
2. PSE email
3. Program website
4. Smart thermostat or smart thermostat app
5. Neighbors, family, friends, colleagues, or a community organization
6. Do not remember
97. Other, please specify: ________________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK ALL RESPONDENTS]

Q8. What were the main reasons why you chose to participate in the Demand Response program? Please select all that apply.

[MULTIPLE RESPONSE, RANDOMIZE 1-6]

1. Getting an incentive from PSE

2. Helping PSE avoid power shortages/outages
3. Helping fight global warming/climate change
4. Helping the environment
5. Saving my energy
6. Helping PSE avoid building more power plants, transmission lines, and infrastructure

97. Other, please specify: ___________ **[INSERT OPEN-ENDED RESPONSE]**

98. Don't know

[ASK IF Q8 = 1-6 OR 97]

Q9. What was the most important reason why you chose to participate in the Demand Response program?

[SINGLE RESPONSE, INSERT SELECTED OPTIONS 1-6 AND ENTERED TEXT FOR 97 FROM Q8]

7.2.2.5 Outcomes/Satisfaction

[DISPLAY FOR ALL RESPONSENTS]: The next section will ask you questions about your experiences with PSE's Demand Response program.

[ASK ALL RESPONDENTS]

Q10. Please indicate how strongly you agree or disagree with the following statements about PSE's Demand Response program.

[INSERT "STRONGLY AGREE", "SOMEWHAT AGREE", "NEITHER AGREE NOR DISAGREE", "SOMEWHAT DISAGREE", "STRONGLY DISAGREE", AND DON'T KNOW CHOICES FOR ALL OPTIONS, RANDOMIZE STATEMENTS]

1. This program is relevant to me
2. The benefits of this program (energy savings/environmental benefits) are important to me
3. The costs for this program were clear
4. Participating in this program is easy
5. Support for the program was readily available
6. This program aligns with my values

[ASK ALL RESPONDENTS]

Q11. Please indicate your level of satisfaction with each of the following:

[INSERT "VERY SATISFIED", "SOMEWHAT SATISFIED", "NEUTRAL/TOO SOON TO TELL", "SOMEWHAT DISSATISFIED", "VERY DISSATISFIED", NOT APPLICABLE, AND DON'T KNOW CHOICES FOR ALL OPTIONS, RANDOMIZE STATEMENTS]

1. The enrollment and start-up experience
2. Communication with PSE to learn more about the program

3. The financial incentive for participating in PSE's Demand Response program
4. How you received the financial incentive for participating in PSE's Demand Response program
5. The number of Demand Response events **[INSERT HOVER TEXT]**
6. The duration of Demand Response events **[INSERT HOVER TEXT]**
7. The impacts this program had on your energy bill
8. Your overall program experiences

[ASK FOR ANY ITEMS IN Q11 WITH SOMEWHAT OR VERY DISSATISFIED RESPONSE]

Q12. Do you have any suggestions for how we can improve the following program element(s) for which you indicated dissatisfaction?

[INSERT OPEN-ENDED RESPONSE OPTION FOR ITEMS IN Q11 WITH VERY DISSATISFIED RESPONSE]

[ASK ALL RESPONDENTS]

Q13. Did you experience any issues during the season that led you to seek help?

[SINGLE RESPONSE, FORCE RESPONSE]

1. Yes
2. No

[ASK IF Q13 = 1]

Q14. Was the issue(s) successfully resolved?

[SINGLE RESPONSE, FORCE RESPONSE]

1. Yes
2. No

[ASK IF Q14 = 1]

Q15. Which resource proved to be the most helpful for resolving your issue?

[SINGLE RESPONSE]

1. Visiting the program website or FAQs
2. Emailing customer support
3. Calling customer support

97. Other, please specify: _________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK IF Q14 = 2]

Q16. We are sorry you experienced an issue(s) during the season that was not successfully resolved. Do you have any feedback you want to provide on the issue(s) and how it could have been resolved?

[SINGLE RESPONSE]

1. No
2. Yes, please specify: _________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK ALL RESPONDENTS]

Q17. Since joining the Demand Response program, have you changed the typical temperature setting that you use for cooling your house?

[SINGLE RESPONSE]

1. Yes, I have lowered my thermostat(s) setting
2. Yes, I have raised my thermostat(s) setting
3. Yes, I have raised and lowered my thermostat(s) setting on different days or times of day
4. No, I have not changed my thermostat(s) setting

97. Other, please specify: _________ **[INSERT OPEN-ENDED RESPONSE]**
98. Don’t know

[ASK ALL RESPONDENTS]

Q18. Is there anything you would change about PSE’s Demand Response thermostat program?

[SINGLE RESPONSE]

1. No
2. Yes, please specify: ________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK ALL RESPONDENTS]

Q19. How likely are you to participate in PSE’s Demand Response program in future seasons?

[SINGLE RESPONSE]

1. Very likely
2. Somewhat likely
3. Neither likely nor unlikely
4. Somewhat unlikely
5. Very unlikely

[ASK IF Q19Q19 = 4 OR 5]

Q20. Is there anything PSE’s Demand Response program could improve to make you more likely to participate in future seasons?

[SINGLE RESPONSE]

1. No
2. Yes, please specify: ________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK ALL RESPONDENTS]

Q21. How likely are you to recommend PSE's Demand Response program to someone you know?

[SINGLE RESPONSE]

1. Very likely
2. Somewhat likely
3. Neither likely nor unlikely
4. Somewhat unlikely
5. Very unlikely

[ASK ALL RESPONDENTS]

Q22. Do you believe your participation in this program helps support PSE's Beyond Net Zero Carbon goal?

[SINGLE RESPONSE]

1. Yes
2. No
3. I am not aware of PSE's Beyond Net Zero Carbon goal

98. Don't know

[ASK IF Q22 = 2]

Q23. Please explain why you don't believe your participation in this program supports PSE's Beyond Net Zero Carbon goal.

[INSERT OPEN-ENDED RESPONSE]

[ASK ALL RESPONDENTS]

Q24. As a result of participating in this program, how has your opinion of PSE changed? Is it…

[SINGLE RESPONSE]

1. Much more positive
2. Somewhat more positive
3. Neither more positive nor more negative
4. Somewhat more negative
5. Much more negative

7.2.2.6 Event Experiences

[DISPLAY FOR ALL RESPONSENTS]: Next, we will ask you questions about your experiences with Demand Response Events. **[INSERT HOVER TEXT]**

[ASK ALL RESPONDENTS]

Q25. Since enrolling in this program, do you recall any times when PSE remotely adjusted your smart thermostat(s)?

[SINGLE RESPONSE, FORCE RESPONSE]

1. Yes
2. No

[ASK IF Q25 = 1]

Q26. About how many times do you think PSE remotely adjusted your smart thermostat(s)?

[SINGLE RESPONSE]

1. 1-3
2. 4-6
3. 7-9
4. 10-12
5. 13-15
6. More than 15

98. Don't know

[ASK IF Q25 = 1]

Q27. How clear were the Demand Response Event **[INSERT HOVER TEXT]** notification(s)?

[SINGLE RESPONSE]

1. Extremely clear
2. Very clear
3. Somewhat clear
4. Slightly clear
5. Not at all clear

[ASK IF Q25 = 1]

Q28. Do you have any suggestions for how we can improve Demand Response Event **[INSERT HOVER TEXT]** notification(s)?

[SINGLE RESPONSE]

1. No
2. Yes, please specify: ________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK IF Q25 = 1]

Q29. How strongly do you agree or disagree with this statement: Demand Response Event **[INSERT HOVER TEXT]** notifications gave me enough advance notice about the upcoming event.

[SINGLE RESPONSE]

1. Strongly agree
2. Somewhat agree
3. Neither agree nor disagree
4. Somewhat disagree
5. Strongly disagree

[ASK IF Q25 = 1]

Q30. Do you have any suggestions for how we can improve the timeliness of Demand Response Event **[INSERT HOVER TEXT]** notification(s)?

[SINGLE RESPONSE]

1. No
2. Yes, please specify: ________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK IF Q25 = 1]

Q31. During Demand Response Events **[INSERT HOVER TEXT]**, how would you rate your overall comfort level?

[SINGLE RESPONSE]

1. Very comfortable
2. Somewhat comfortable
3. Neither comfortable nor uncomfortable
4. Somewhat uncomfortable
5. Very uncomfortable
6. It depends on *what day* the Demand Response Events occurred
7. It depends on *what time of day* the Demand Response Events occurred

[ASK IF Q25 = 1]

Q32. Before today, were you aware that you can override the remote adjustment to your smart thermostat(s) during Demand Response Events **[INSERT HOVER TEXT]**?

[SINGLE RESPONSE]

1. Yes
2. No
98. Don't know

[ASK IF Q32 = 1]

Q33. How many times did you or someone in your household override the remote adjustment to your smart thermostat(s) during Demand Response Events **[INSERT HOVER TEXT]**?

[SINGLE RESPONSE]

1. 0
2. 1-3
3. 4-6
4. 7-9
5. 10-12
6. 13-15
7. More than 15
98. Don't know

[ASK IF Q33 = 2-7]

Q34. Why did you override the remote adjustment to your smart thermostat(s) during Demand Response Events **[INSERT HOVER TEXT]**? Please select all that apply.

[MULTIPLE RESPONSE]

1. I/we were too cold
2. Party/family event
3. Household members or visitors with cold sensitivity/health problems
4. Long period of cold weather
5. Heating system having problems
6. Peak Event was too long
97. Other, please specify: ____________ **[INSERT OPEN-ENDED RESPONSE]**
98. Don't know

Q35. What strategies, if any, did you undertake to stay comfortable at the time of Demand Response Events **[INSERT HOVER TEXT]**? Please select all that apply.

[MULTIPLE RESPONSE, RANDOMIZE 1-7

1. Wore more layers and/or blankets *during* Peak Event(s)
2. Ate and drank warm foods and drinks *during* Peak Event(s)
3. Pre-heat home by raising thermostat(s) *before* Peak Event(s)
4. Used major heat-generating appliances (such as the oven) *during* Peak Event(s)
5. Turned on a space heater to keep warm *during* Peak Event(s)
6. Left the house to go somewhere warmer (such as a coffee shop, workplace, etc.) *during* Peak Event(s)
7. I did not use any strategy to stay comfortable

97. Other, please specify: __________ **[INSERT OPEN-ENDED RESPONSE]**
98. Don't know

7.2.2.7 Household Characteristics

[DISPLAY FOR ALL RESPONSENTS]: Please answer the next set of questions so we can better understand who participates and make sure we have reached a variety of households. Your responses will remain anonymous.

[ASK ALL RESPONDENTS]

Q36. About when was the building you live in first built? Your best guess is fine.

[SINGLE RESPONSE]

1. 2020 or later
2. 2010 to 2019
3. 2000 to 2009
4. 1990 to 1999
5. 1980 to 1989
6. 1970 to 1979
7. 1960 to 1969
8. 1950 to 1959
9. 1940 to 1949
10. 1939 or earlier

98. Don't know
99. Prefer not to answer

[ASK ALL RESPONDENTS]

Q37. Which of the following best describes the type of home you live in?

[SINGLE RESPONSE]

1. Single family, detached (e.g., freestanding house)
2. Single family, attached (e.g., townhouse or row house)
3. Apartment in multi-unit structure of 2-4 units
4. Apartment in multi-unit structure of 5 or more units
5. Mobile home

98. Don't know
99. Prefer not to answer

[ASK ALL RESPONDENTS]

Q38. What is the total square footage of your home? Your best guess is fine.

[SINGLE RESPONSE]

1. Less than 500
2. 500 to 749
3. 750 to 999
4. 1,000 to 1,499
5. 1,500 to 1,999
6. 2,000 to 2,499
7. 2,500 to 2,999
8. 3,000 to 3,999
9. 4,000 or more

98. Don't know
99. Prefer not to answer

[ASK ALL RESPONDENTS]

Q39. How many smart thermostats do you have in your home?

[SINGLE RESPONSE]

1. 1
2. 2
3. 3
4. 4

5. 5 or more
98. Don't know
99. Prefer not to answer

Q40. How many people in your household are over 65 years of age?

[SINGLE RESPONSE]

1. 0
2. 1
3. 2
4. 3
5. 4
6. 5 or more
98. Don't know
99. Prefer not to answer

[ASK ALL RESPONDENTS]

Q41. How many people in your household are under 5 years of age?

[SINGLE RESPONSE]

1. 0
2. 1
3. 2
4. 3
5. 4
6. 5 or more
98. Don't know
99. Prefer not to answer

Q42. What is the highest degree or level of school you have completed? If you're currently enrolled in school, please indicate the highest degree you have received.

[SINGLE RESPONSE]

1. Less than a high school diploma
2. High school degree or equivalent
3. Vocational/trade school or associate degree

4. Bachelor’s degree (e.g., BA, BS)
5. Master’s degree (e.g., MA, MS, MEd)
6. Doctorate (e.g., PhD, MD, EdD)
97. Other (please specify)
99. Prefer not to answer

Q43. What is the primary household language?

[SINGLE RESPONSE]

1. English
2. Spanish
3. Chinese (including Mandarin and Cantonese)
4. Tagalog
5. Vietnamese
6. Korean
97. Other (please specify)
99. Prefer not to answer

[ASK ALL RESPONDENTS]

Q44. Please select all races/ethnicities below with which members of your household identify:

[MULTIPLE RESPONSE]

1. White or Caucasian
2. Black or African American
3. Hispanic, Latino/Latina, or Spanish
4. Asian (Chinese, Japanese, Korean, Filipino, Vietnamese, Indian)
5. American Indian or Alaska Native
6. Two or more races
97. Other, please specify: [OPEN-ENDED RESPONSE]
99. Prefer not to answer

[ASK ALL RESPONDENTS]

Q45. Please select the range that best describes your household’s annual 2022 income before taxes:

[SINGLE RESPONSE]

1. $Less than $10,000
2. $10,000 to $14,999
3. $15,000 to $24,999
4. $25,000 to $34,999
5. $35,000 to $49,999
6. $50,000 to $74,999
7. $75,000 to $99,999
8. $100,000 to $149,999
9. $150,000 to $199,999
10. $200,000 or more
98. Don't know
99. Prefer not to answer

7.2.2.8 Closing

[ASK ALL RESPONDENTS]

Q46. Is there anything else you want to tell us about your experience with PSE's Demand Response program or this survey?

[SINGLE REPONSE]

1. No
2. Yes. Please share your comments: _________________ **[INSERT OPEN-ENDED RESPONSE]**

[ASK ALL RESPONDENTS]

Q47. As a thank you for your participation in this research, your response will be entered into a drawing for an Amazon e-gift card of **up to** $200. If selected for the e-gift card, you will be notified by email (please check your spam filter). Would you like to be included in the incentive drawing? **[SINGLE REPONSE]**

1. Yes
2. No

[ASK IF Q47 = 1]

Q48. Please provide your contact information for the drawing:

[SINGLE REPONSE]

1. First name: _________________ **[INSERT OPEN-ENDED RESPONSE]**
2. Last name: _________________ **[INSERT OPEN-ENDED RESPONSE]**
3. Email address: _________________ **[INSERT OPEN-ENDED RESPONSE]**

[DISPLAY FOR ALL RESPONDENTS]: You have completed the survey and your responses have been submitted. Your contribution to this survey helps Puget Sound Energy to evaluate and improve its program offerings. Thank you for your participation and time.

7.3 Appendix C: Comparison of DNV and Autogrid impact estimates

To give a sense of the effect of these discrepancies in actual energy use data, we included tables comparing both DNV baselines versus Autogrid baselines for all customers (Table 7-1) and DNV baselines versus Autogrid baselines for *only* customers where DNV and Autogrid had the same actual event-day energy usage data (Table 7-2).

In Table 7-1, we see that both Autogrid's baseline and its actual event-day usage are contributing to its underestimated impact estimates. Actual event-day usage is overestimated, which causes impacts to be underestimated regardless of baseline. Additionally, as we have discussed, Autogrid baselines are lower than DNV baselines, which also leads to smaller impact estimates.

In Table 7-2, we reduce the set of customers to those with the same energy usage data in the DNV and Autogrid data. However, the Autogrid baselines in this case actually get lower, causing more baseline-based bias, probably because they do not have the artificially inflated energy usage data in the pre-period that is used to create the baseline. We see that Autogrid's baseline methodology substantially underestimates event impacts even if customers with data issues are removed.

In Table 7-3, we reduce the set of customers to those with different energy usage data in the DNV and Autogrid data. In this case, Autogrid actuals are consistently higher, Autogrid baselines vary by fuel type, and impacts also vary accordingly. We see that our impact estimates for these customers are consistent with the program overall impacts, while Autogrid's estimates fluctuate depending on the event.

Table 7-1. Comparison of DNV and Autogrid baselines, all customers

			Avg. Baseline			Avg. Actual				Avg. Impacts	
Fuel Type	Event Date	# Sites	DNV	Autogrid	% Diff.	# Actual Diff.	DNV	Autogrid	% Diff.	DNV	Autogrid
Gas	12/20/22	39	0.257	0.253	-1%	0	0.106	0.106	0%	0.151	0.147
	12/22/22	41	0.354	0.246	-30%	1	0.222	0.223	0%	0.131	0.024
	1/30/23	64	0.285	0.191	-33%	2	0.189	0.193	2%	0.096	-0.003
	1/31/23	65	0.244	0.188	-23%	2	0.146	0.147	1%	0.098	0.041
	2/23/23	66	0.268	0.195	-27%	1	0.165	0.165	0%	0.103	0.030
Electric	12/5/22	34	5.18	5.13	-1%	5	2.49	3.05	22%	2.69	2.08
	12/20/22	48	5.26	5.30	1%	8	2.98	3.84	29%	2.28	1.45
	12/22/22	47	6.73	5.49	-19%	8	4.67	5.92	27%	2.06	-0.43
	1/30/23	57	6.31	4.39	-30%	8	4.88	5.71	17%	1.43	-1.31
	1/31/23	57	5.26	4.39	-16%	8	4.01	4.67	17%	1.25	-0.28
	2/23/23	62	5.47	5.00	-9%	8	4.03	5.03	25%	1.44	-0.04

Table 7-2. Comparison of DNV and Autogrid baselines, customers with same actuals

			Avg. Baseline			Avg. Actual		Avg. Impacts	
Fuel Type	Event Date	# Sites	DNV	Autogrid	% Diff.	DNV	Autogrid	DNV	Autogrid
Gas	12/20/22	39	0.257	0.253	-1%	0.106	0.106	0.151	0.147
	12/22/22	40	0.357	0.248	-31%	0.228	0.228	0.130	0.020
	1/30/23	62	0.292	0.194	-33%	0.190	0.190	0.101	0.004
	1/31/23	63	0.250	0.191	-23%	0.149	0.149	0.101	0.042
	2/23/23	65	0.268	0.196	-27%	0.167	0.167	0.101	0.029
Electric	12/5/22	29	5.28	4.56	-14%	2.55	2.55	2.73	2.01
	12/20/22	40	5.53	4.52	-18%	3.04	3.04	2.50	1.49
	12/22/22	39	7.08	4.67	-34%	4.91	4.91	2.16	-0.25
	1/30/23	49	6.67	3.89	-42%	5.11	5.11	1.56	-1.22
	1/31/23	49	5.51	3.89	-30%	4.28	4.28	1.23	-0.39
	2/23/23	54	5.60	4.22	-25%	4.13	4.13	1.47	0.09

Table 7-3. Comparison of DNV and Autogrid baselines, customers with different actuals

			Avg. Baseline			Avg. Actual				Avg. Impacts	
Fuel Type	Event Date	# Sites	DNV	Autogrid	% Diff.	# Actual Diff.	DNV	Autogrid	% Diff.	DNV	Autogrid
Gas	12/22/22	1	0.142	0.182	28%	1	0.010	0.020	100%	0.132	0.162
	1/30/23	2	0.232	0.081	-65%	2	0.140	0.280	100%	0.092	-0.199
	1/31/23	2	0.062	0.081	30%	2	0.030	0.060	100%	0.032	0.021
	2/23/23	1	0.119	0.117	-2%	1	0.020	0.040	100%	0.099	0.077
Electric	12/5/22	5	4.63	8.46	82%	5	2.16	5.98	176%	2.47	2.48
	12/20/22	8	3.89	9.17	135%	8	2.70	7.89	193%	1.20	1.28
	12/22/22	8	5.06	9.50	88%	8	3.50	10.83	209%	1.55	-1.33
	1/30/23	8	4.15	7.51	81%	8	3.50	9.38	168%	0.65	-1.87
	1/31/23	8	3.70	7.51	103%	8	2.32	7.06	205%	1.38	0.44
	2/23/23	8	4.60	10.25	123%	8	3.35	11.12	232%	1.24	-0.87

7.4 Appendix D: Cumulative curtailment by event

The following plots show event specific cumulative hourly impacts for both electric and gas participants. The x-axis is hour of the day from 0 (12:00 am) to 23 (11:00 pm). The y-axis is the estimated cumulative total curtailment based on average hourly curtailment and the number of participants enrolled in the event. See Section 4.1 for more information.

7.4.1 Electric impacts

Figure 7-2. Total load reduction, pre-heating, and snapback by event hour for the 12/5/2022 event

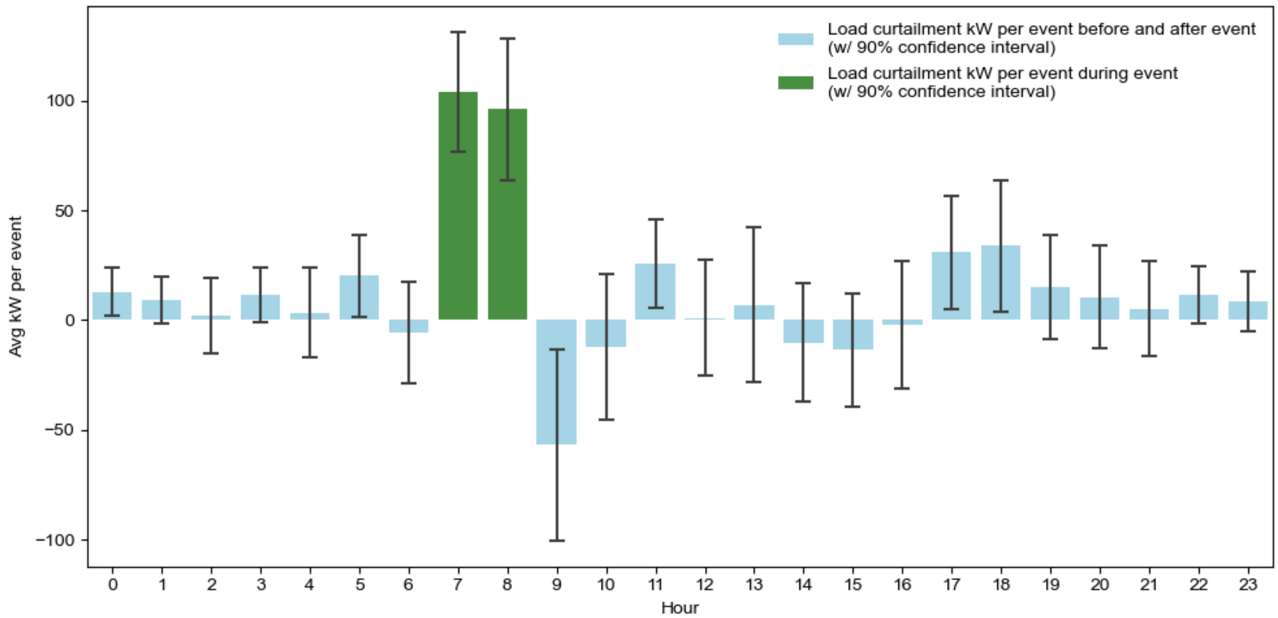


Figure 7-3. Total load reduction, pre-heating, and snapback by event hour for the 12/20/2022 event

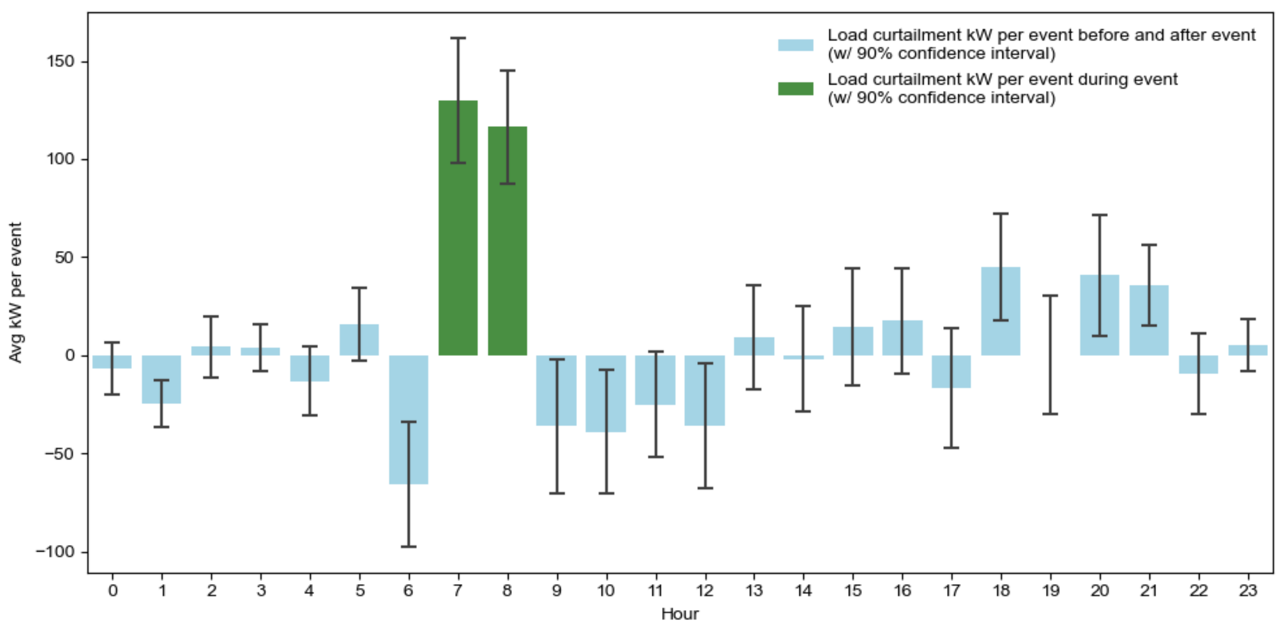


Figure 7-4. Total load reduction, pre-heating, and snapback by event hour for the 12/22/2022 event

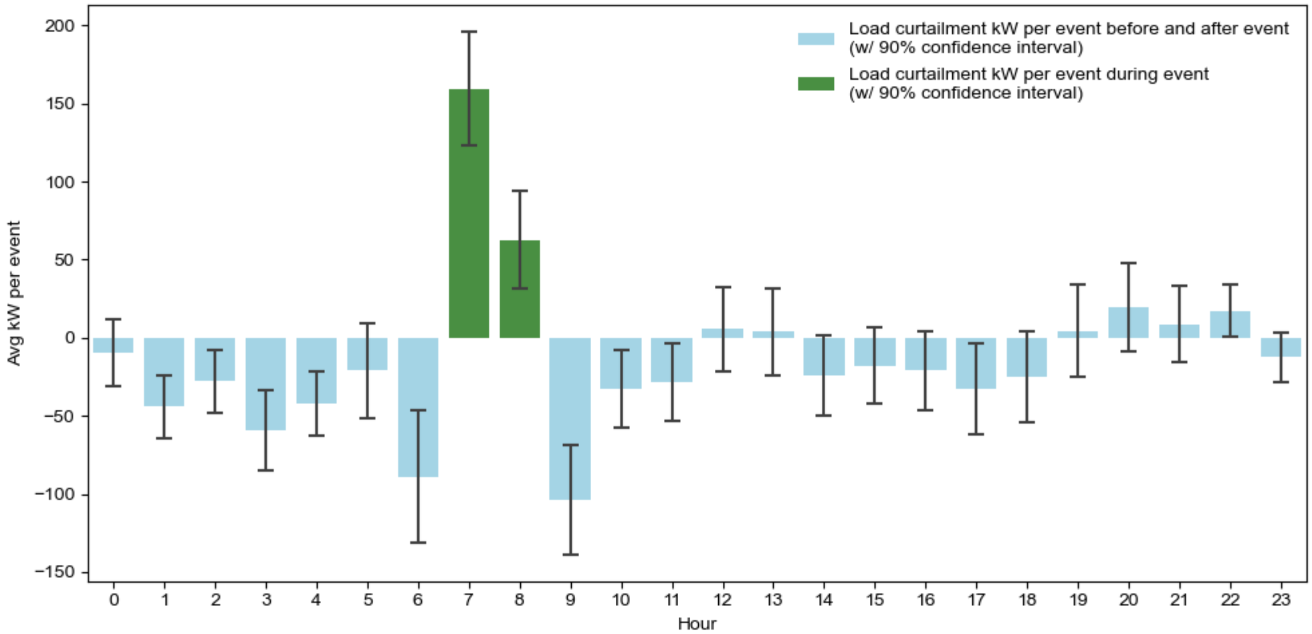


Figure 7-5. Total load reduction, pre-heating, and snapback by event hour for the 1/30/2023 event

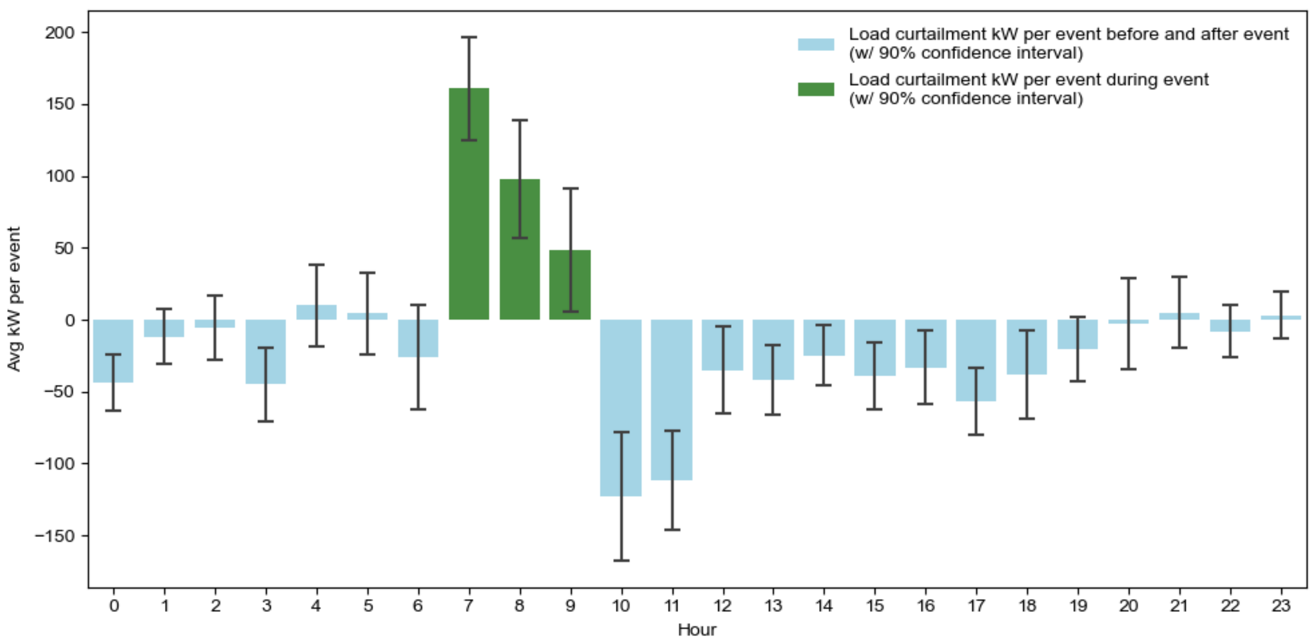


Figure 7-6. Total load reduction, pre-heating, and snapback by event hour for the 1/31/2023 event

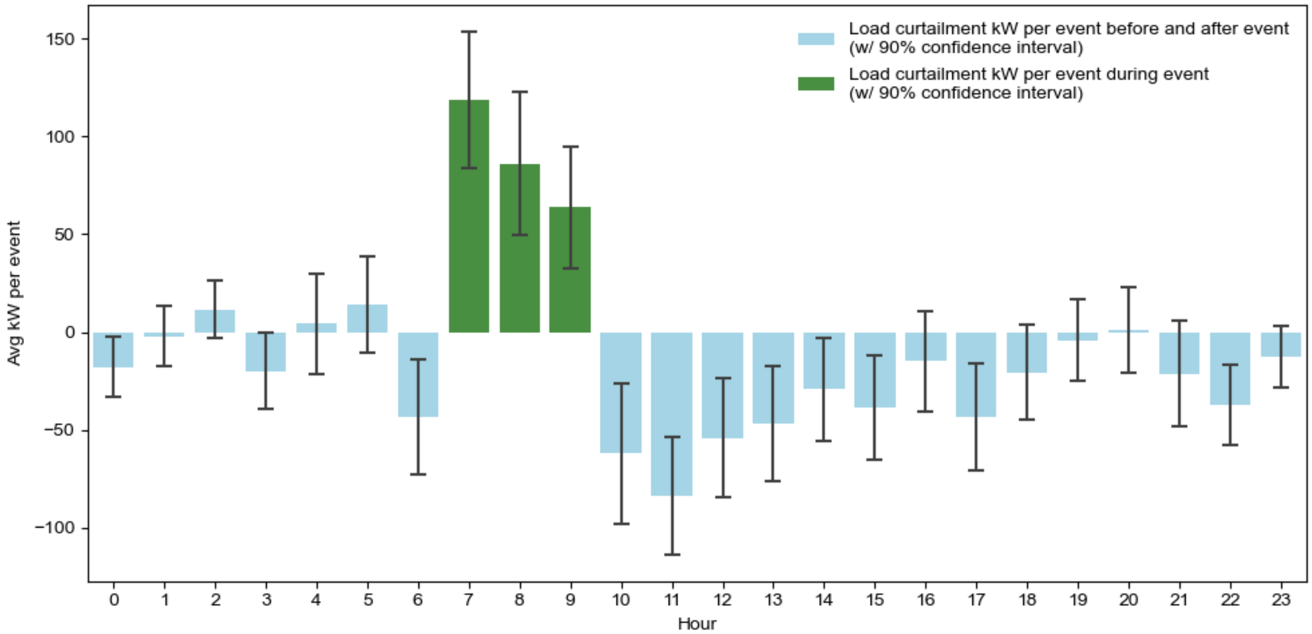


Figure 7-7. Total load reduction, pre-heating, and snapback by event hour for the 2/23/2023 event

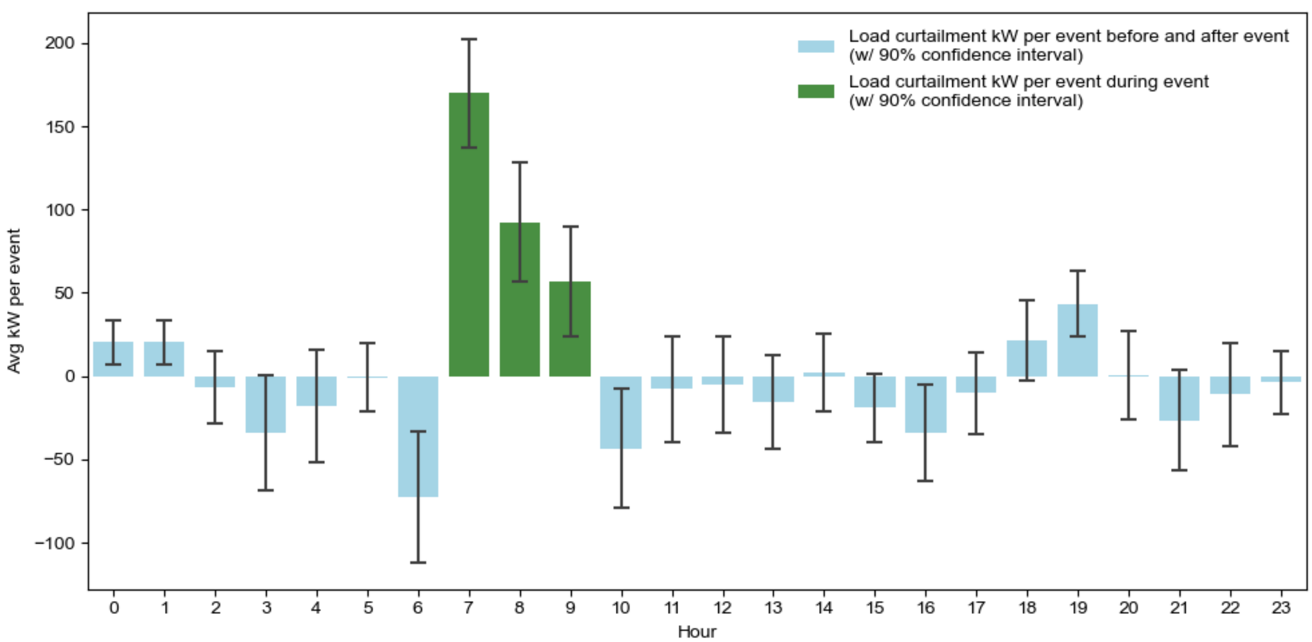
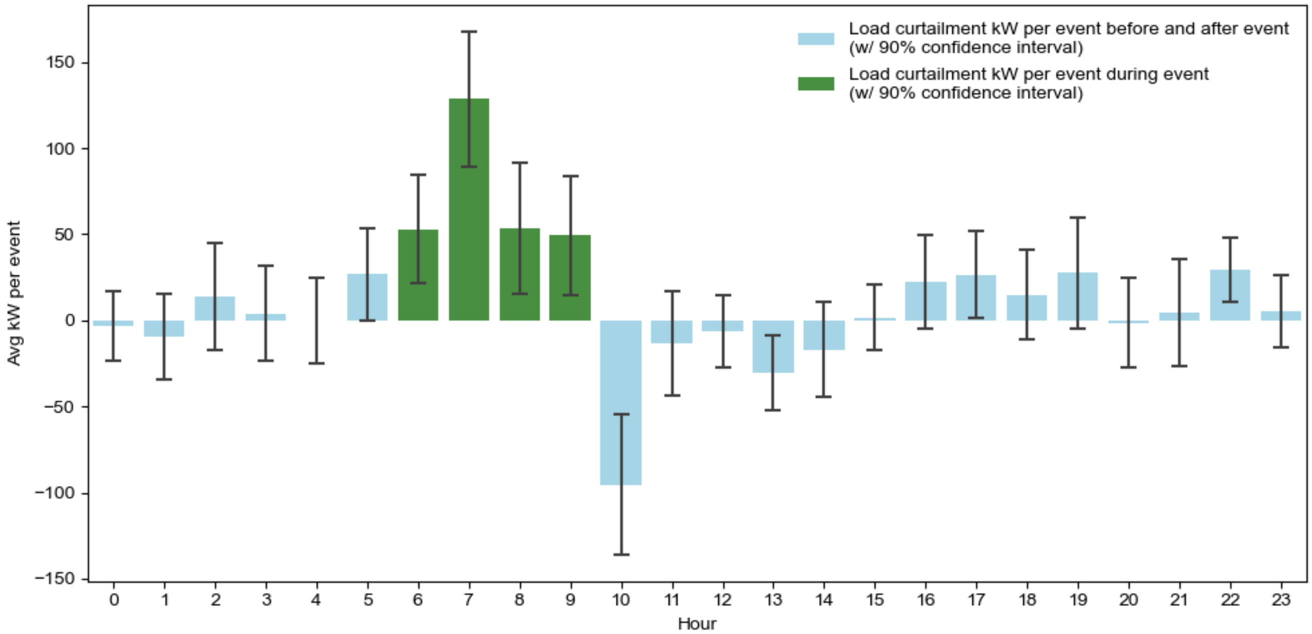


Figure 7-8. Total load reduction, pre-heating, and snapback by event hour for the 2/24/2023 event



7.4.2 Gas impacts

Figure 7-9. Total load reduction, pre-heating, and snapback by event hour for the 12/20/2022 event

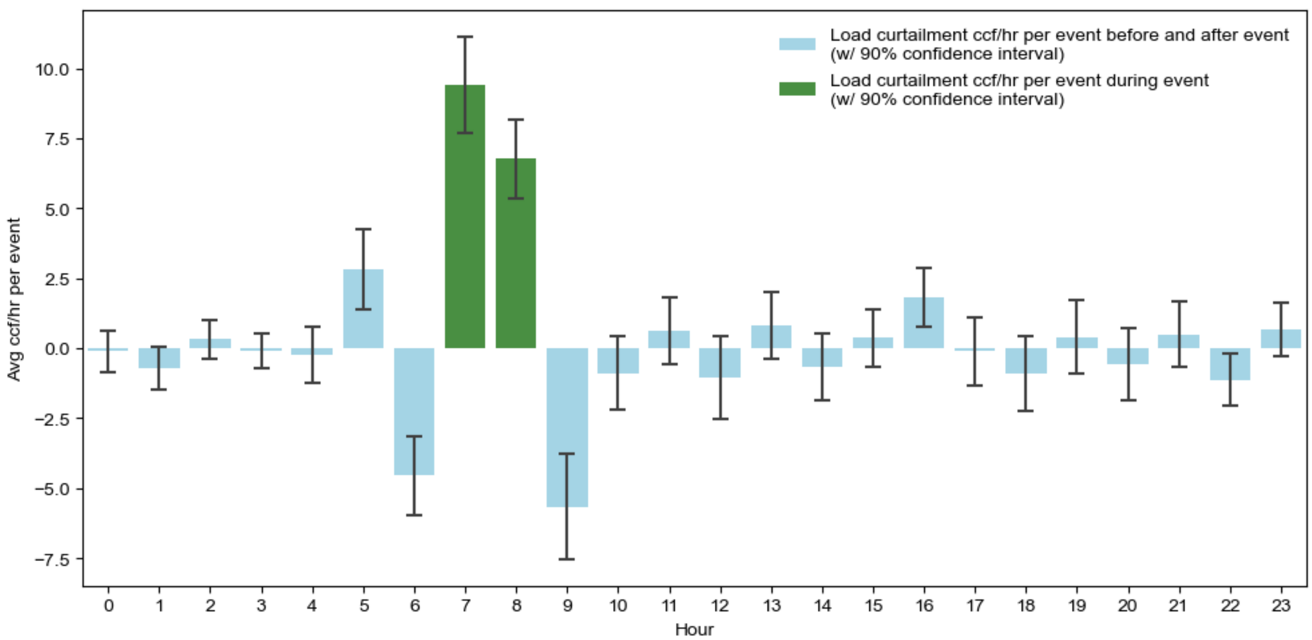


Figure 7-10. Total load reduction, pre-heating, and snapback by event hour for the 12/22/2022 event

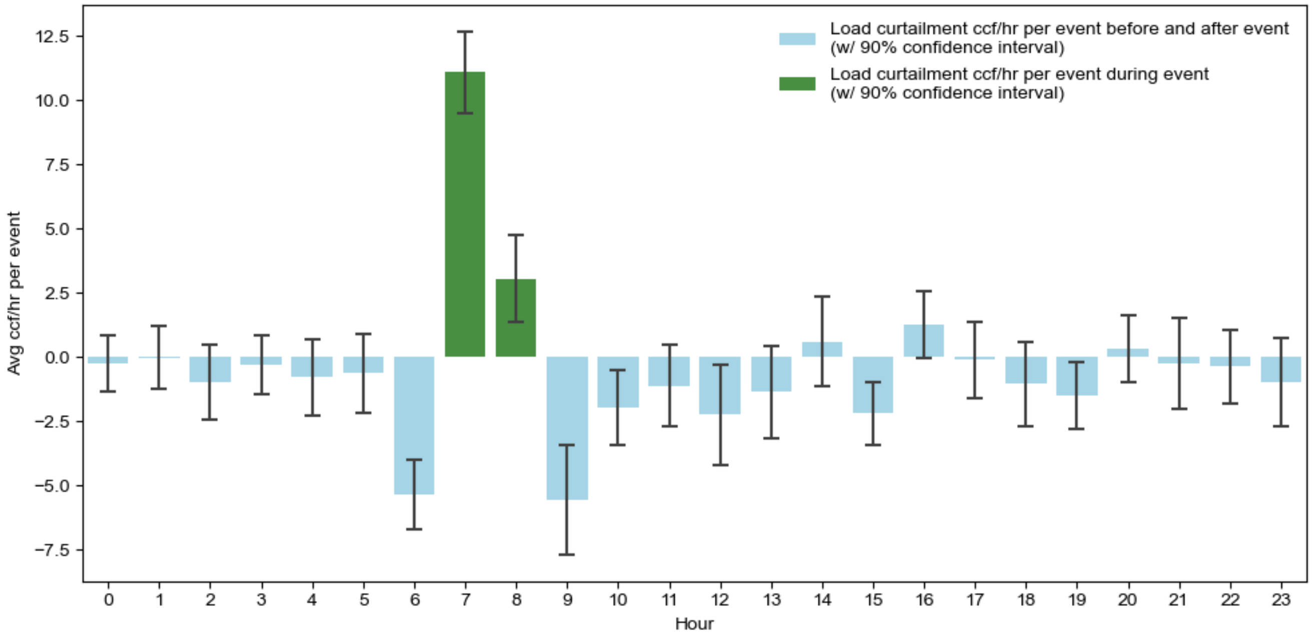


Figure 7-11. Total load reduction, pre-heating, and snapback by event hour for the 1/30/2023 event

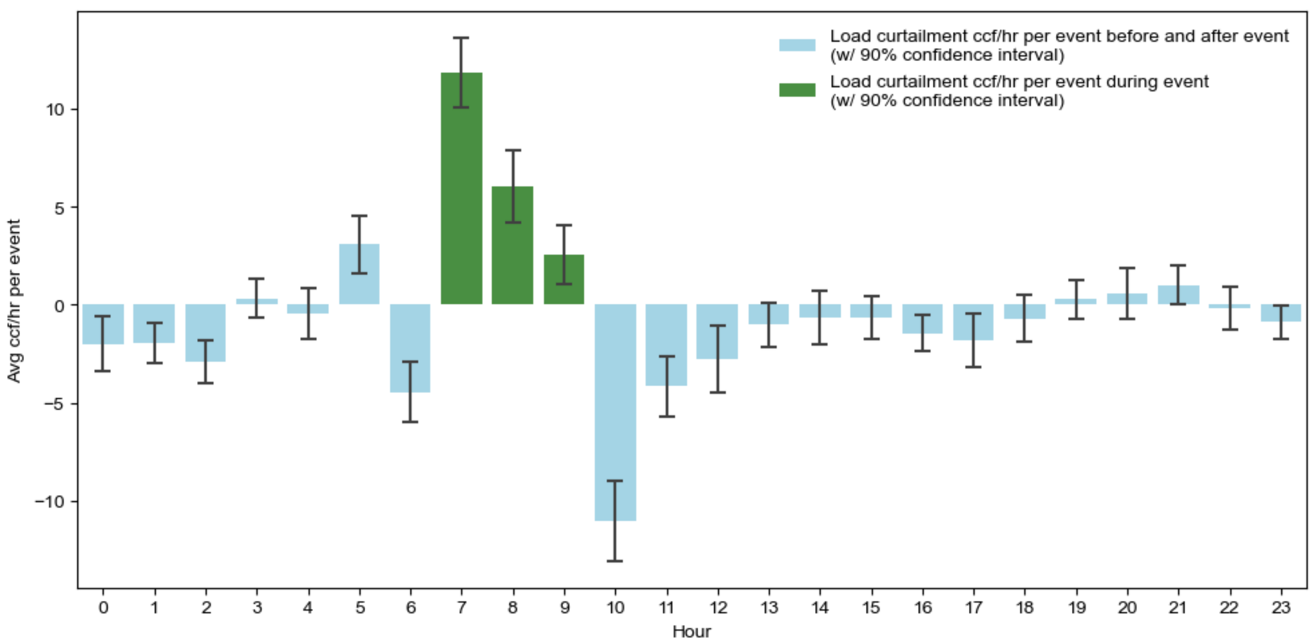


Figure 7-12. Total load reduction, pre-heating, and snapback by event hour for the 1/31/2023 event

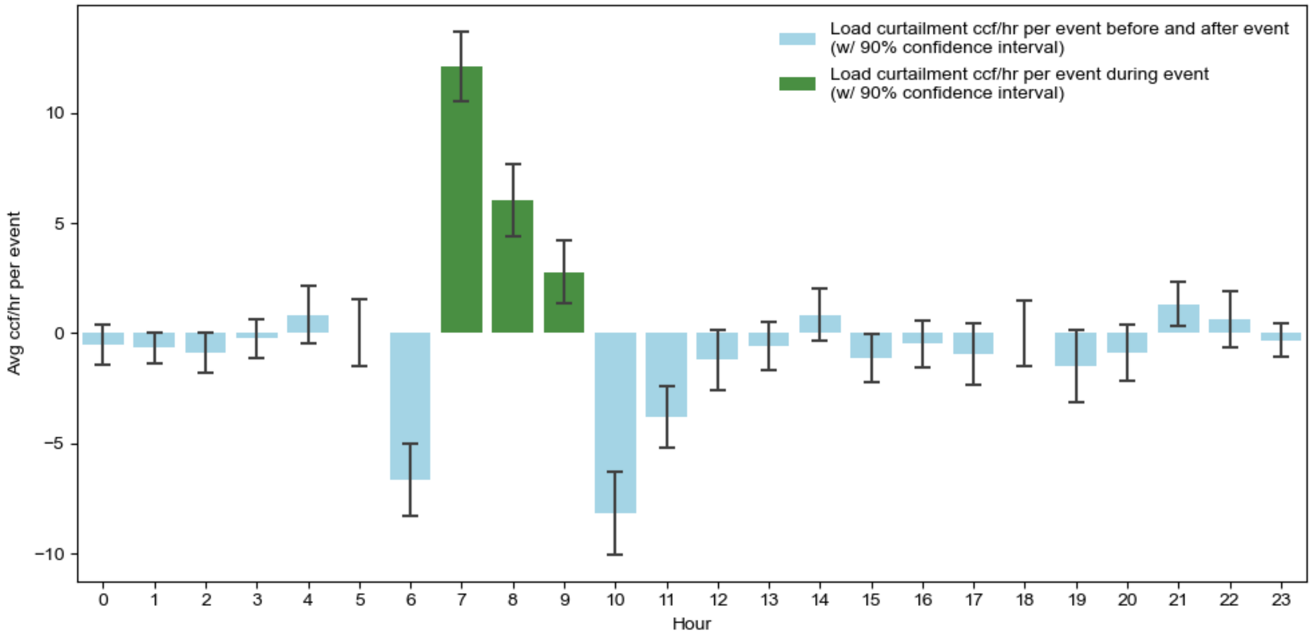


Figure 7-13. Total load reduction, pre-heating, and snapback by event hour for the 2/23/2023 event

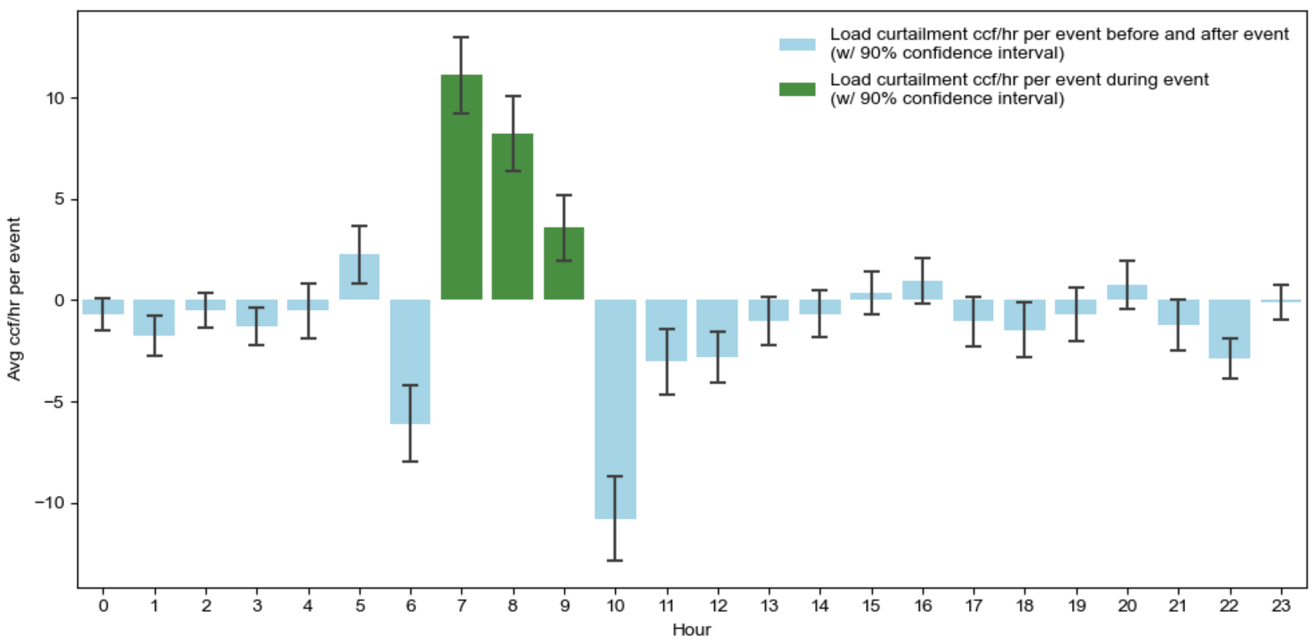
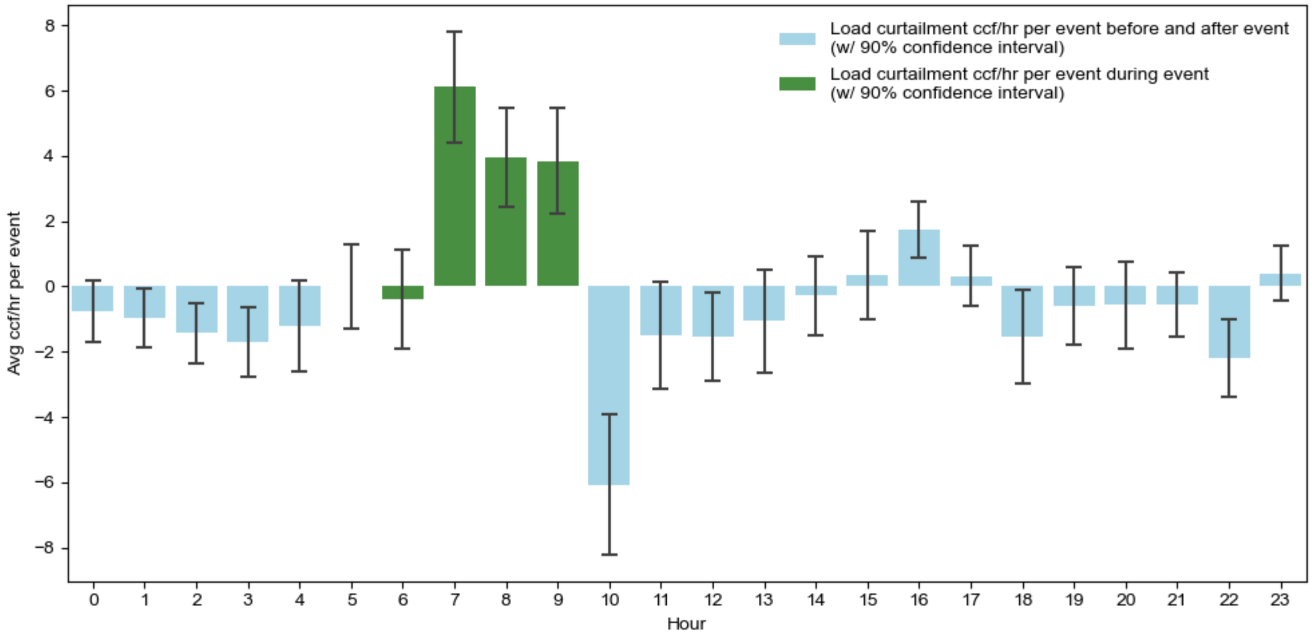


Figure 7-14. Total load reduction, pre-heating, and snapback by event hour for the 2/24/2023 event



7.5 Appendix E: Event-specific impact plots

The following plots show event specific average hourly impacts for both electric and gas participants. The x-axis is hour of the day from 0 (12:00 am) to 23 (11:00 pm). The y-axis is the average difference between participant hourly event load and the calculated difference-in-difference adjusted baseline load discussed in Section 3.1.3.

7.5.1 Electric impacts

Figure 7-15. Average hourly impacts by premise for the 12/5/2022 event

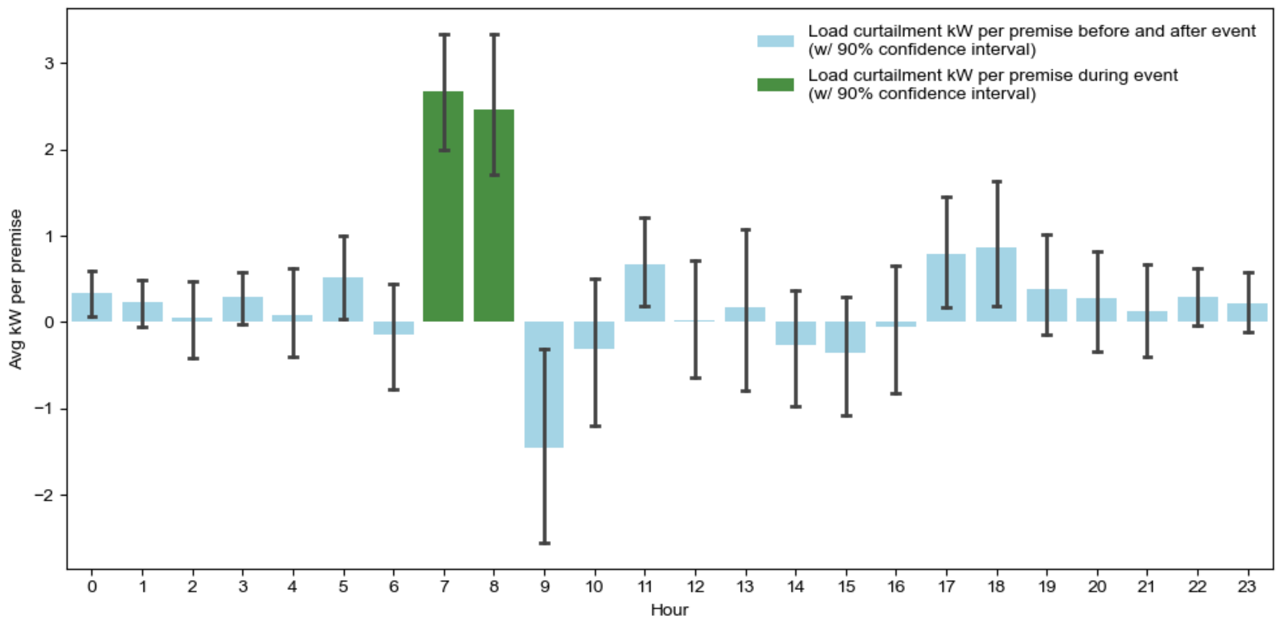


Figure 7-16. Average hourly impacts by premise for the 12/20/2022 event

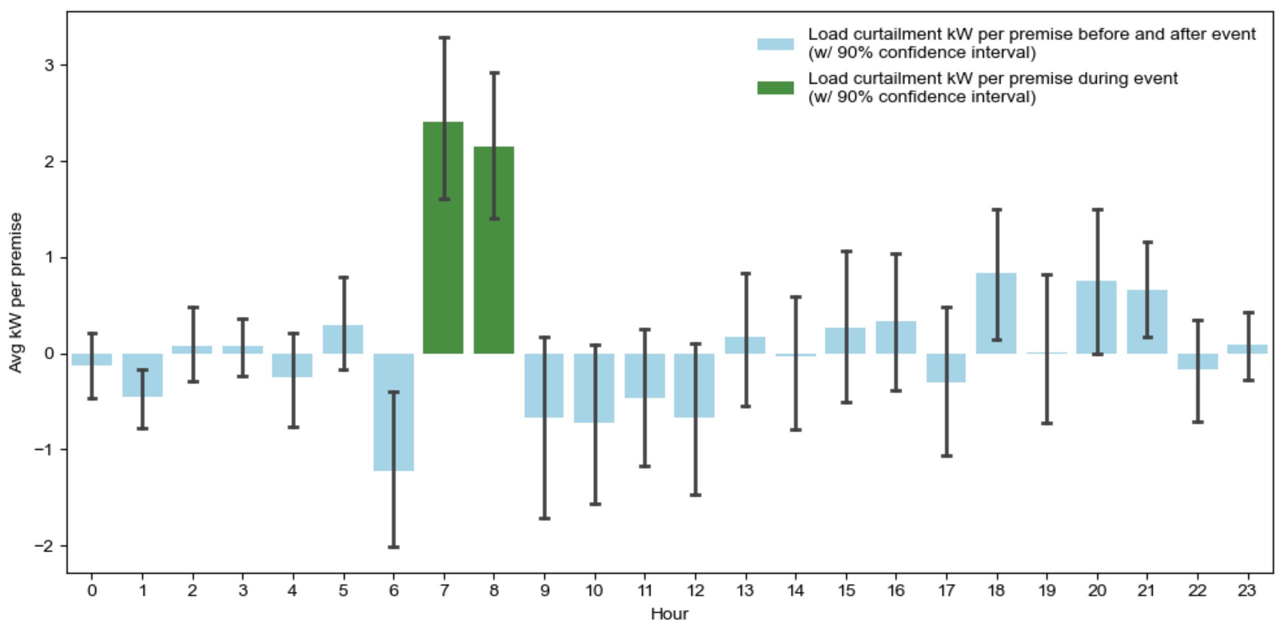


Figure 7-17. Average hourly impacts by premise for the 12/22/2022 event

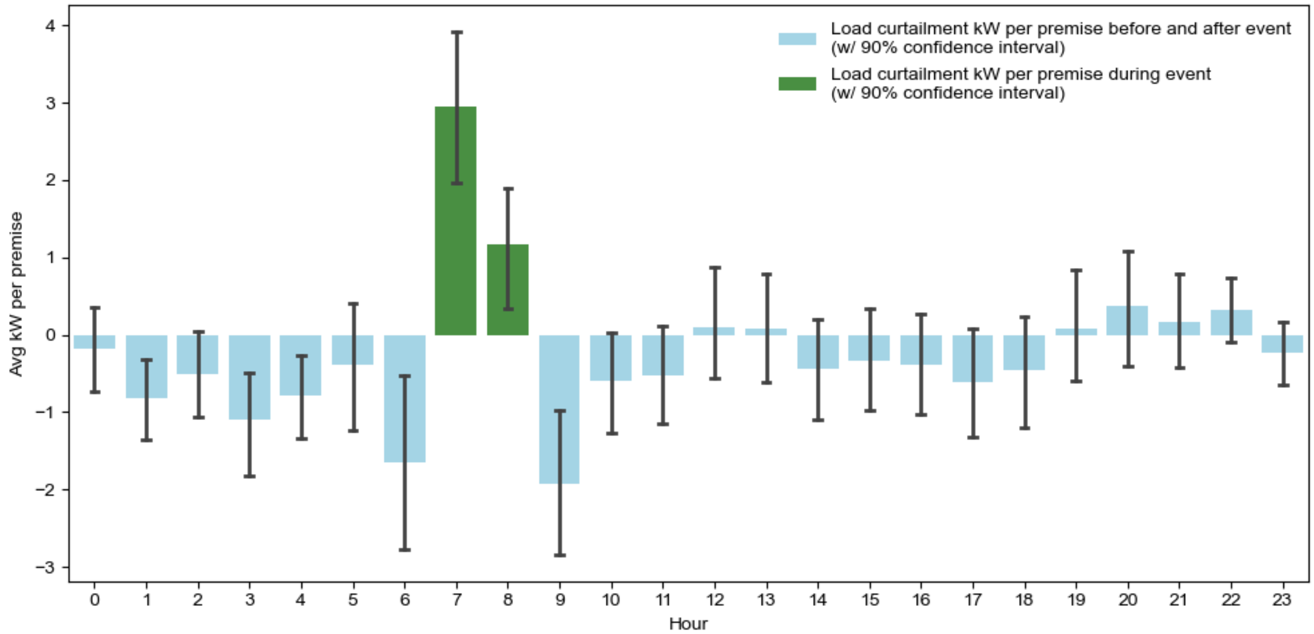


Figure 7-18. Average hourly impacts by premise for the 1/30/2023 event

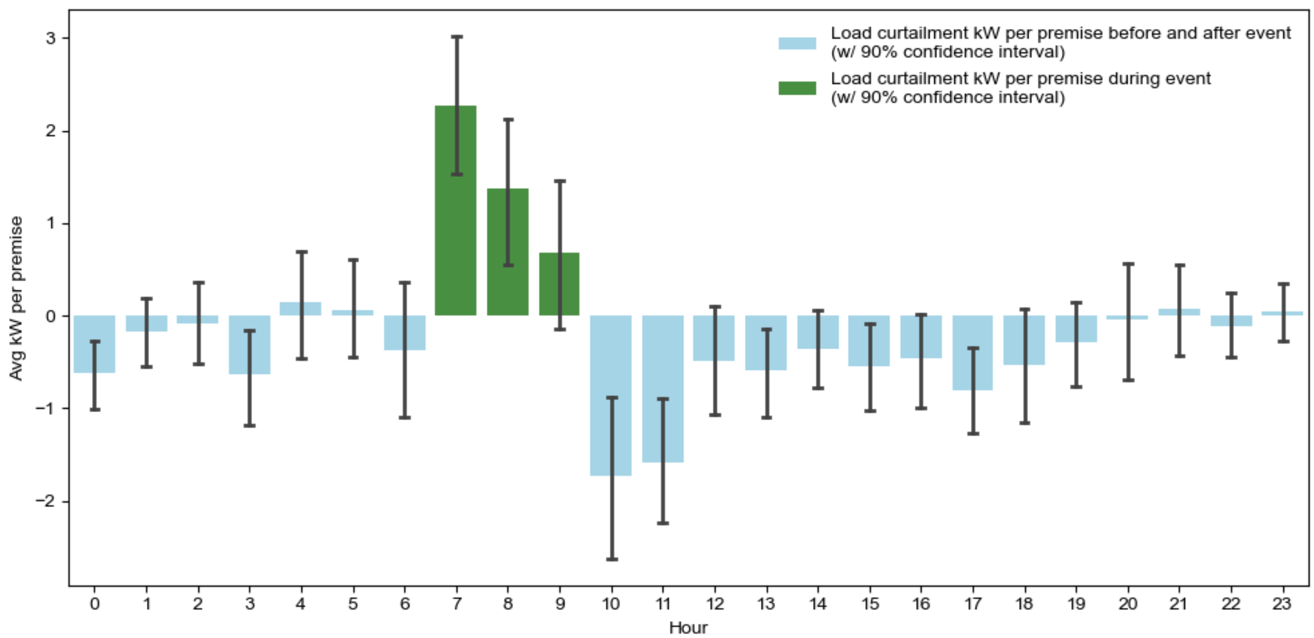


Figure 7-19. Average hourly impacts by premise for the 1/31/2023 event

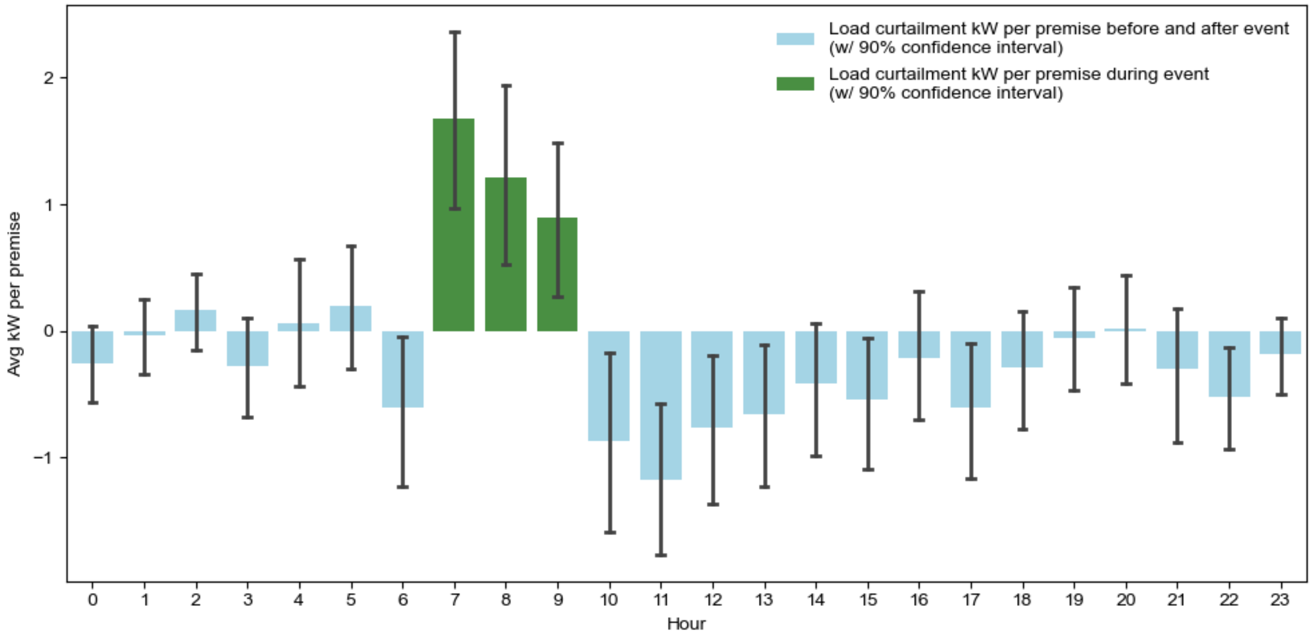


Figure 7-20. Average hourly impacts by premise for the 2/23/2023 event

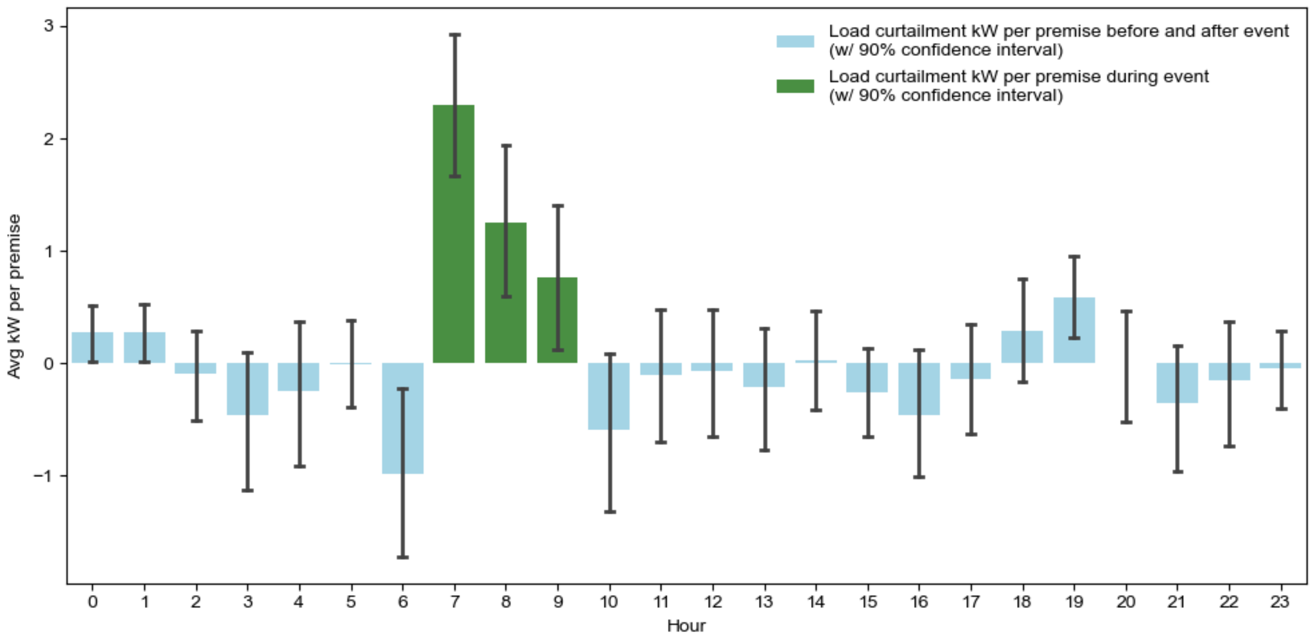
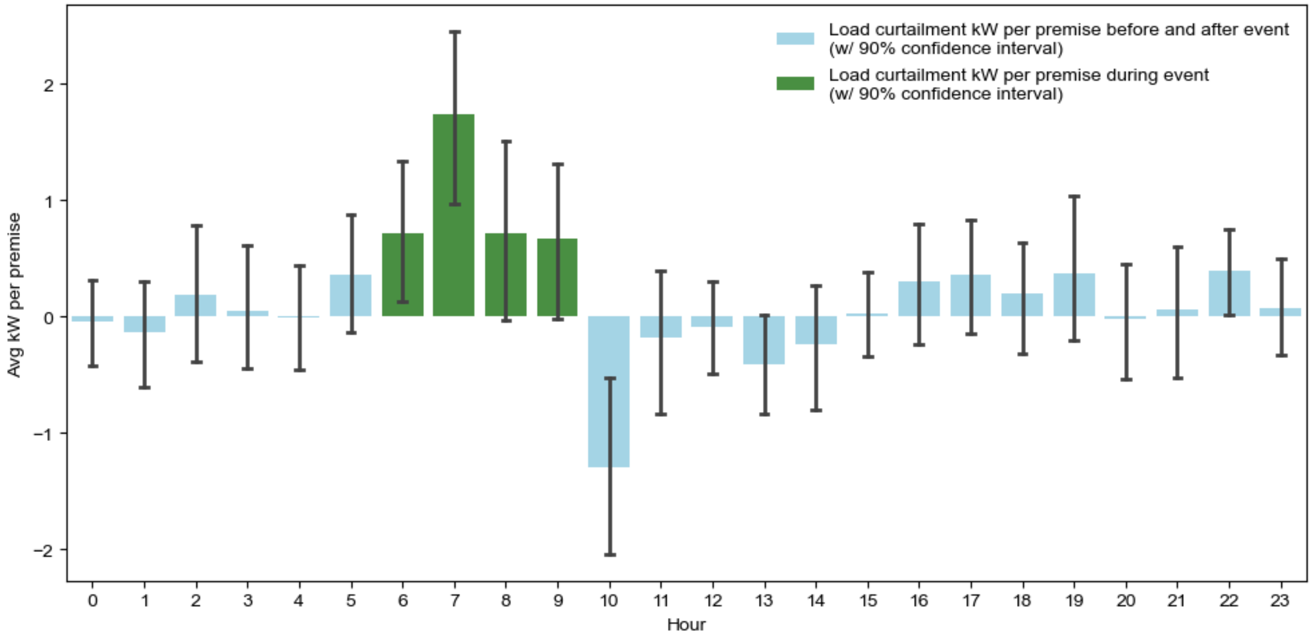


Figure 7-21. Average hourly impacts by premise for the 2/24/2023 event



7.5.2 Gas impacts

Figure 7-22. Average hourly impacts by premise for the 12/20/2022 event

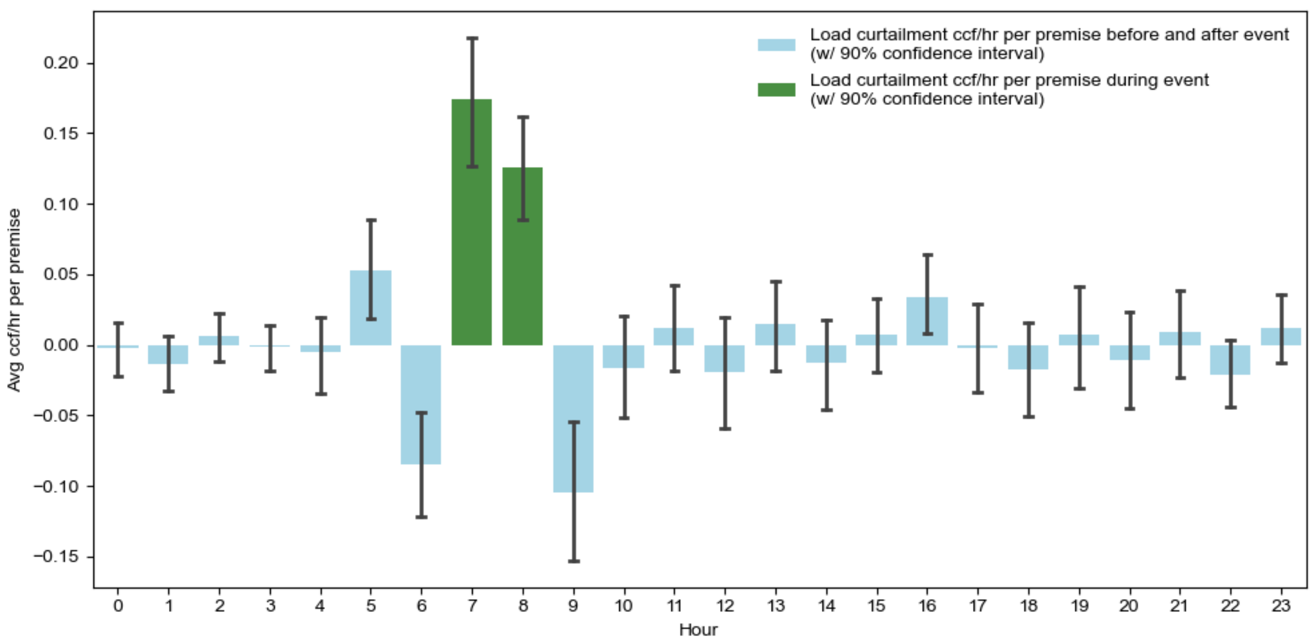


Figure 7-23. Average hourly impacts by premise for the 12/22/2022 event

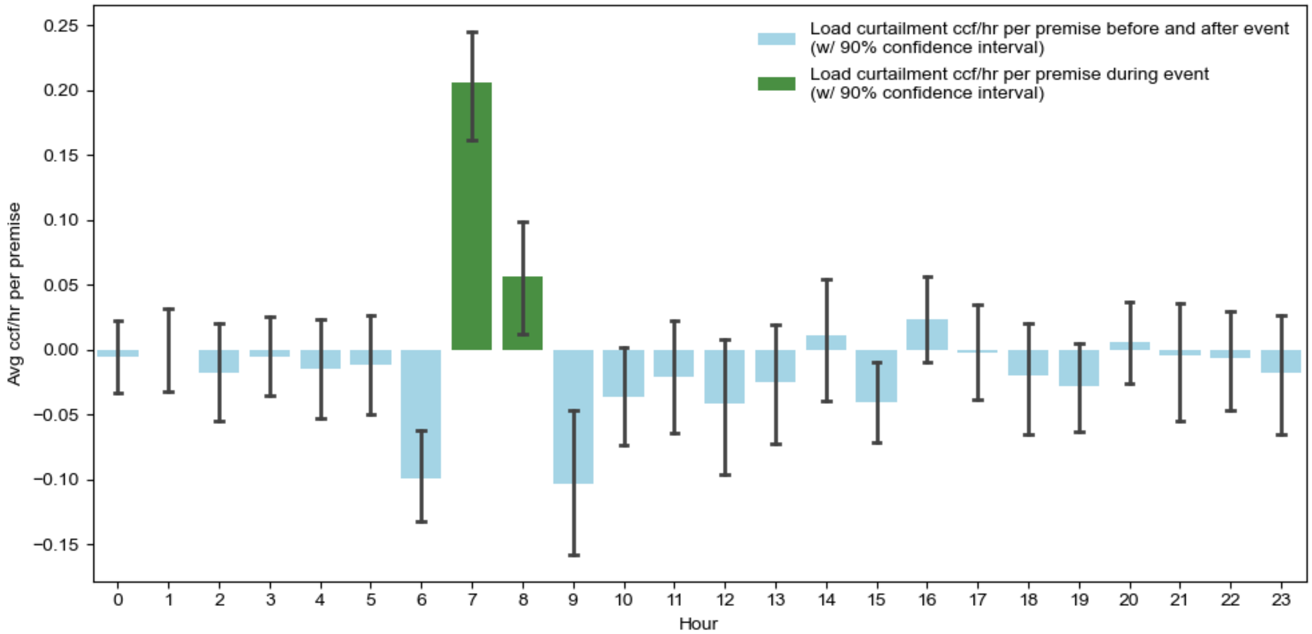


Figure 7-24. Average hourly impacts by premise for the 1/30/2023 event

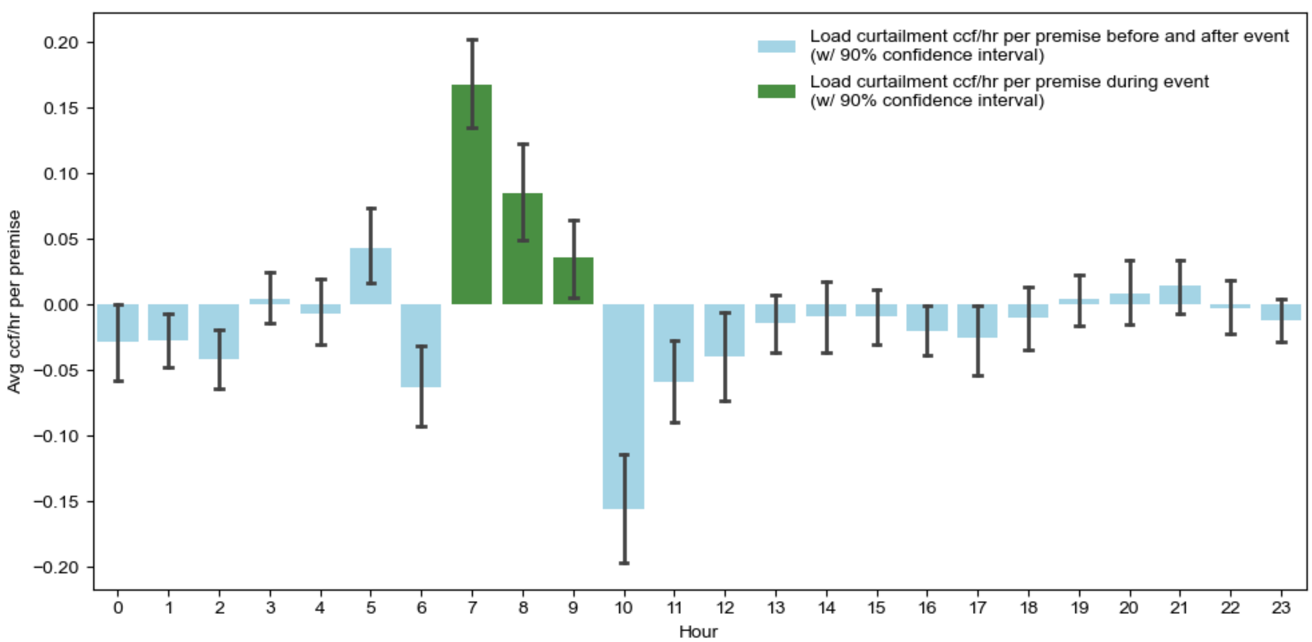


Figure 7-25. Average hourly impacts by premise for the 1/31/2023 event

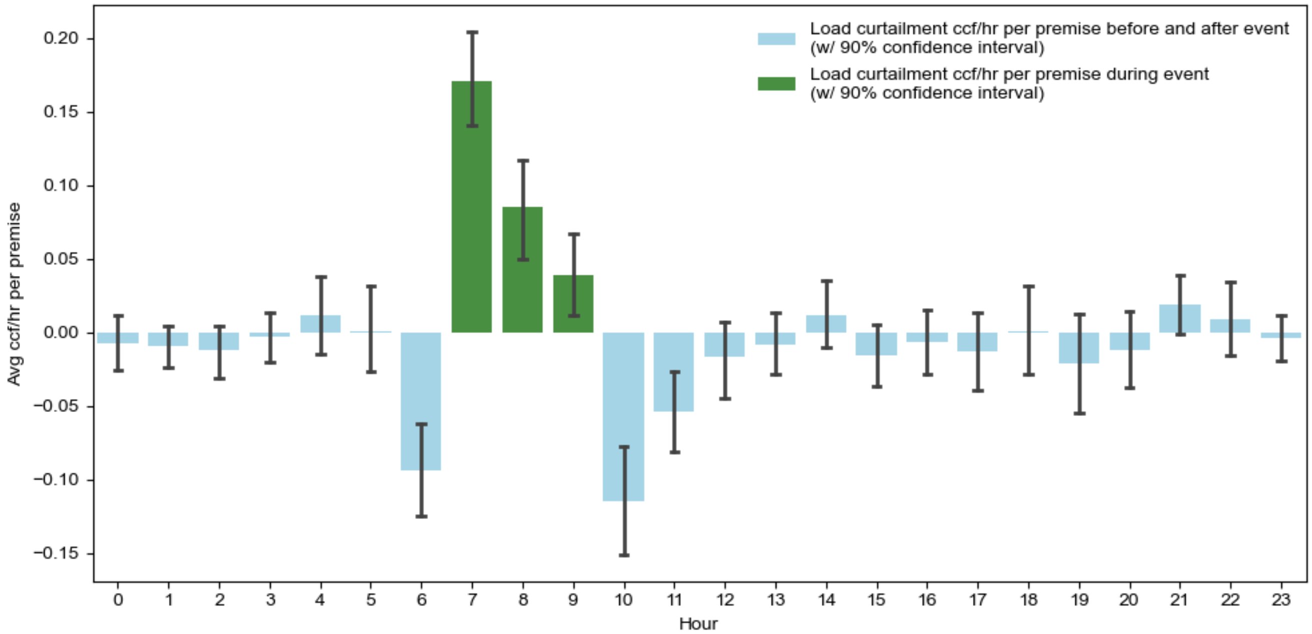


Figure 7-26. Average hourly impacts by premise for the 2/23/2023 event

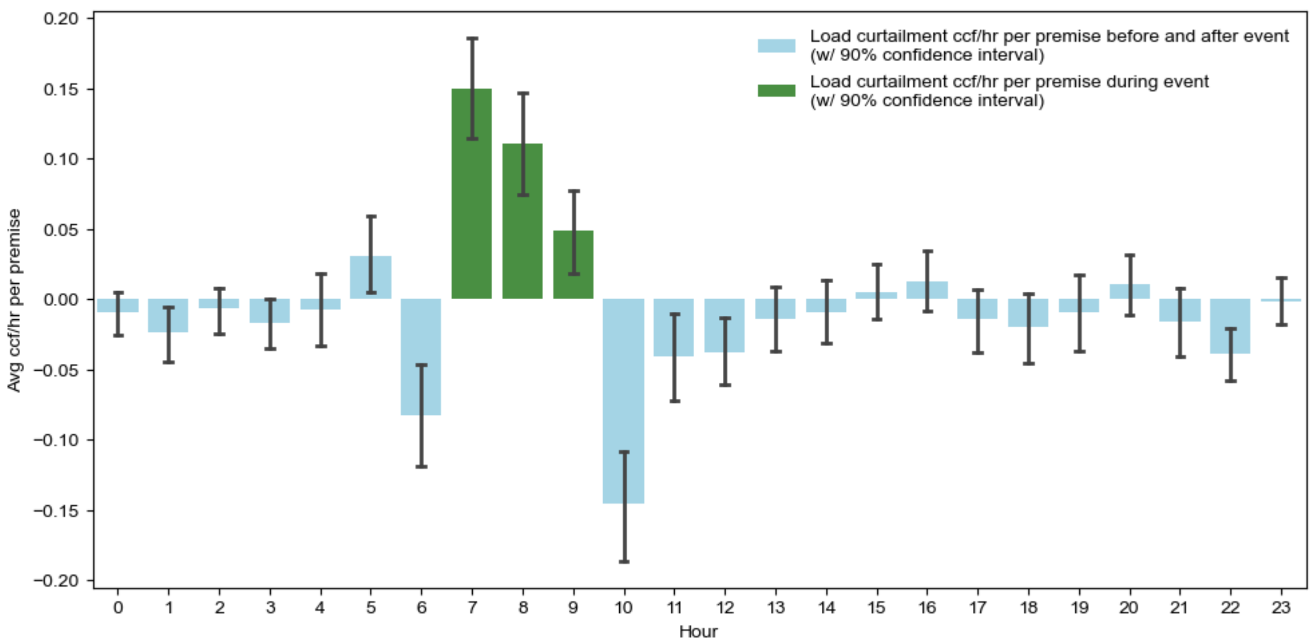
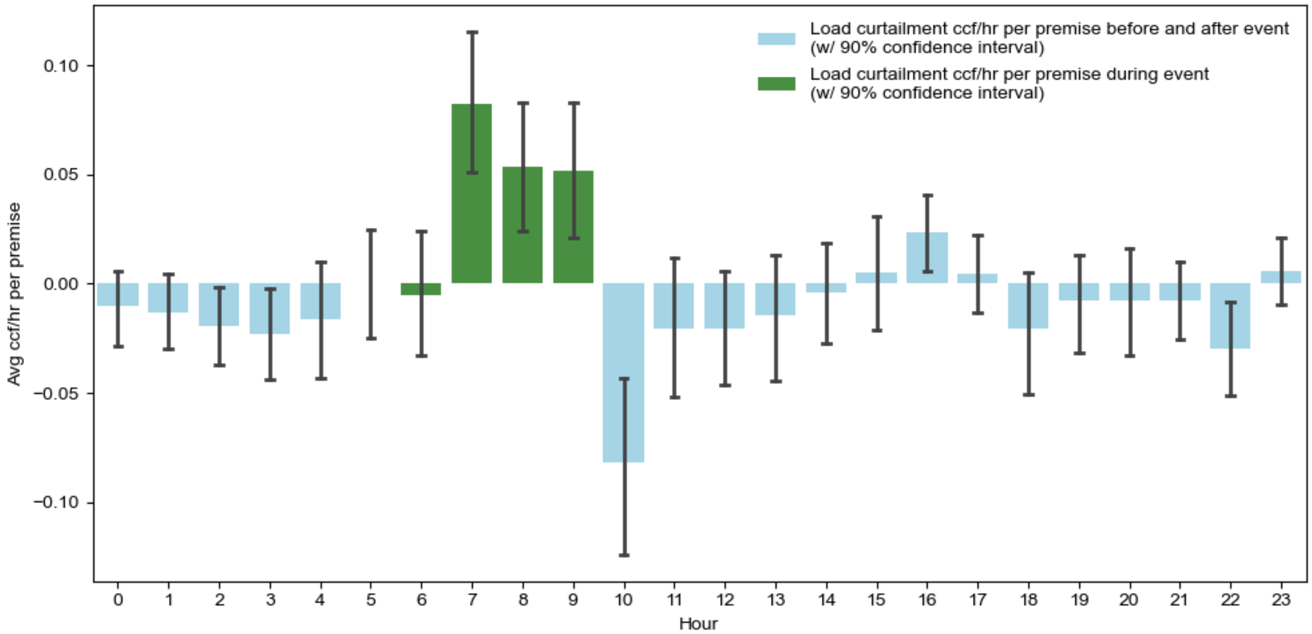


Figure 7-27. Average hourly impacts by premise for the 2/24/2023 event



7.6 Appendix F: Event-specific matching load shapes

The following plots show event specific matching load shapes for both electric and gas participants. The x-axis is hour of the day from 0 (12:00 am) to 23 (11:00 pm). The y-axis is the average hourly load for either the treatment or match group. More detail can be found in Section 3.1.2.

7.6.1 Electric load shapes

Figure 7-28. Average pre-event matching day load for the 12/5/2022 event

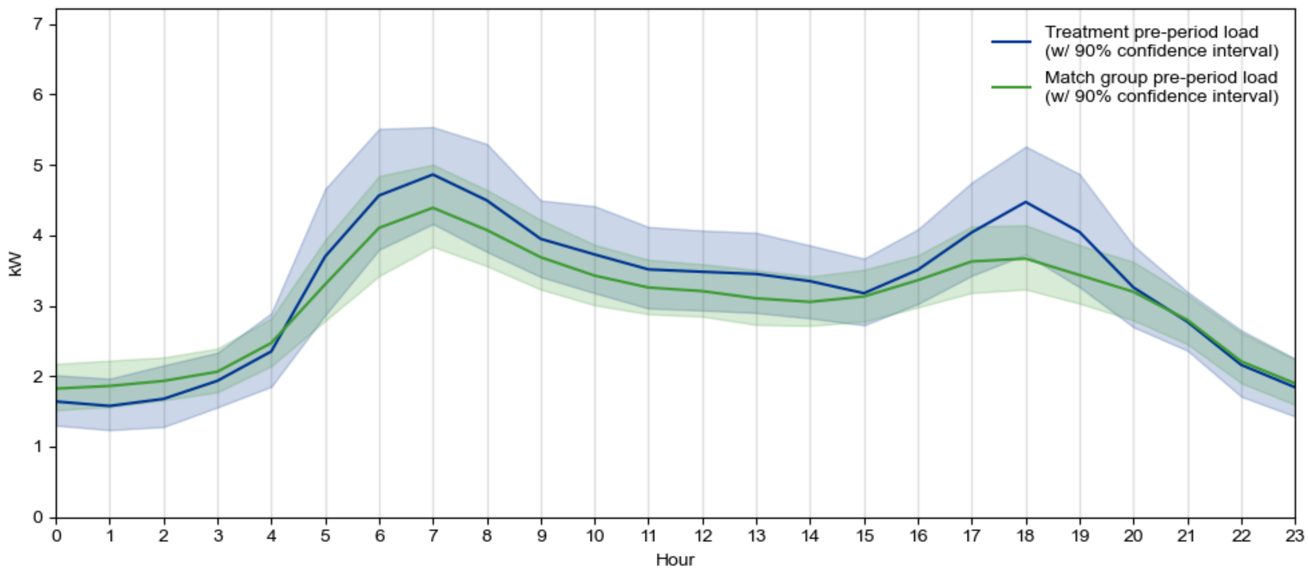


Figure 7-29. Average pre-event matching day load for the 12/20/2022 event

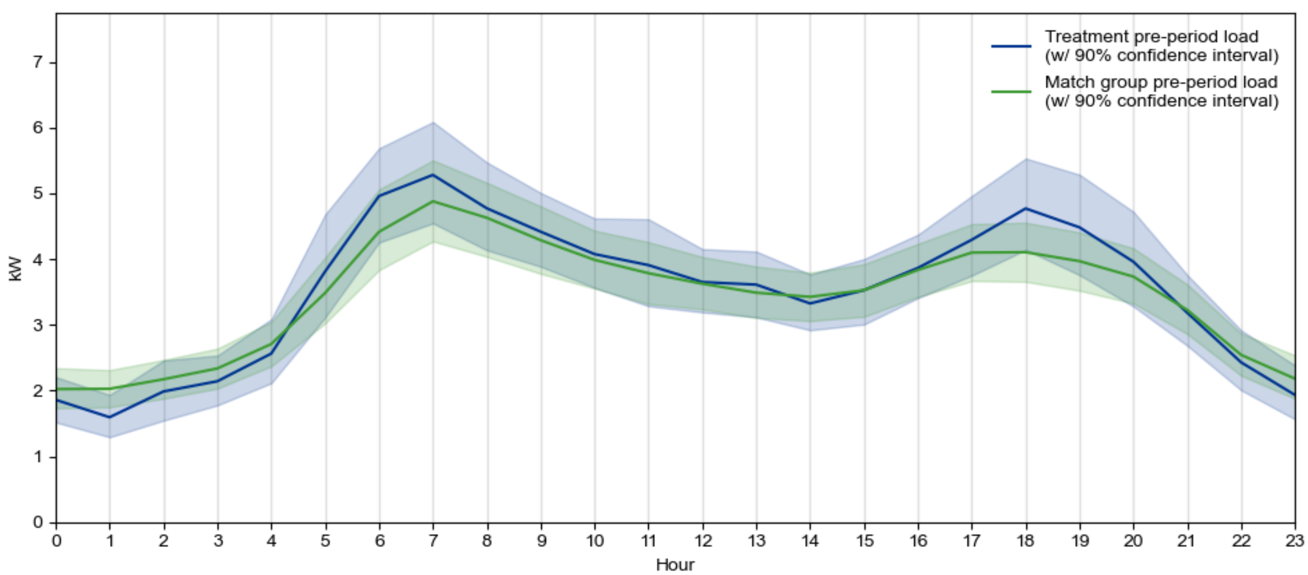


Figure 7-30. Average pre-event matching day load for the 12/22/2022 event

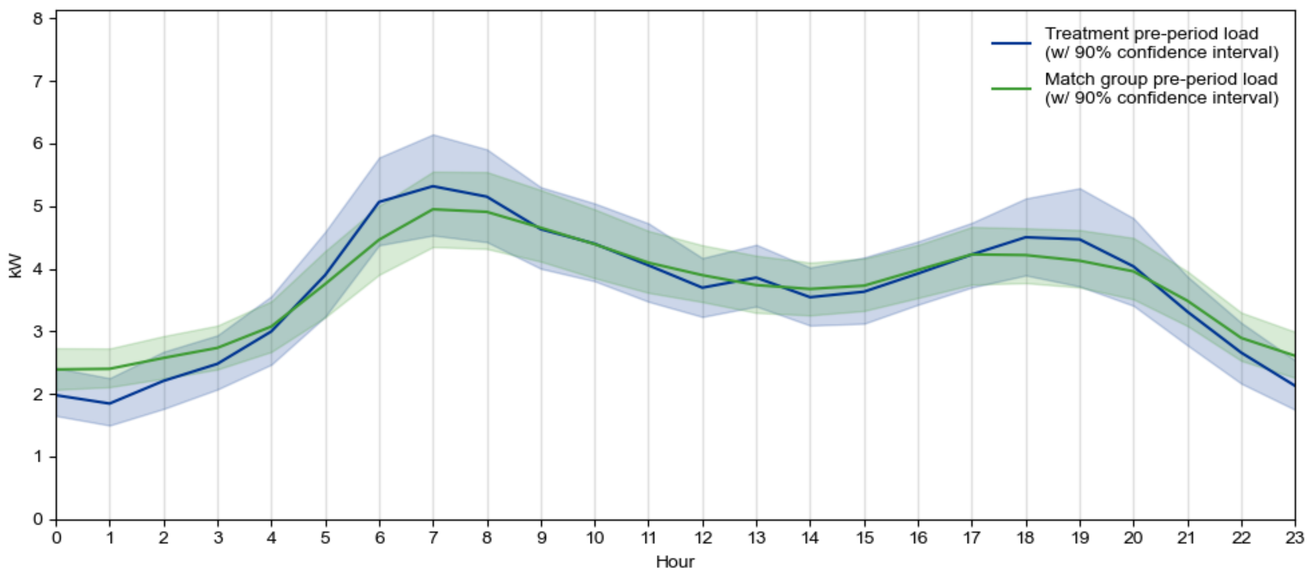


Figure 7-31. Average pre-event matching day load for the 1/30/2023 event

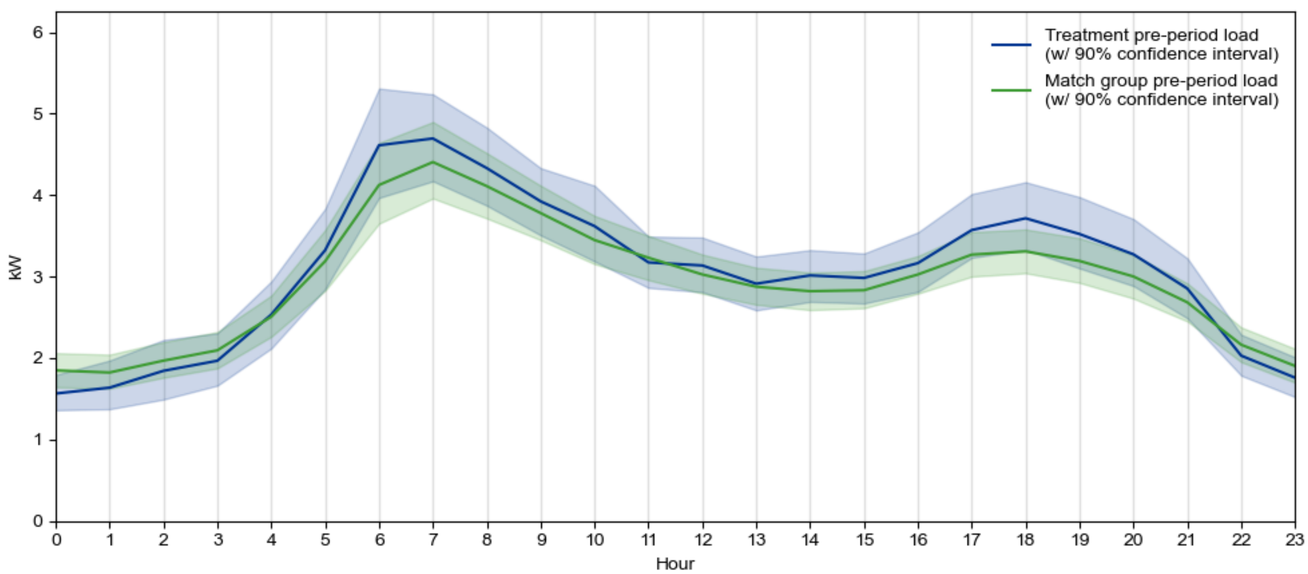


Figure 7-32. Average pre-event matching day load for the 1/31/2023 event

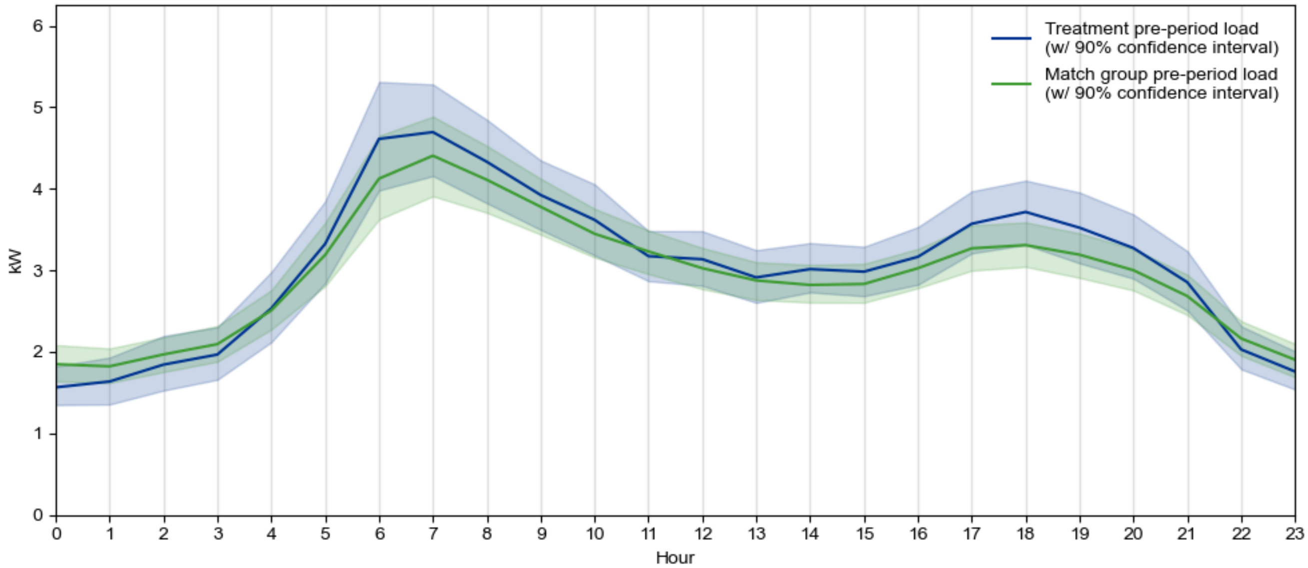


Figure 7-33. Average pre-event matching day load for the 2/23/2023 event

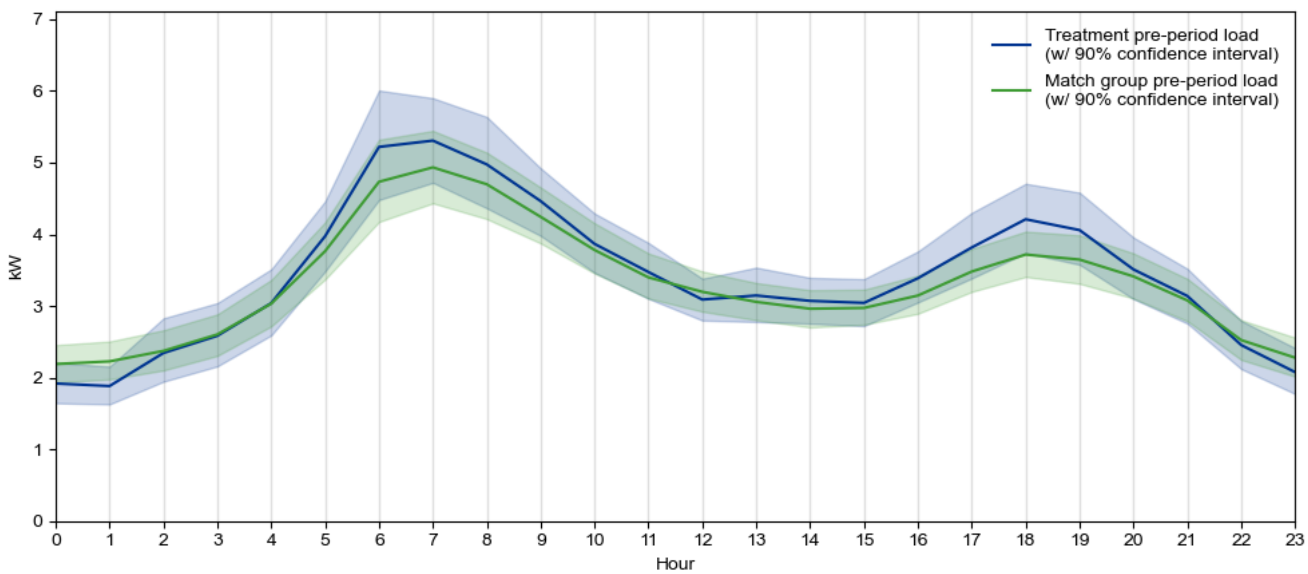
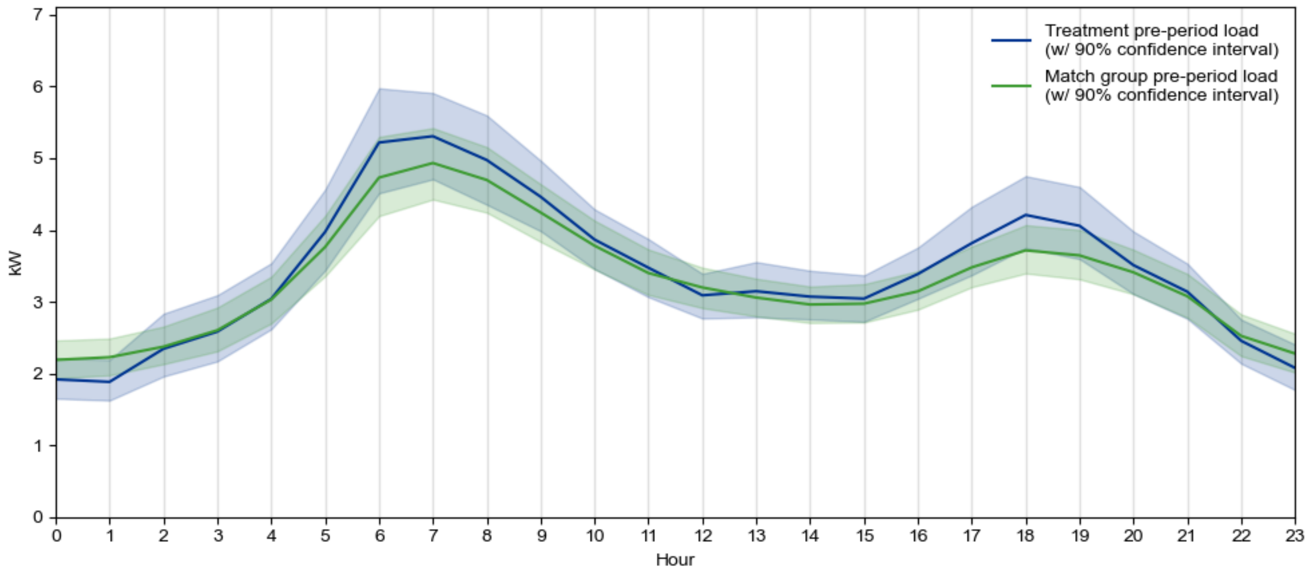


Figure 7-34. Average pre-event matching day load for the 2/24/2023 event



7.6.2 Gas load shapes

Figure 7-35. Average pre-event matching day load for the 12/20/2022 event

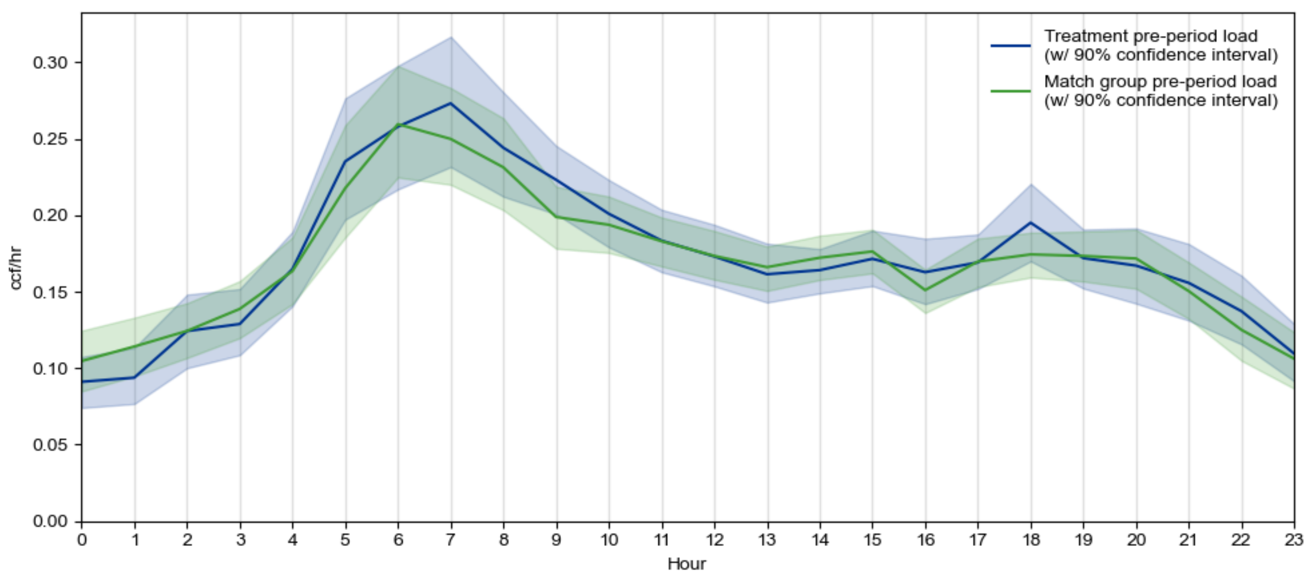


Figure 7-36. Average pre-event matching day load for the 12/22/2022 event

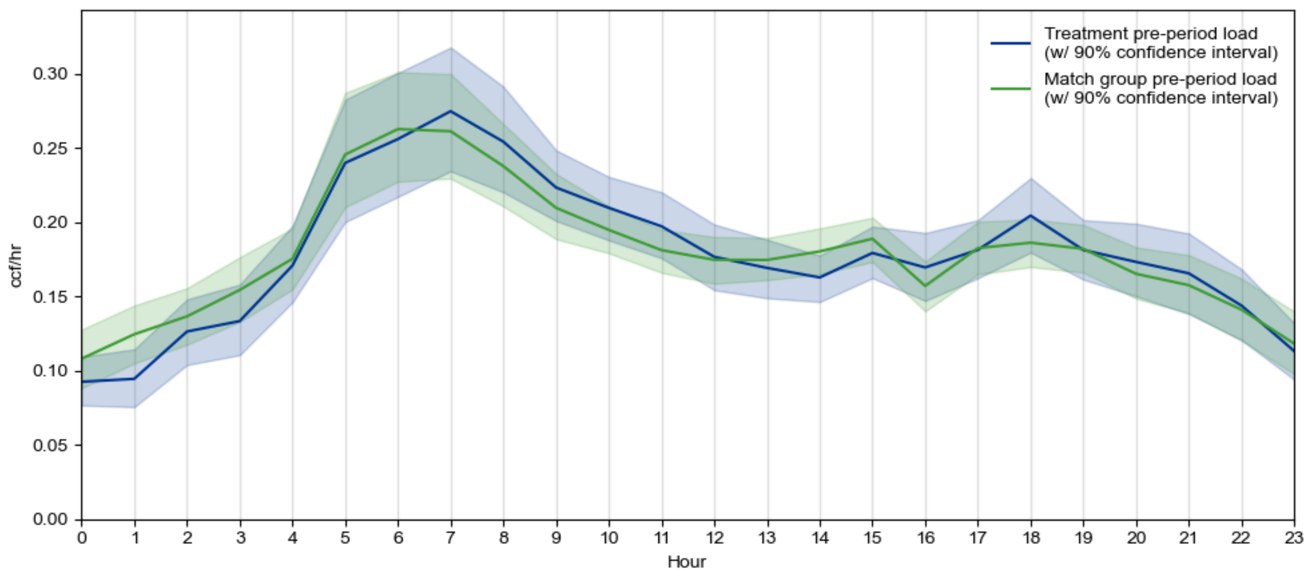


Figure 7-37. Average pre-event matching day load for the 1/30/2023 event

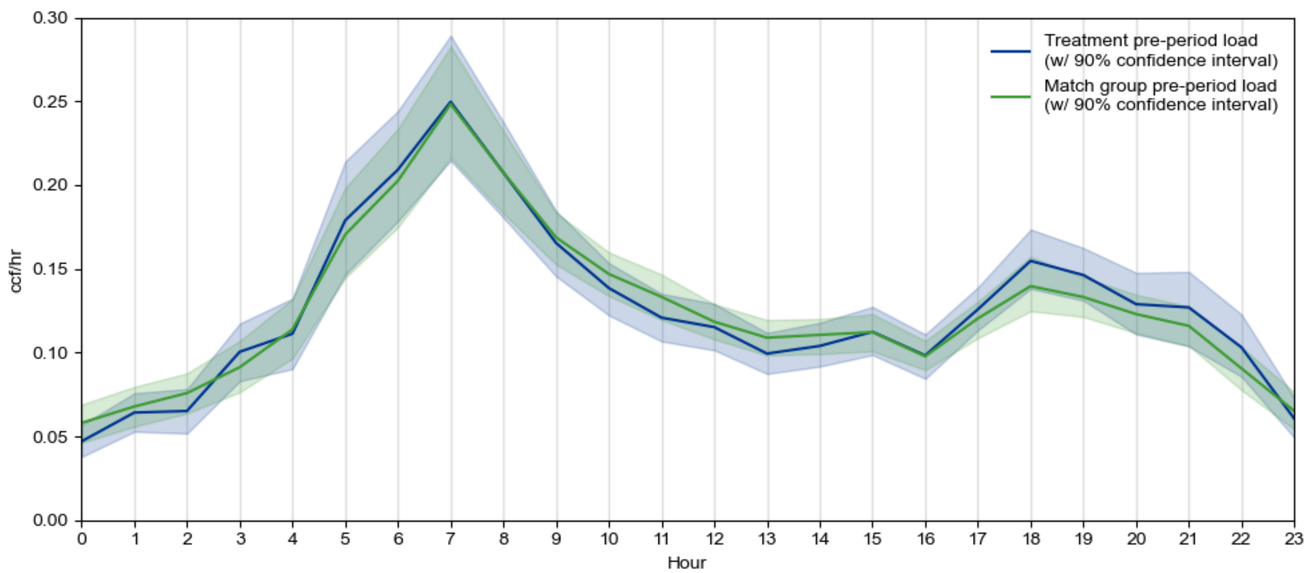


Figure 7-38. Average pre-event matching day load for the 1/31/2023 event

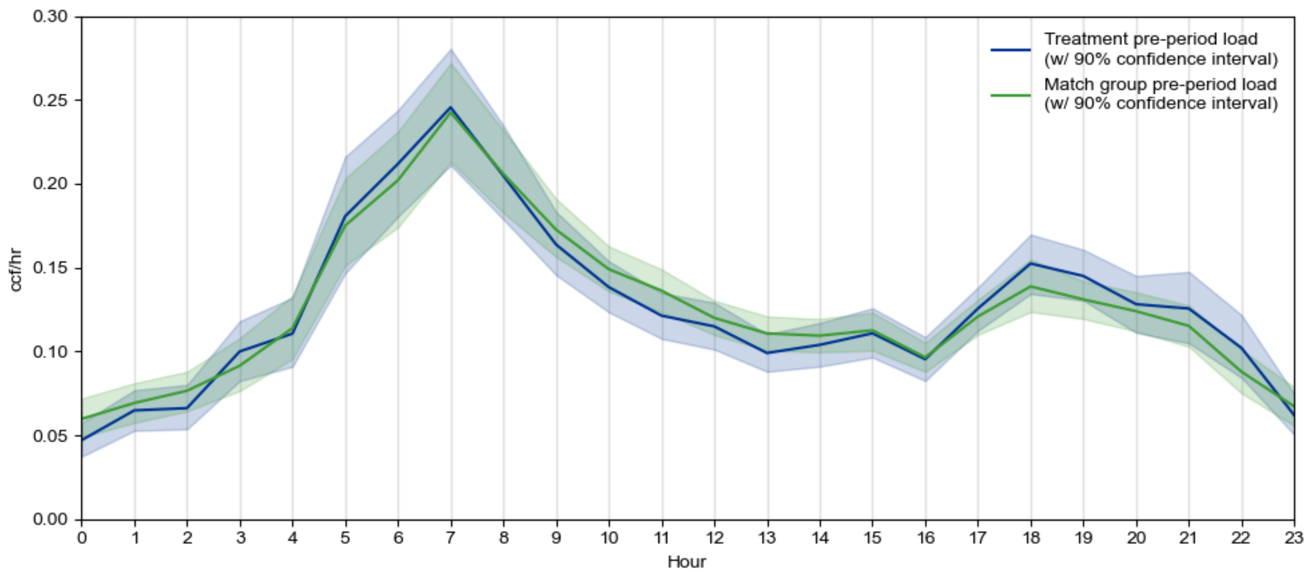


Figure 7-39. Average pre-event matching day load for the 2/23/2023 event

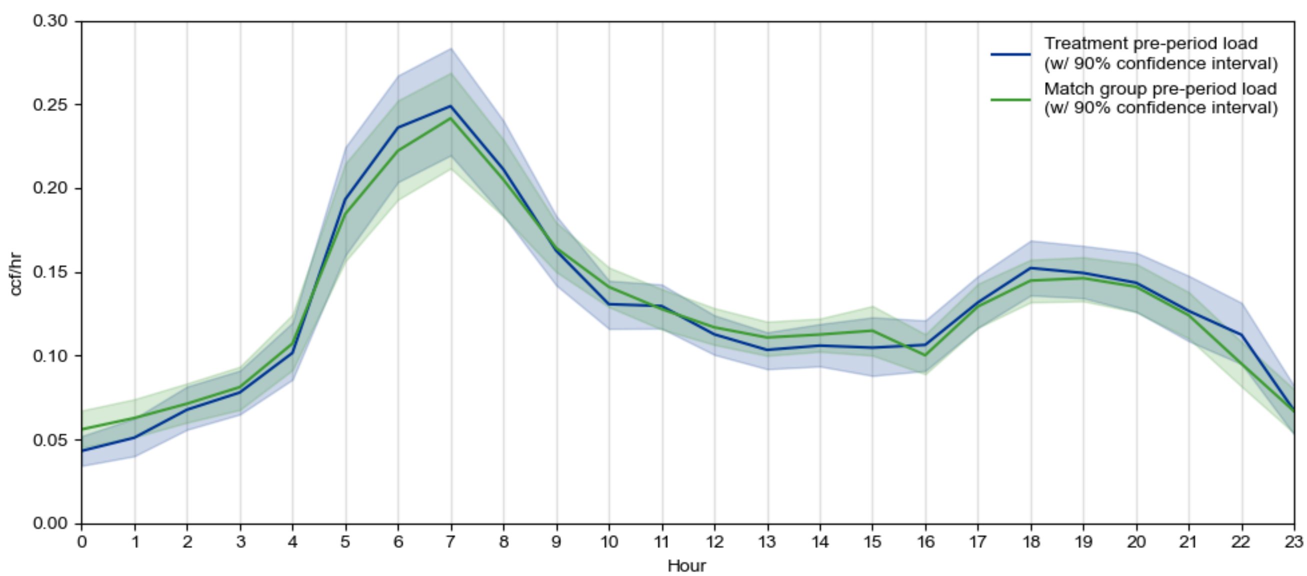
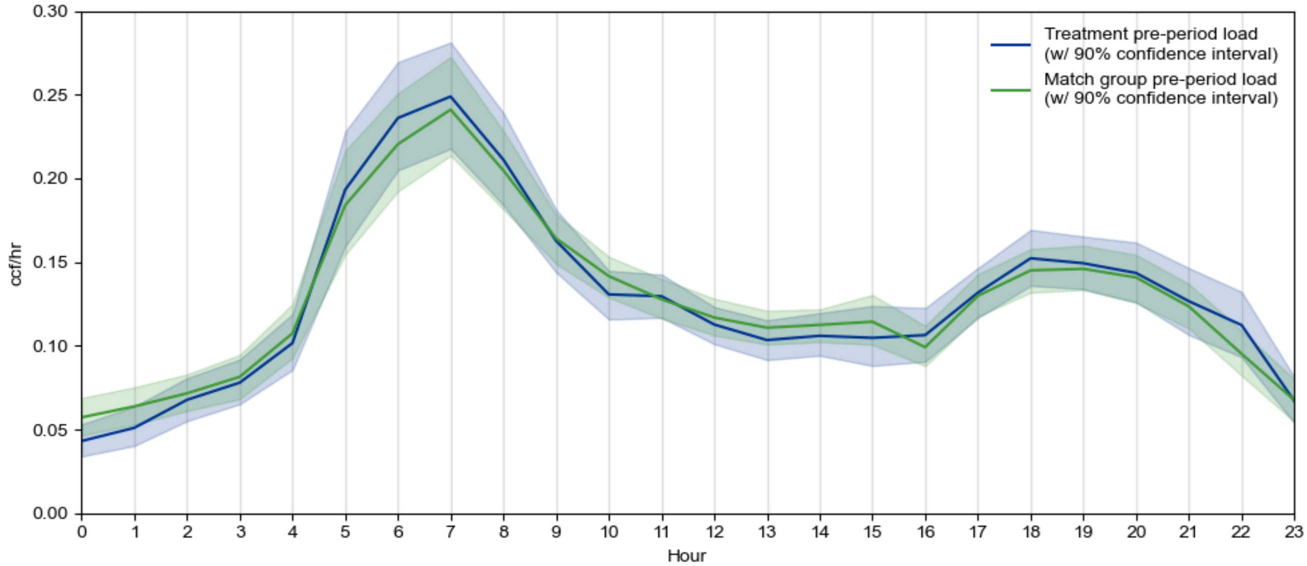


Figure 7-40. Average pre-event matching day load for the 2/24/2023 event



7.7 Appendix G: Hourly impacts by device brand

These plots show average per-premise hourly impacts, averaged across all events (except the 2/24 event). Impacts are separated by device brand to explore different behavioral patterns or algorithm designs associated with each device.

Figure 7-41. Average hourly impacts by device brand, averaged over event days, electric participants

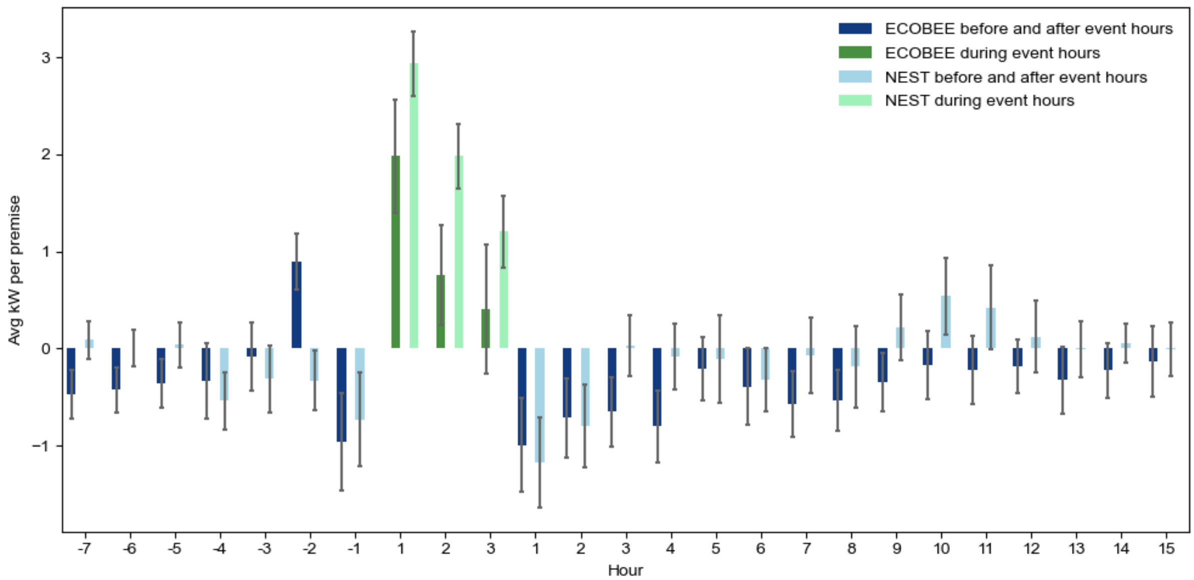


Figure 7-42. Average hourly impacts for Mysa device, averaged over event days, electric participants

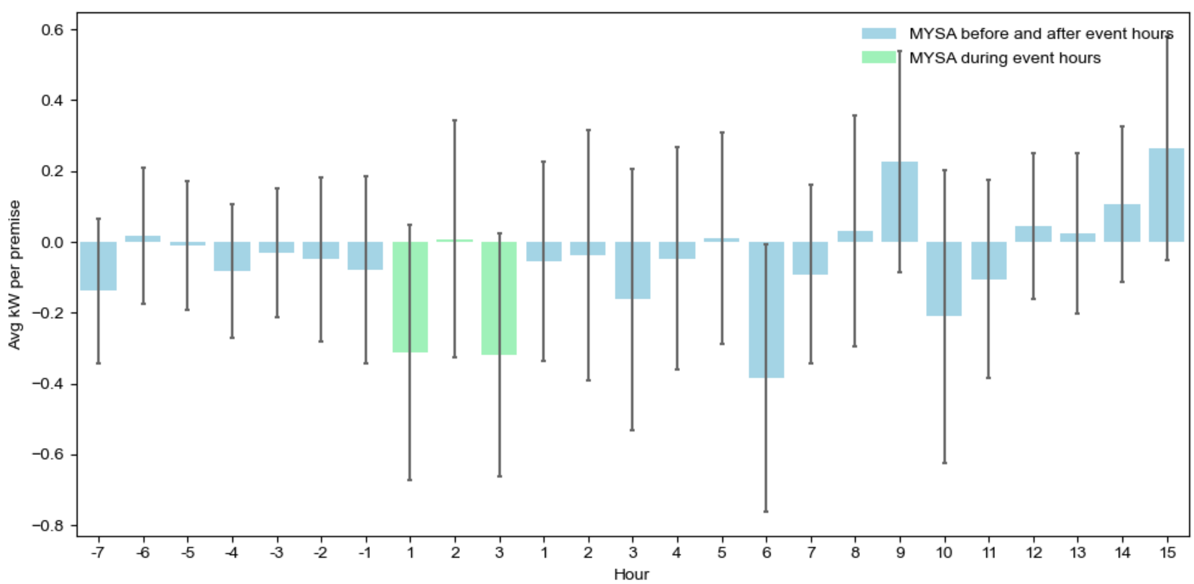
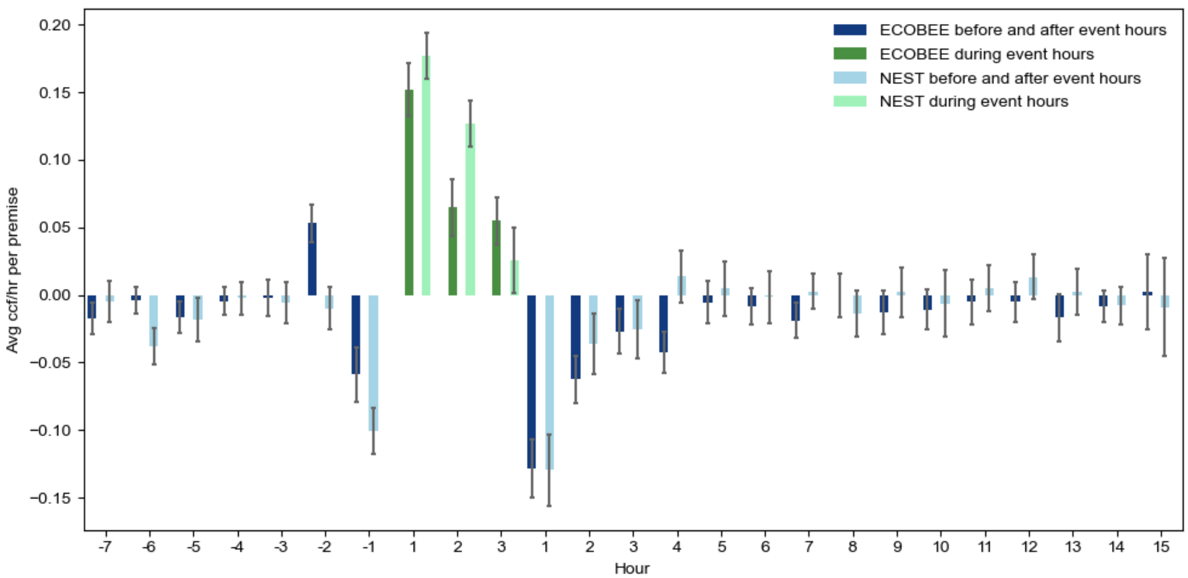


Figure 7-43. Average hourly impacts by device brand, averaged over event days, gas participants



About DNV

DNV is a global quality assurance and risk management company. Driven by our purpose of safeguarding life, property and the environment, we enable our customers to advance the safety and sustainability of their business. We provide classification, technical assurance, software and independent expert advisory services to the maritime, oil & gas, power and renewables industries. We also provide certification, supply chain and data management services to customers across a wide range of industries. Operating in more than 100 countries, our experts are dedicated to helping customers make the world safer, smarter and greener.