2019 National Grid Behavioral Demand Response Evaluation Findings

Findings Report - FINAL

Prepared for:

National Grid

Submitted by:
Guidehouse Inc.
77 South Bedford Street
Suite 400
Burlington, MA 01803

(781) 270-8300
guidehouse.com

Reference No.: 194701
April 28, 2020

Table of Contents

1. Introduction and Background 1
2. Literature Review 3
2.1 BDR Programs Nationwide 3
2.2 BDR Program Case Studies 5
Consumers Energy 10
DTE…….. 11
Efficiency Vermont/Green Mountain Power 12
PGE….. 13
2.3 Literature Review Findings 15
3. Methodology 17
3.1 Data Review 17
3.2 Within-Subject Regression 18
3.3 Quasi-Control Regression 21
4. Results 24
5. Findings & Conclusion 27
6. References 28
Appendix 29
A.1 Example BDR Email 29
A.2 SES Pilot Details 30
A.3 Additional Model Specification Results 30

1. Introduction and Background

The objective of this study was to estimate demand reductions from National Grid's 2019 behavioral demand response (BDR) program via a literature review and an analysis of a subset of National Grid Massachusetts customers with interval utility metering.

Within this report, BDR is broadly defined as a program where the utility asks customers to voluntarily reduce usage during certain hours, without direct financial incentives or the installation of technology.

In 2018, National Grid designed and implemented a BDR program for residential and small / medium business (SMB) customers. The Company sends an email to eligible customers for whom email addresses are available, encouraging them to voluntarily reduce usage on the following day during a specific window (e.g., 2–5 pm) when the annual peak demand is projected.[1]

National Grid called six events from 2–5 pm in 2018, and in 2019 they called two events from 2–6 pm. A sample email can be seen in Figure 7 in Appendix A.1. The Company's goal is to call an event that includes the annual ISO-NE system peak hour. As a result, BDR events can be called on weekends and holidays if needed. The National Grid Massachusetts program sends an email to approximately 400,000 residential and 21,000 SMB customers. National Grid does not have advanced metering infrastructure (AMI) data for most of their residential and SMB customers which limits their ability to provide customized data (i.e., post-event communications) or a direct monetary incentive to participants and the ability to evaluate hourly savings from the program.

As part of this evaluation, Guidehouse conducted a literature review of BDR programs nationwide and based on case studies of four BDR programs most similar to National Grid's, savings for residential BDR participants during event hours range from 1.3-4.0%, and average 2.3%, of whole home electricity usage. However, all four of these programs include personalized post-event communications with social comparisons of customers' savings during the event. Since this type of communication is not possible in National Grid's program, and evidence suggests that these types of comparisons are important in behavior programs,[2] these savings may be high compared to what National Grid can achieve with generic messaging.

To evaluate the impacts of National Grid's BDR program in 2019, Guidehouse used AMI data for customers from the Smart Energy Solutions (SES) pilot run in Worcester, Massachusetts from 2014 to 2018. Guidehouse conducted a within-subject regression analysis using similar non-event days to estimate the demand response impacts during event hours. Additionally, Guidehouse tested using SES customers who were not part of the BDR program (i.e., did not have an email address on file with National Grid) as a quasi-control group for estimating savings.

[1] In 2018, the program auto-enrolled all residential and SMB customers with an email address on file with National Grid; the email addresses are updated every week such that every customer with an email currently on file is auto-enrolled for each event. Customers have the option to opt-out of the program at any time.

[2] See, for example, Navigant 2015 or Erhardt-Martinez et. al. 2010.

For each model, the estimated event savings were not statistically significant. It is possible that an evaluation with access to a larger group of participants, or with a randomized control trial (RCT), may detect statistically significant savings in the future. Likewise, should National Grid have access to AMI data for more customers, the program design could be improved with customer-specific, post-event communications. An improved program design could lead to increased savings in the future.

The rest of this report presents the findings of our literature review and then discusses the methodology and results from the evaluation of 2019 BDR savings.

2. Literature Review

2.1 BDR Programs Nationwide

Guidehouse conducted a literature review to identify BDR programs that have been run by other utilities nationwide. We identified 10 programs from the last several years that are not combined with either real time data provision[3] or dynamic pricing, which offers a higher price or rebate during BDR events. The key attributes of these programs are summarized in Table 1.

[3] For example, programs where customers were provided an in-home display or mobile app that showed them their usage in (near) real time.

Table 1. Key Components of BDR Programs

Utility	Vendor	Independent Impact Evaluation	Process Evaluation of Behaviors Taken During Events	Opt-Out	Communication	Event Timing	Number of Events
National Grid	**In-house**	**NA**	**NA**	✓	**Not personalized pre-event**	**2018: 2-5 pm 2019: 2-6 pm**	**2018: 6 2019: 2**
Consumers Energy	Oracle	✓*	x	✓	Personalized pre- and post-event	2-6 pm	3
DTE Energy	Oracle	✓	x	✓	Personalized pre- and post-event	3-6 pm	2015: 6 2016: 10
Efficiency Vermont/Green Mountain Power	Oracle	✓*	x	✓	Personalized pre- and post-event	1-5 pm	4
Glendale Water & Power	Oracle	✓*	x	✓	Personalized pre- and post-event	1-5 pm	3
New Hampshire Electric Cooperative	AutoGrid	x	x	x	Not personalized pre-event	No more than 5 hours between 12-8 pm	No more than 15
Over 90 Electric Co-ops nationwide**	Apogee	x	x	x***	Varies***	Varies***	Varies***
Pacific Gas & Electric	Oracle	✓	x	✓	Personalized pre- and post-event	5-8 pm	4
Penn Power	Oracle	✓	x	✓	Personalized pre- and post-event	2-6 pm	3
Portland General Electric	AutoGrid	✓	✓	✓	Personalized pre- and post-event	4 hours between 3-8 pm	6
Hydro Ottawa	Oracle	✓	x	✓	Personalized pre- and post-event	12-5 pm	4

* Evaluation done by an academic organization for Oracle.
**Including Delaware Electric Cooperative, Vermont Electric Co-op, Duck River Electric Membership Cooperative, Old Dominion Electric Cooperative, Howell-Oregon Electric Cooperative, and Shenandoah Valley Electric Cooperative.
*** Implementation varies across the different co-ops.
Source: Guidehouse

Unfortunately, none of these BDR programs are directly comparable to National Grid's program. The biggest difference is the communication type; almost all of the other programs provide personalized pre- and post-event communication, whereas the National Grid program only sends pre-event communication that is not personalized due to the AMI data limitation.

To determine the programs to case study we considered the following criteria:

- Whether the program was opt-out
  - Opt-in programs have a fundamentally different calculus for participation that can lead to much different savings
- Whether there was an independent impact evaluation of the program
  - All of the opt out programs had one
- Whether there was a process evaluation assessing the types of actions customers took during events in response to the program
  - Only Portland General Electric (PGE) had one
- Type of communication
  - All of the opt out programs had personalized communications before and after the event, typically showing normative comparisons similar to Home Energy Reports (HERs)
- Representation across implementers
  - Of the opt-out programs, only PGE had an implementer other than Oracle[4]
- Geography
  - We prioritized evaluations in the Northeast and Midwest over evaluations on the West Coast
- Event timing
- Number of events
- Thoroughness of the impact evaluation report

These criteria led us to the following programs to conduct case studies of:

- Consumers Energy
- DTE Energy (DTE)
- Efficiency Vermont/Green Mountain Power
- PGE

2.2 BDR Program Case Studies

Guidehouse reviewed the program components (see Table 2 below) and impact evaluation methods and findings (see Table 3 below) for each of the four programs we conducted case studies of: Consumers Energy in Michigan, DTE in Michigan, Efficiency Vermont for Green Mountain Power, and PGE in Oregon. For PGE, we also reviewed the methods and findings of the process evaluation conducted for the program.

Across all four programs, savings for residential BDR participants during event hours range from 1.3-4.0%, and average 2.3%, of whole home electricity usage. There are some important similarities across these programs:

[4] Many of the programs run by Oracle were implemented by Opower before they were purchased by Oracle.

1. Each program included pre- and post-event communication including a personalized social-normative peer comparison
2. The programs were for residential customers only
3. Participants in all four programs had AMI meters and thus hourly data available for evaluation
4. A random control group was used for evaluation in each of the four programs

Particularly, the differences in communication may mean that these savings are high compared to what National Grid can achieve without personalized messaging. Additionally, available evidence from Business Energy Report programs (like HERs but for commercial customers) suggests that these programs are less effective in the commercial sector than the residential sector. Multiple studies of Business Energy Report programs have found no statistically significant savings.[5] Process evaluation research suggests that behavioral programs for commercial customers need to be precisely tailored to produce savings.[6] Although we cannot say for sure whether the BDR savings would be higher or lower for SMB customers as compared to residential, this evidence causes us to hypothesize that savings are likely lower.

[5] See, for example, Stewart 2015 or Nexant 2017.

[6] See, for example Cornish 2016.

Table 2. Program Components Summary

	National Grid	Consumers Energy	DTE	Efficiency Vermont/Green Mountain Power	PGE
Year(s) Evaluated	2018 & 2019 (Years run as this program has not been evaluated)	2014	2015 & 2016	2014	2016 & 2017
Customer Segments	Residential and SMB	Residential	Residential	Residential	Residential
Participation	Opt-out	Opt-out	Opt-out	Opt-out	Opt-out
Number of Participants	Approximately 400,000 residential and 21,000 SMB; no controls	109,612 (Treatment) 19,849 (Control)	2015 cohort: 39,984 (Treatment), 19,961 (Control) 2016 cohort: 71,932 (Treatment), 84,578 (Control)	31,958 (Treatment) 21,828 (Control)	2016: 6,233 (Treatment), 6,186 (Control) 2017: 10,089 (Treatment), 10,087 (Control)
Participant Targeting	All residential and SMB customers with an email address on file with National Grid	Random customer selection from eligible HER recipients*	2015 cohort: random selection from customers not receiving HERs or in the AC direct load control program 2016 cohort: random selection from HER recipients and not in AC direct load control program	Random selection from eligible customers who had not previously received HERs*	Random selection from eligible customers**
Participants had AMI?	No	Yes	Yes	Yes	Yes
Incentives for Participation	None	Some participants got an unconditional $5 gift voucher at the start of season	None	None	None

	National Grid	Consumers Energy	DTE	Efficiency Vermont/Green Mountain Power	PGE
Pre-event Communication	Day ahead email and/or automated phone call alerting participants of the event timing and providing them tips to save energy during the peak event	Day ahead email and/or automated phone call alerting participants of the event timing and providing them tips to save energy during the peak event and (for email only) information comparing their usage on the previous event day to their neighbors	Day ahead email and/or automated phone call alerting participants of the event timing and providing them tips to save energy during the peak event and (for email only) information comparing their usage on the previous event day to their neighbors	Day ahead email and/or automated phone call alerting participants of the event timing and providing them tips to save energy during the peak event and (for email only) information comparing their usage on the previous event day to their neighbors	Day ahead email alerting participants of the event and appeal to participate in collective actions to reduce usage. It is unclear if this pre-event email contained tips for savings or comparison information
Post-event Communication	None	Email and/or automated phone call after the event providing information on how participants performed during the event compared to their neighbors	Email and/or automated phone call after the event providing information on how participants performed during the event compared to their neighbors	Email and/or automated phone call after the event providing information on how participants performed during the event compared to their neighbors	Email after the event providing participants with information on their event performance results including social-normative peer comparisons.
Event Description	2018: 6 events between June and August from 2-5 pm 2019: 2 events in July from 2-6 pm	Three events between July and August called from 2-6 pm	2015: 6 events between July and September from 3-6 pm 2016: 10 events between July and September from 3-6 pm	Four events between July and September from 1-5 pm	2016: 6 events between July and August from 4-7 pm 2017: 6 events between July and September between 3 and 8 pm lasting 3 to 4 hours
Vendor	In-house	Oracle	Oracle	Oracle	AutoGrid

* The eligibility criteria included eliminating outliers and customers who had less than 12 months of usage data prior to the program.
** The eligibility criteria included (1) receive PGE electric service with at least 12 months of history, (2) not a solar customer, (3) not a participant in the Rush Hour Rewards thermostat control demand response program, (4) have a valid email address with PGE, and (5) have an AMI meter. Overlap with PGE's HER program was not discussed in the evaluation report.

Source: Guidehouse

Table 3. Impact Evaluation Methods and Findings Summary

	Consumers Energy	DTE	Efficiency Vermont/Green Mountain Power	PGE
Independent Evaluator	Academics from University of Chicago and Georgia State University (Brandon et. al., 2014)	Navigant (Kircher et. al., 2017)	Academics from University of Chicago and Georgia State University (Brandon et. al., 2014)	Cadmus (2018)
Impact Evaluation Description	Difference-in-difference model with a random control group	Lagged dependent variable model with a random control group	Difference-in-difference model with a random control group	Lagged dependent variable model with a random control group
Data	AMI data for treatment and control households for 3 events days and 12 non-event days	AMI data for treatment and control households on event days with controls for pre-program usage from the summer prior to the launch of BDR and the month prior to the BDR season	AMI data for treatment and control households for 4 events days and 12 non-event days	AMI data for treatment and control households on event days with controls for pre-program usage from the months before the BDR season
Impact During Event Hours (% of Whole Home Load)	1.3-2.0%	2015: 4.0% 2016: 3.0%	3.6%	2016: 1.3%* 2017: 2.3%
Impact During Event Hours (kW)	0.02-0.03	2015: 0.04 2016: 0.06	0.03	2016: 0.03 2017: 0.05
Impact During Entire Event Day (% of Whole Home Load)**	0.7-1.0%	NA	1.2%	NA

* In 2016, enrollment was relatively low, and PGE was still making major improvements to the program delivery.
** The entire event day considers any pre-cooling or other preparation before the event, snapback after the event, and any other intra-day load shifting that may occur.
Source: Guidehouse

In addition to the impact evaluation, Cadmus (2017) conducted an independent process evaluation of PGE's BDR program. Their findings indicate that customers who take action during events do many things other than just adjust their thermostat-controlled HVAC. For PGE, shifting chores like laundry and cooking (70% of respondents), reducing lighting (56%), and closing window blinds and curtains (54%) topped the list while raising the thermostat temperature (29%), using fans rather than AC (28%), and turning off the AC (24%) were lower down; note that customers were able to select multiple actions taken during events in this survey and as such the percentages are not additive. These findings align with process findings from another anonymous process evaluation of BDR where 40-60% (depending on the year) of customers who reported taking actions during events did something other than thermostat-controlled HVAC adjustment.[7] For that utility, other actions customers reported taking included adjusting lighting, turning off or avoiding usage of appliances like TVs, delaying laundry or dishwashing, and BBQing or cooking outside. These results are also supported by survey results from the SES pilot (which was not BDR but had some similar components) which show that approximately 35-40% of customers said they avoided the use of electricity intensive devices while only approximately 10% said they raised the AC temperature, see Figure 8 in Appendix A.2.

The remainder of this section provides more detailed descriptions of each of the four BDR programs and their evaluations.

Consumers Energy

Consumers Energy in Michigan evaluated their BDR program, implemented by Oracle, in the summer of 2014. That year they had three event days with events running from 1-6 pm.[8] The program included 129,461 residential customers who were randomly divided between 109,612 participants and 19,849 controls. The customers for the BDR program were randomly selected from households that had previously received Oracle HERs and passed basic eligibility criteria;[9] they were auto-enrolled in the program and given the opportunity to opt-out at any time.[10]

Participants in the program were randomly divided between four different treatments:

1. Social norm treatment (participants' usage compared to the neighborhood average)
2. Social rank treatment (participants usage rank within the neighborhood out of 100)
3. Social norm treatment and an unconditional $5 gift provided at the beginning of the season
4. Social rank treatment and an unconditional $5 gift provided at the beginning of the season

All participants received both pre- and post-event communications which they were randomly assigned to receive via email, interactive voice response phone calls, or both. Participants also received a post card at the beginning of the event season alerting them about the program. The pre-event communications came a day-ahead and told participants when the event would be occurring and gave them tips for how to reduce their energy usage. If they received information

[7] Of the same customers, 50-70% reported adjusting their AC.

[8] Events ran on 7/23/2014, 8/21/2014, and 8/22/2014.

[9] The eligibility criteria included eliminating outliers and customers who had less than 12 months of usage data prior to the program.

[10] Over the course of the summer approximately 3.5% of customers opted out of the program.

via email, the pre-event communication (with the exception of the first event) also contained information about their usage during the previous events compared to their neighbors.[11] The post-event communication told participants how they performed during the event compared to their neighbors. Across the three events, approximately 71% of both the pre- and post-event automated phone calls landed,[12] while the email open rate averaged 26% for pre-event emails and 41% for post-event emails.[13]

The program was evaluated by Brandon et. al. (2014) using AMI data comparing usage between the randomly assigned treatment and control customers. The evaluators used a difference-in-difference model which compared the difference in energy usage for participants during the event hours and over the entire day across event and non-event days to the same difference for controls. The model included usage on 12 non-event days which occurred on the same days of the week as the BDR events in the 4 weeks prior to the BDR season.[14] Separate models were run for the social norm and social rank treatment groups and each model also included coefficients to control for the effect of receiving the unconditional gift.

The evaluators found that participants decreased their usage during events by 1.3-2.0% compared to control customers. The savings did not differ statistically significantly between the social norm and social rank groups and the unconditional gift did not have a statistically significant impact on the savings. The impacts were relatively stable across the hours of the events and did not statistically differ between the different events of the season. The evaluators also found that participants saved energy across the entire event day compared to the controls at approximately half the magnitude of savings during event hours, i.e., 0.7-1.0% savings.

DTE

DTE in Michigan evaluated their BDR program, implemented by Oracle, in the summers of 2015 and 2016. They had six event days in 2015 and ten in 2016, all of which ran from 3-6 pm.[15] In 2015 the BDR program targeted customers who were not report recipients in DTE's HER program while in 2016 the BDR program targeted customers who were, see Figure 1. The customers for the BDR program were randomly selected from households who were or were not HER recipients, depending on the year, and were not part of DTE's AC direct load control program; they were auto-enrolled in the program and given the opportunity to opt-out at any time.[16]

[11] The presentation of this information differed across the social norm and social comparison group. The social norm group received information comparing them to the average usage of their neighbors while the social rank group received information ranking them among their neighbors.

[12] According to Brandon et. al. (2014), "calls landed refers to the proportion of customers for whom messages were left, so it is a measure of customers who ultimately received voice communication messages from Opower."

[13] The click-through rate was approximately 0.4% on pre-event emails and 0.9% on post-event emails.

[14] The BDR season began 6/16/2014, and the first event was not called until 7/22/2014. This meant the non-event days spanned 5/20/2014 to 6/13/2014 and there was approximately a five week gap between the non-event days and the first event day.

[15] Events in 2015 ran on 7/17, 7/28, 7/29, 8/14, 8/18, and 9/2. Events in 2016 ran on 7/6, 7/22, 7/27, 8/4, 8/5, 8/10, 8/11, 8/19, 8/30, and 9/7.

[16] In 2016 approximately 6.4% of participants opted-out of the program. This information is not available for 2015.

Figure 1. Sample Size of Participants in DTE BDR Analysis

	HER treatment group	HER control group	Neither	Total
2015 BDR treatment group	2,283	1,394	36,307	39,984
2015 BDR control group	1,139	635	18,187	19,961
2016 BDR treatment group	53,932	17,986	14	71,932
2016 BDR control group	63,620	20,929	29	84,578

Note: While the program targeted customers (1) for whom AMI data were available, and (2) were not on the interruptible air conditioning rate code, there were a small number of customers which were not screened out. The counts presented in this table, and the analysis that follows, exclude these customers. *Source:* Navigant.

Source: Kirchner et. al. (2017)

All participants received both pre- and post-event communications via email, interactive voice response phone calls, or both. Participants also received a letter at the beginning of the event season telling them about the program and providing tips for event days. The pre-event communications came a day-ahead and told participants when the event would be occurring and gave them tips for how to reduce their energy usage. If they received information via email, the pre-event communication (with the exception of the first event) also contained information about their usage during the previous events compared to their neighbors. The post-event communication told participants how they performed during the event compared to their neighbors. Phone call landing rates and email open rates are not available.

The program was evaluated by Navigant (Kirchner et. al., 2017) using AMI data comparing usage between the randomly assigned treatment and control customers. The evaluators used a lagged dependent variable model which compared event day energy usage for participants to controls. The model only included event days and controlled for non-event usage via a pre-period usage lag which controlled for average usage in the same hour and month as the event day from the year prior to the BDR program and a pre-season lag which controlled for average usage in the same hour during the most recent month without an event.[17]

The evaluators found that BDR participants decreased their usage during events by 4.0% in 2015 and 2.9% in 2016 compared to control customers.[18] The impacts were relatively stable across the hours of the events although the first hour had the highest savings after which impacts trended downward. The evaluators did not find evidence of event fatigue as impacts were relatively steady throughout the season. The evaluators also found evidence that participants saved energy in the hours shortly before and after the event but did not report out the event day savings.

Efficiency Vermont/Green Mountain Power

Efficiency Vermont evaluated their BDR program run for Green Mountain Power, implemented by Oracle, in the summer of 2014. That year they had four event days with events running from

[17] Note that weather differences from year to year are controlled for by virtue of having a random control group.
[18] Savings for the BDR plus HER group were slightly smaller than for BDR-only (2.9% versus 3.1%), though these differences were not statistically significant.

1-5 pm.[19] The program included 53,786 residential customers who were randomly divided between 31,958 participants and 21,828 controls. The customers for the BDR program were randomly selected from Green Mountain Power's households that had not previously received Oracle HERs after considering basic eligibility criteria;[20] they were auto-enrolled in the program and given the opportunity to opt-out at any time.[21] All participants in the program received the same treatment.

All participants received both pre- and post-event communications which they were randomly assigned to receive via email, interactive voice response phone calls, or both. Participants also received a post card at the beginning of the event season alerting them about the program. The pre-event communications came a day-ahead and told participants when the event would be occurring and gave them tips for how to reduce their energy usage. If they received information via email, the pre-event communication (with the exception of the first event) also contained information about their usage during the previous events compared to their neighbors. The post-event communication told participants how they performed during the event compared to their neighbors. Across the three events, approximately 79% of both the pre- and post-event automated phone calls landed, while the email open rate averaged 37% for pre-event emails and 53% for post-event emails.[22]

The program was evaluated by Brandon et. al. (2014) using AMI data comparing usage between the randomly assigned treatment and control customers. The evaluators used a difference-in-difference model which compared the difference in energy usage for participants during the event hours and over the entire day across event and non-event days to the same difference for controls. The model included usage on 12 non-event days which occurred on the same days of the week as the BDR events in the 4 weeks prior to the BDR season.[23]

The evaluators found that participants decreased their usage during events by 3.6% compared to control customers. The impacts were relatively stable across the hours of the events and did not statistically differ between the different events of the season. The evaluators also found that participants saved energy across the entire event day compared to the controls at approximately one-third the magnitude of savings during event hours.

PGE

In 2016 and 2017, PGE ran a Residential Pricing Pilot, implemented by AutoGrid, with many different configurations to test customer's acceptance and response to DR program design. One of the components they tested was opt-out BDR in the absence of any time varying rate, which is the portion of the program this memo will focus on. There were 6 events in each summer;

[19] Events ran on 7/23/2014, 8/20/2014, 8/26/2014, and 9/5/2014.

[20] The eligibility criteria included eliminating outliers and customers who had less than 12 months of usage data prior to the program.

[21] Over the course of the summer approximately 4% of customers opted out of the program.

[22] The click-through rate was approximately 0.8% on pre-event emails and 1.8% on post-event emails.

[23] The BDR season began 7/7/2014, and the first event was not called until 7/23/2014. This meant the non-event days spanned 6/10/2014 to 7/4/2014 and there was approximately a two week gap between the non-event days and the first event day.

events were also called in the winter but those will not be discussed in this memo. In 2016, all events ran from 4-7 pm, while in 2017 events ran for 3 to 4 hours between 3 and 8 pm.[24]

The BDR program has 6,233 participants in 2016 and 6,186 controls, while in 2017 it had 10,089 participants and 10,087 controls. Participants enrolled in 2016 were also enrolled in 2017 unless they opted-out or moved away. Over the two years of the pilot, the opt-out rate was approximately 3%. All participants in the BDR program received the same treatment.

All participants received both pre- and post-event communications via email. Participants also received a welcome email and postcard in mid-June 2016 alerting them about the program and an end of season report each summer detailing how much electricity was saved during the season (by them individually and by the program in aggregate). The pre-event communications came a day-ahead and told participants when the event would be occurring and included an appeal to participate in the collective action to reduce usage. It is unclear if the pre-event communication included tips to save energy or any social comparison. The post-event communication told participants how they performed during the event compared to their peers. Email open rates are not available, however according to a 2016 survey 77% of customers recalled receiving the event notifications.

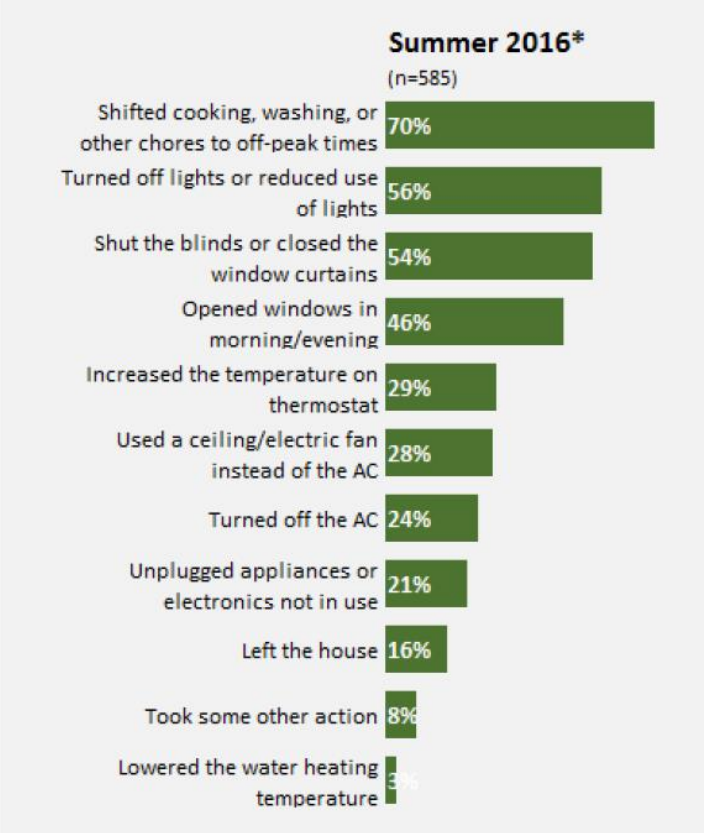
The program was evaluated by Cadmus (2017) using AMI data comparing usage between the randomly assigned treatment and control customers. The evaluators used a lagged dependent variable model which compared event day energy usage for participants to controls. The model only included event days and controlled for non-event usage via a pre-period usage lag which controlled for average usage in the morning and evening from April 1 through the last non-holiday weekday before the first event of the season in each year.

The evaluators found that participants decreased their usage during events by 1.3% in 2016 and 2.3% in 2017 compared to control customers. Cadmus' report primarily focused on 2017 as participation was relatively low in 2016 and the program was still undergoing major program delivery improvements (e.g., in deploying events, messaging customers, and providing participants with feedback). The evaluators did not provide information regarding impacts across the hours of the event, across the events of the season, or in the non-event hours of event days.

Cadmus also conducted a process evaluation including several surveys throughout the two year pilot. In a survey at the end of the 2016 summer event season, they asked customers whether they took action to save during events and if so, what actions they took. The survey had 160 completes for a 4% response rate. For BDR, 49% of respondents indicated that they took some action to save during the events. Figure 2 shows actions that customers self-reported taking during events; note that this graphic shows responses over all of PGE's event-based pilots, not just the BDR portion. Adjusting the thermostat does not occur until the fifth response; the top four responses are delaying chores like laundry and cooking (70%), adjusting lighting (56%), closing window blinds and curtains (54%), and opening windows (46%). The next three responses all related to AC usage: raising the thermostat temperature (29%), using fans rather than AC (28%), turning off the AC (24%).

[24] In 2016, events ran on 7/27, 7/29, 8/11, 8/12, 8/18, and 8/25. In 2017, events ran 7/25 4–7 pm, 8/1 5–8 pm, 8/3 4–8 pm, 8/9 3–6 pm, 8/28 4–8 pm, and 9/5 4:30–7:30 pm.

Figure 2. How Customers Conserved During PGE Events



Survey Question: How did you and your household conserve energy during Flex Time events? (Select all that apply)

*The Summer 2017 experience survey did not ask the event participation question. Results from the Summer 2016 event survey are reported here instead. Appendix F contains the survey results for Winter 2016/2017.

Note: This survey question was asked to customers in the event-based treatments (PTR-only, Hybrids, and Opt-Outs).

Source: Cadmus (2017)

2.3 Literature Review Findings

Across the four case studies, Guidehouse found savings for residential BDR participants during event hours range from 1.3-4.0%, and average 2.3%, of whole home electricity usage. Like National Grid's program, these programs are all opt-out, send pre-event notifications via email, and call fewer than ten events during the summer for three to four hours in the late afternoon. However, these programs had two important components in common that differ for National Grid:

1. Customers had AMI data allowing the program implementer to provide them with personalized communications telling them how they performed during events compared to other customers
2. Residential customers only

Behavior program research suggests that personalized peer comparisons are an important component to drive savings,[25] indicating that National Grid's program may achieve more limited savings from generic messaging compared to these other programs. We found no evidence to suggest how SMB customers perform in BDR programs, but evidence from Business Energy Report programs (like HERs but for commercial customers) suggests that these programs are less effective in the commercial sector than the residential sector.[26] Therefore, we hypothesize that savings are likely lower for SMB customers as compared to residential.

[25] See, for example, Navigant 2015 or Erhardt-Martinez et. al. 2010.
[26] See, for example, Stewart 2015 or Nexant 2017.

3. Methodology

The evaluation consisted of two approaches to estimate demand savings for each 2019 BDR event:

- Within-Subject Regression
- Quasi-Control Regression

This section describes the data used for the evaluation and the methodology for each of the evaluation team's approaches.

3.1 Data Review

Guidehouse utilized AMI data from customers in the SES pilot to estimate impacts from the BDR program in 2019. Since the SES pilot ran from 2014 to 2018, with its own demand response events, it was not possible to estimate 2018 BDR savings with this dataset.[27]

Table 4 summarizes the overlap between the SES pilot participants and BDR participants.[28] For this BDR evaluation, Guidehouse only utilized SES customers who did not opt-in to additional SES technology offerings and, as a result, had less exposure to the SES pilot.[29] In SES, all eligible customers were auto-enrolled[30] onto a Critical Peak Pricing (CPP) rate and could choose to switch to a Peak Time Rebate (PTR) rate or to opt out of SES by switching back to the Basic Rate. Level 1 customers in SES only had access to the pilot's web portal while Levels 2-4 also had access to enabling technology like in-home displays (IHDs), smart thermostats, and load control devices. Thus, our evaluation of BDR only included customers in Level 1 of SES (the level with no enabling technology). Note that since the SES pilot ended in 2018, these customers were not on CPP or PTR rates in 2019.

[27] As the effect would be confounded with that from the SES pilot.

[28] The number of BDR participants is based on cross-referencing account numbers for BDR participants with AMI data provided by Itron for June 1-16, 2019 and SES participation data. Only customers in all three datasets are included in this column.

[29] Guidehouse focused their analysis on Level 1 customers, which is the largest group who was least engaged with the SES pilot. Although the predicted baselines fit better for this group of customers, no group was found to have statistically significant savings.

[30] That is, the SES pilot was offered on an opt-out, as opposed to opt-in, basis.

Table 4. BDR SES Participant Overlap

SES Level*	Level Description**	SES Price Plan	Number of Overlapping BDR Participants***	Total Number of SES Participants†	Included in BDR Evaluation‡
Level 1 Active	AMI + web portal + mobile app; accessed web portal at least once during SES pilot	CPP	1,863	2,152	Yes
		PTR	127	139	Yes
Level 1 Passive	AMI + web portal + mobile app; never accessed web portal	CPP	4,399	7,659	Yes
		PTR	497	883	Yes
Level 2	Level 1 + IHD	CPP	607	726	No
		PTR	59	71	No
Level 3	Level 1 + smart thermostat	CPP	29	30	No
		PTR	3	4	No
Level 4	Level 1 + Level 2 + Level 3 + load control devices	CPP	215	240	No
		PTR	19	20	No
		Total	**7,818**	**11,924**	**-**
Total Included in BDR Study (Level 1)			**6,886**	**10,833**	**-**

* In Level 1, Active participants are those who visited the pilot web portal at least once.
** Complete descriptions of all the SES levels are available in Section 2.2 of Navigant (2017).
*** The number of BDR participants is based on cross-referencing account numbers for BDR participants with AMI data provided by Itron for June 1-16, 2019 and SES participation data. Only customers in all three datasets are included in this column, which is a subset of the approximately 50,000 total BDR participants in the Worcester area.
† Based on customer counts from Navigant (2019).
‡ Guidehouse focused their analysis on Level 1 customers, which is the largest group who was least engaged with the SES pilot. Although the predicted baselines fit better for this group of customers, no group was found to have statistically significant savings.
Source: Guidehouse analysis

The biggest disadvantage of the SES AMI dataset is that it is limited to areas in or near Worcester, MA where the SES pilot was implemented, whereas the BDR program participants are across National Grid's Massachusetts service area. The Worcester weather may differ from other parts of the service area or Worcester residents may not be representative, in terms of demographics or energy usage, of the broader customer population causing the impacts for the SES customers to be different from the BDR program as a whole. Additionally, because of their (up to four years of) participation in the SES pilot, these customers may be better informed about how to save or shift energy usage than the general BDR population. Conversely, passive participants from the SES pilot have already shown a penchant for being disengaged with National Grid programs.

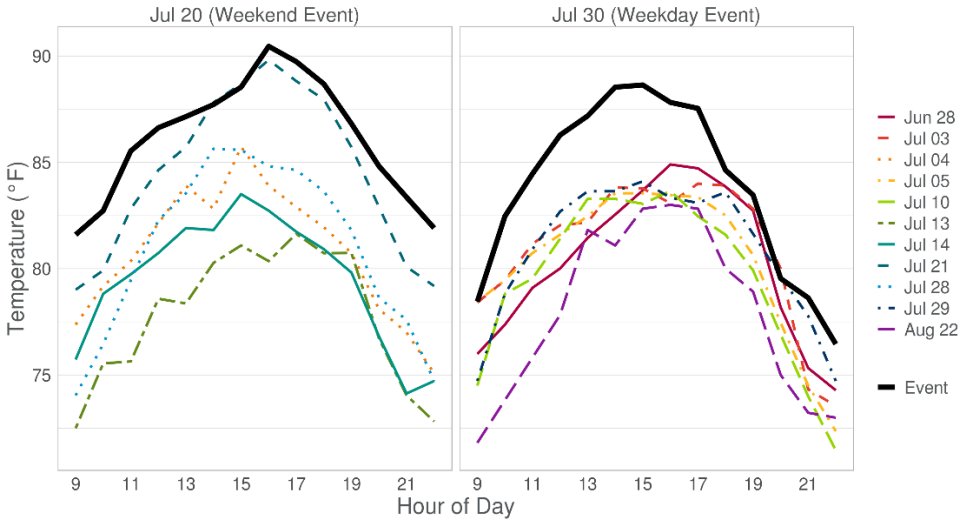
3.2 Within-Subject Regression

Because there is no random control group for the BDR program, Guidehouse first utilized a within-subject analysis to estimate 2019 BDR savings. The within-subject analysis utilizes only BDR program participants and constructs counterfactual usage (i.e., usage in the absence of a BDR event) based on participant's usage during other hot days of the 2019 summer. The difference between the predicted baseline usage and actual usage during BDR events is the

BDR program savings. We estimated the average treatment effect (ATE) which represents demand savings for all households called to participate in the BDR event.[31]

Guidehouse selected event-like non-event days to construct the counterfactual baseline usage for each event. The weekday event was matched to non-event, non-holiday weekdays and the weekend event was matched to non-event weekends or holidays. We chose non-event days that minimize the sum of the squared Euclidean distance in the temperature across the 14 hours of the day from 8 am to 10 pm, excluding days with unusual weather patterns (e.g. passage of cold fronts after thunderstorms). Figure 3 plots the weather on each event day in bold, along with the top five weather matches for each event day.[32]

Figure 3. Weather Matches



Source: Guidehouse analysis of NOAA data

Guidehouse then applied a linear regression framework using the event and selected non-event days to estimate the savings during each event. The regression model fits energy use as a function of a customer-specific fixed effect, events for participants, weather, and hour of the day. The model also captures up to 2 hours of (manual) precooling before the event starts and up to 3 hours of snapback after the event. Equation 1 provides the detailed model specification.

[31] We cannot estimate a treatment effect on the treated as we do not know who responds to the BDR emails by making changes to their usage and who does not.

[32] Weather matching uses weather data from the closes weather station to each participant's zip code. Therefore, participants in different zip codes may have different matched days, resulting in greater than five days in the "top five" weather matches.

Equation 1. BDR Event Within-Subject Regression Model

$$kW_{it} = \alpha_i + \sum_{j=1}^{24} \beta_{1j} H_{jt} + \sum_{k=1}^{2} \beta_{2k} pre_cool_counter_{kt} \cdot event_dummy_t + \sum_{m=1}^{4} \beta_{3m} event_counter_{mt} \cdot event_dummy_t + \sum_{n=1}^{3} \beta_{4n} sb_counter_{nt} \cdot event_dummy_t + \beta_5 cdh70_{it} + \beta_6 NHBU_{it} + \beta_7 MA4_cdh70_{it} + \beta_8 MA24_cdh70_{it} + \varepsilon_{it}$$

Where:

kW_{it}	is consumption (kW) in hour *t* by participant *i*
α_i	is a customer-specific fixed effect for participant *i*
H_{jt}	is an hourly dummy equal to 1 in hour *t*
$pre_cool_counter_{kt}$	is a dummy equal to 1 when hour *t* is in the k^{th} hour before the event
$event_dummy_t$	is an event dummy equal to 1 when hour *t* falls during the event
$event_counter_{mt}$	is a dummy equal to 1 when hour *t* is in the m^{th} hour of the event
$sb_counter_{nt}$	is a dummy equal to 1 when hour *t* is in the n^{th} hour of snapback after the event
$cdh70_{it}$	is cooling degree hours with base 70 for participant *i* in hour *t*
$NHBU_{it}$	is normalized heat buildup for participant *i* in hour *t*
$MA4_cdh70_{it}$	is the four-hour moving average of cooling degree hours with base 70 for participant *i* in hour *t*
$MA24_cdh70_{it}$	is the 24-hour moving average of cooling degree hours with base 70 for participant *i* in hour *t*
ε_{it}	is the error term

This model yields per participant average demand savings (or dissavings) during each hour of precooling, the event, and snapback. The values can be used to determine demand savings for the population during each hour of the event and energy savings for the day across precooling, the event, and snapback.

3.3 Quasi-Control Regression

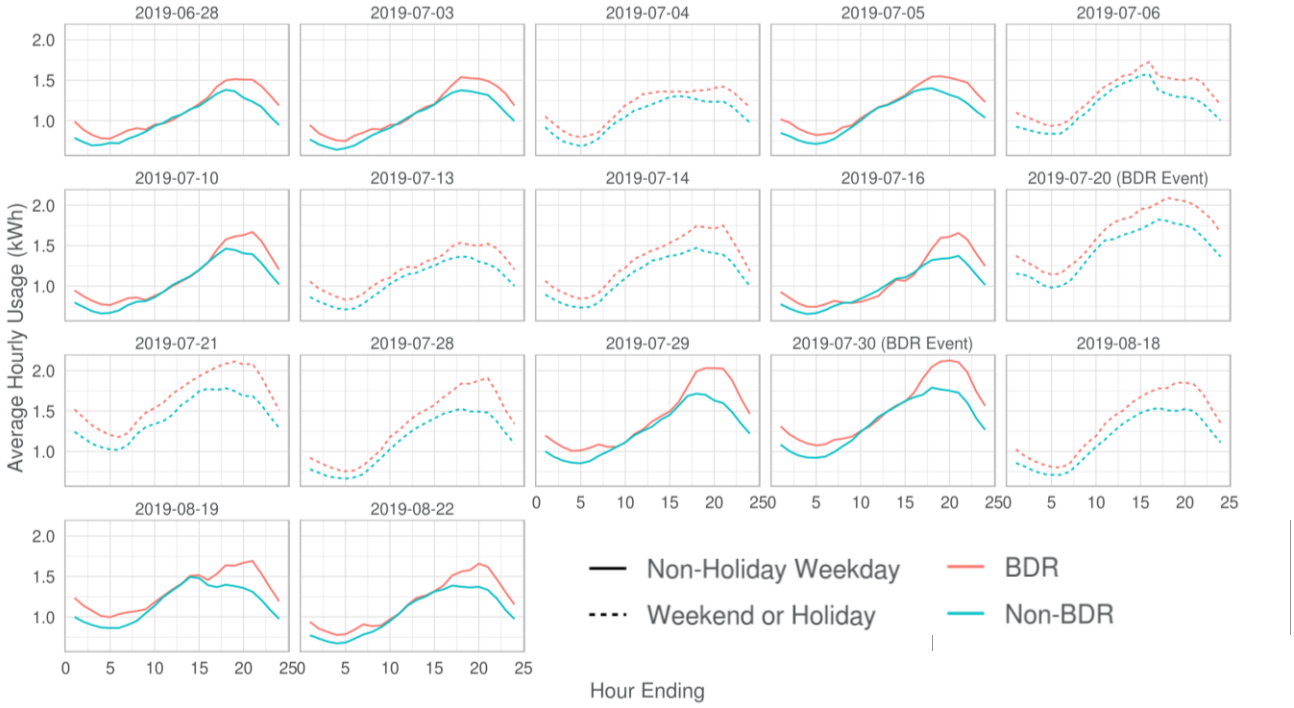
Guidehouse observed that the weather on event days was considerably hotter than the weather on the best matched control days. Therefore, in addition to the within-subject model, Guidehouse explored a regression specification that utilized the SES participants without email addresses as a quasi-control group for the BDR participants. Table 4 from the previous section contains the counts of participants from the SES pilot who were enrolled in the BDR program. The Level 1 Passive group was the only group with enough non-BDR participants to serve as an adequate control group (3,646 quasi-controls compared to 4,896 treatment customers). Even for this group, there were not enough non-BDR participants to conduct participant matching. Instead, Guidehouse used the entire non-BDR group as the quasi-control group.

Before using the non-BDR participants as quasi-controls, Guidehouse checked that their average usage was comparable to that of the BDR participants for hot days during the summer of 2019. Figure 4 shows the average usage profiles for the BDR and non-BDR participants on selected hot days,[33] including the event days, for the SES Level 1 Passive group. The BDR participants' usage is consistently higher than the non-BDR participants' usage, especially during weekday peak periods. To reduce this difference, Guidehouse removed the top 3% of the BDR group and bottom 3% of the non-BDR group by usage, which brought the usage curves closer together during non-event hours.[34]

[33] Figure 4 includes the top five matched control days from the within-subject analysis, as well as any other day that fell within the top best matched days.

[34] Guidehouse tested a range of threshold values for removing participants on the tails, from 1% to 10%. The 3% threshold resulted in the most overlap in usage curves between the two groups on event days outside of event hours.

Figure 4. Average Usage Profiles for All Level 1 Passive BDR and Non-BDR Participants



Source: Guidehouse analysis

Guidehouse then applied a linear regression framework using the BDR and non-BDR participants on the event and selected non-event days to estimate the savings during each event. Like the within-subject model, the regression model fits energy use as a function of a customer-specific fixed effect, events for participants, weather, and hour of the day. The model also captures up to 2 hours of (manual) precooling before the event starts and up to 3 hours of snapback after the event. Equation 2 provides the detailed model specification.

Equation 2. BDR Event Quasi-Control Regression Model

$$kW_{it} = \alpha_i + \sum_{j=1}^{24} \beta_{1j} H_{jt} \cdot event_dummy_t + \sum_{k=1}^{2} \beta_{2k} pre_cool_counter_{kt} \cdot event_dummy_t \cdot treatment_i + \sum_{m=1}^{4} \beta_{3m} event_counter_{mt} \cdot event_dummy_t \cdot treatment_i + \sum_{n=1}^{3} \beta_{4n} sb_counter_{nt} \cdot event_dummy_t \cdot treatment_i + \beta_5 cdh70_{it} + \beta_6 NHBU_{it} + \beta_7 MA4_cdh70_{it} + \beta_8 MA24_cdh70_{it} + \varepsilon_{it}$$

Where all variables are the same as Equation 1 and:

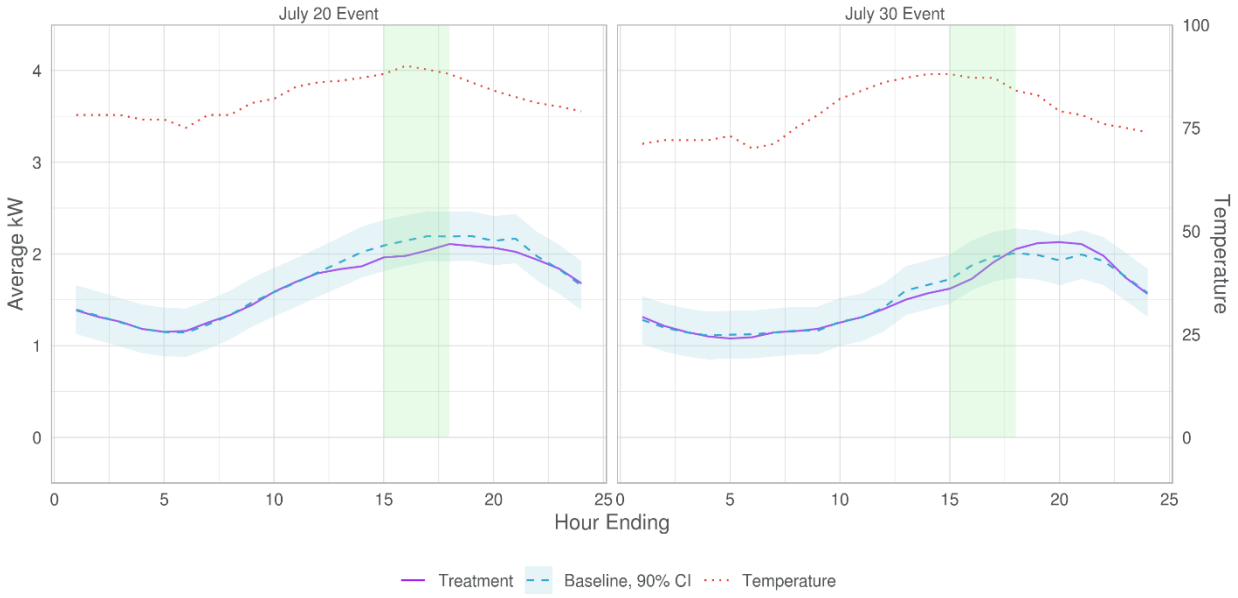
$treatment_i$ is a dummy equal to 1 when participant *i* is a BDR participant

Like the within-subject model, this model yields per participant average demand savings (or dissavings) during each hour of precooling, the event, and snapback. The values can be used to determine demand savings for the population during each hour of the event and energy savings for the day across precooling, the event, and snapback.

4. Results

Guidehouse did not find statistically significant savings for the 2019 BDR events from any method. Figure 5 shows the average usage of the BDR participants (Treatment) for the July 20th and July 30th events. The event period hours, 2-6 pm, are shaded in green. The dashed blue line shows the predicted baseline with a shaded 90% confidence interval. Because the purple line (which represents actual load of BDR participants) is completely contained within the blue shaded confidence interval, the BDR participants usage is not statistically different from what it would have been in the absence of the event.

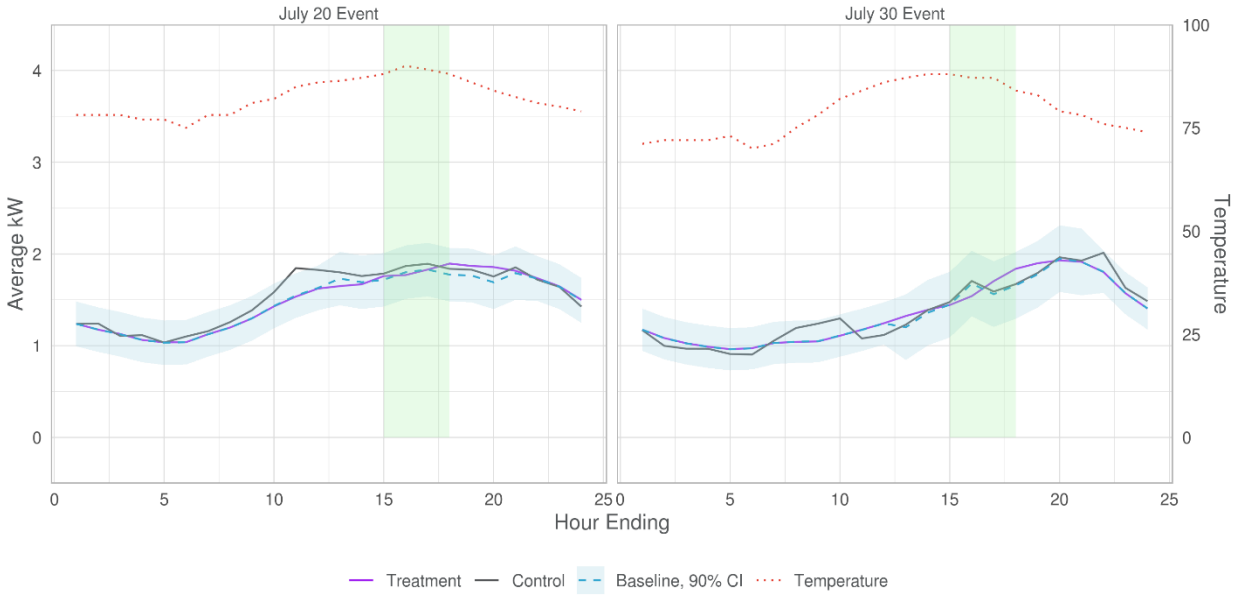
Figure 5. Within-Subject Predicted Baseline for 2019 Events*



* Level 1 Passive customers, with the single best matched non-event day for each event. We also did not find statistically significant savings for any of the other SES groups.
Source: Guidehouse analysis

Figure 6 shows the average usage for both the treatment and quasi-control groups for the July 20th and July 30th events from the quasi-control model. The predicted baseline is roughly the same as the treatment group's average usage. For this model, the top 3% of the treatment group and bottom 3% of the control group were excluded to better align the usage profiles of the two groups. Adding the quasi-control group creates a slightly more accurate baseline outside of the precooling, event, and snapback hours, but there are still no demand savings during the event hours.

Figure 6. Quasi-Control Predicted Baseline for 2019 Events*



* Level 1 Passive customers, with the single best matched non-event day for each event. Excludes the top 3% of treatment group and bottom 3% of control group. We also did not find statistically significant savings for any of the other SES groups.
Source: Guidehouse analysis

For some model specifications (e.g., the within-subject model for Level 1 Passive customers), the participants' usage is below the predicted baseline. If seen consistently, this could indicate event demand savings, even if they are not statistically significant, but this result was not consistently seen across models. In fact, many predicted baselines were below the participants' actual usage, indicating increased usage during the event. Because of the inconsistency in savings versus dissavings, and because of the wide confidence bands on the predicted baseline, we cannot conclude that there is any change in usage resulting from the BDR treatment. A visualization of the baselines from all model specifications can be found in Appendix A.3.

Although the estimated savings were not statistically significant, Guidehouse calculated point estimates and confidence intervals for the BDR impact for each event. Table 5 shows the calculated savings for each event along with a 90% confidence interval. Because each confidence interval contains zero, the calculated point estimates are not statistically significant.

Table 5. BDR Event Impacts by Modeling Approach for 2019 Events

Approach	Event	DR Impact*	90% CI High	90% CI Low
Within-Subject	July 20	0.13 kW	0.41 kW	-0.14 kW
	July 30	0.07 kW	0.33 kW	-0.20 kW
	Average	**0.10 kW**	**0.37 kW**	**-0.17 kW**
Quasi-Controls	July 20	-0.03 kW	0.26 kW	-0.32 kW
	July 30	-0.05 kW	0.31 kW	-0.40 kW
	Average	**-0.04 kW**	**0.29 kW**	**-0.36 kW**

* Positive values indicate a decrease in usage during the BDR event, 2-6 pm, while negative values indicate an increase in usage.
Source: Guidehouse analysis

5. Findings & Conclusion

National Grid's BDR program is a behavioral program with no enabling technology or customer specific communication. Event savings for similar programs range from 1.3% to 4.0% pf whole home load. Guidehouse conducted within-subject and quasi-controlled regression analyses utilizing customer AMI data from the SES pilot to estimate savings from National Grid's 2019 BDR events. The evaluation team did not find statistically significant demand savings for the BDR program's two events in the summer of 2019.

Based on the literature review, behavioral programs like National Grid's BDR program can achieve demand savings. Guidehouse's analysis relied on a subset of the BDR participants that were enrolled in the National Grid SES pilot from 2014 through 2018. It is possible that an evaluation with access to a larger group of participants, or with a randomized control trial (RCT), may detect statistically significant savings in the future. Likewise, should National Grid have access to AMI data for more customers, the program design could be improved with customer-specific, post-event communications. An improved program design could lead to increased savings in the future.

6. References

Brandon, A., J. List, R. Metcalfe, M. Price. 2014. *The Impact of the 2014 Opower Summer Behavioral Demand Response Campaigns on Peak-Time Energy Consumption*. Report written for Opower. http://www.oracle.com/us/industries/utilities/opower-summer-bdr-2015-5225428.pdf

Cadmus. 2018. *Flex Pricing and Behavioral Demand Response Pilot Program.* Prepared for Portland General Electric. https://edocs.puc.state.or.us/efdocs/HAH/um1708hah16432.pdf

Cornish, L., B. Karlin, A. Sanguinetti, J. Kaufman. 2016. *How Do Small Businesses Experience Energy Reports?* ACEEE Summer Study on Energy Efficiency in Buildings: Pacific Grove, CA. https://aceee.org/files/proceedings/2016/data/papers/8_519.pdf

Erhardt-Martinez, K., K. Donnelly, J. Latiner. 2010. *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities*. https://www.smartgrid.gov/files/ami_initiatives_aceee.pdf

Kirchner, D., D. Brannan, C. Olig, and W. Sierzchula. 2017. *The Reliability of Behavioral Demand Response*. International Energy Program Evaluation Conference: Baltimore, MD. http://www.oracle.com/us/industries/utilities/reliability-bdr-5225436.pdf

Navigant. 2015. *Home Energy Report Pilot Program Evaluation*. Prepared for Eversource New Hampshire. http://www.puc.state.nh.us/electric/Monitoring%20and%20Evaluation%20Reports/Eversource%20NH%20HER%20Evaluation%20Report%202016-03-24%20Final.pdf

Navigant. 2017. *National Grid Smart Energy Solutions Pilot*. Prepared for National Grid. (D.P.U 10-82) https://fileservice.eea.comacloud.net/FileService.Api/file/FileRoom/9179984

Navigant. 2019. *National Grid Smart Energy Solutions 2017 and 2018 – FINAL.* Prepared for National Grid. https://fileservice.eea.comacloud.net/FileService.Api/file/FileRoom/11177198

Nexant. 2017. *Business Energy Reports Evaluation*. Prepared for Pacific Gas and Electric Company. https://www.etcc-ca.com/reports/business-energy-reports-evaluation?dl=1570647505

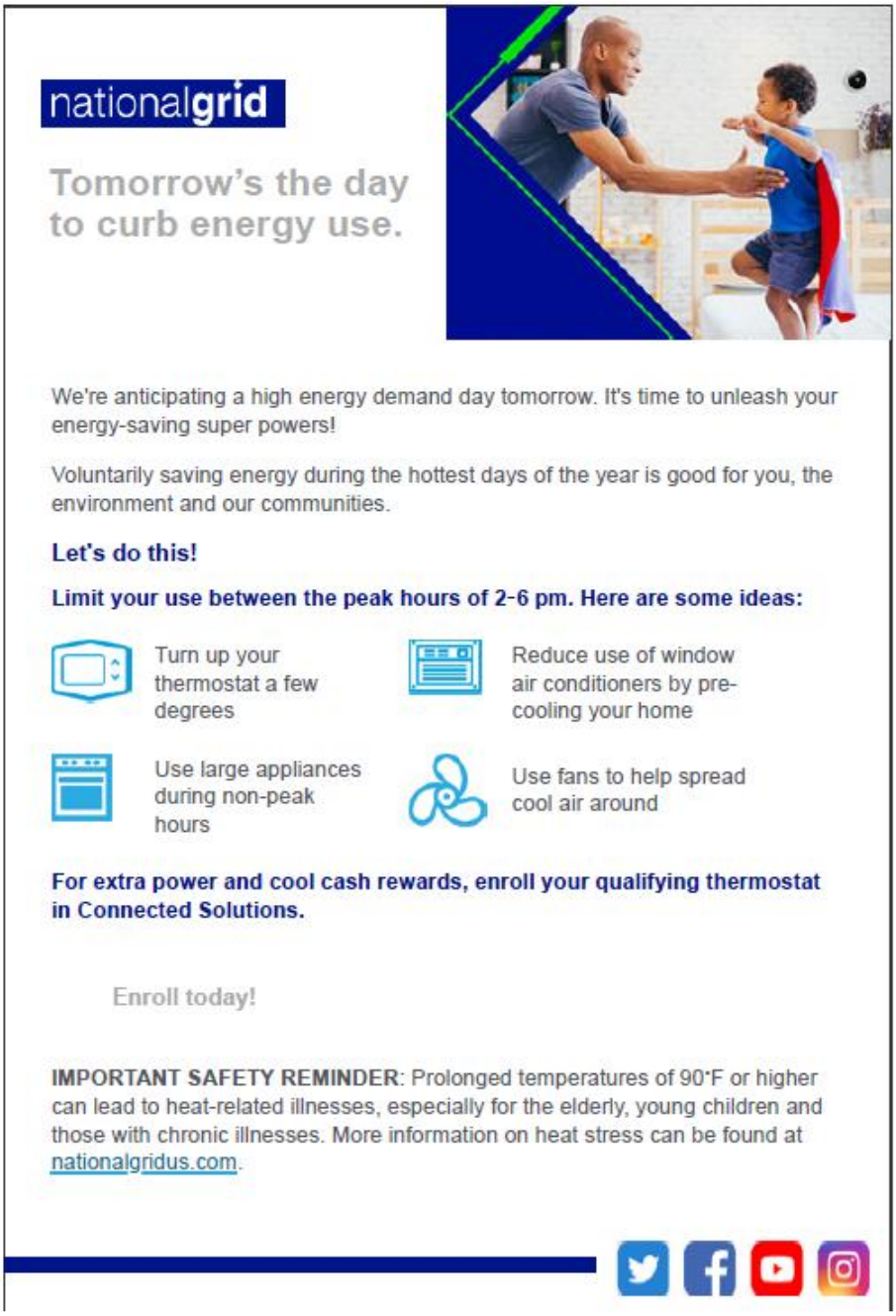
Stewart. 2015. *Energy Savings from Business Energy Feedback.* Behavior, Energy, and Climate Change Conference: Sacramento, CA. https://beccconference.org/wp-content/uploads/2015/10/presentation_stewart.pdf

Appendix

A.1 Example BDR Email

Figure 7 shows the BDR email sent by National Grid in 2019. Note the email was the same for the two events.

Figure 7. 2019 BDR Email

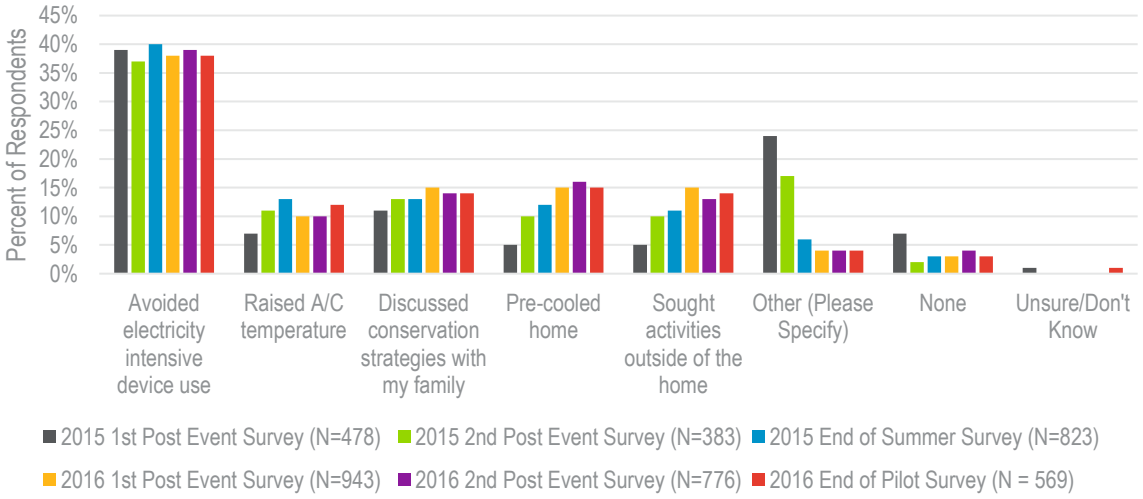


Source: National Grid

A.2 SES Pilot Details

Figure 8 shows reported actions customers took during SES pilot conservation day events. Many more respondents (approximately 40%) reported avoiding electricity intensive devices than raising their AC temperature (approximately 10%).

Figure 8. Actions SES Participants Took to Reduce Electricity Usage on Conservation Days



Source: Navigant (2017)

A.3 Additional Model Specification Results

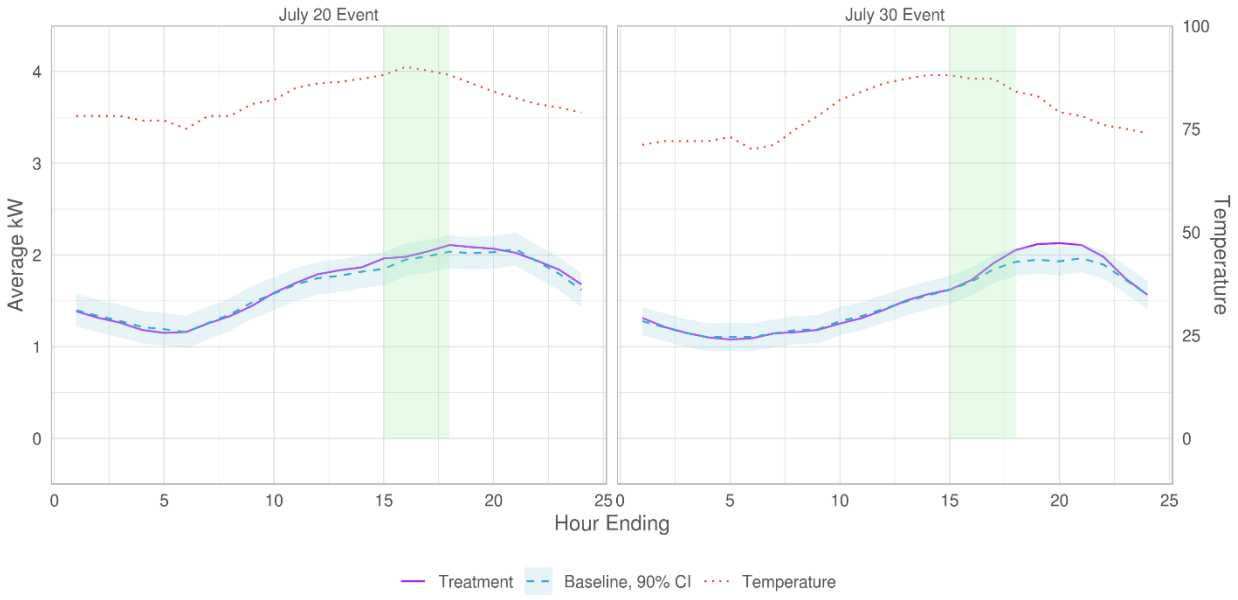
Guidehouse tested six model specifications to estimate event impacts. The **Top Weather Matches** specification was chosen as the final "within-subject" model, and the **Top Weather Matches, Top Quasi-Controls** specification was chosen as the final "quasi-controls" model. These models were chosen because they produced the best-fitting baselines outside of the pre-event, event, and snapback hours. The list below details all six specifications tested. No specification produced statistically significant results.

- **All Weather Matches**: Includes the top five weather matches for each event for the control group only, no quasi-controls.
- **All Weather Matches, All Quasi-Controls**: Includes the top five weather matches for each event, plus all quasi-controls.
- **All Weather Matches, Top Quasi-Controls**: Includes the top five weather matches for each event, excluding the top 3% of the treatment group and bottom 3% of the control group.
- **Top Weather Matches**: Includes the single best weather match for each event for the control group only, no quasi-controls. *This method is included in the "Within-Subject" section of the report.*
- **Top Weather Matches, All Quasi-Controls**: Includes the single best weather match for each event, plus all quasi-controls.

- **Top Weather Matches, Top Quasi-Controls**: Includes the single best weather match for each event, excluding the top 3% of the treatment group and bottom 3% of the control group. *This method is included in the "Quasi-Controls" section of the report.*

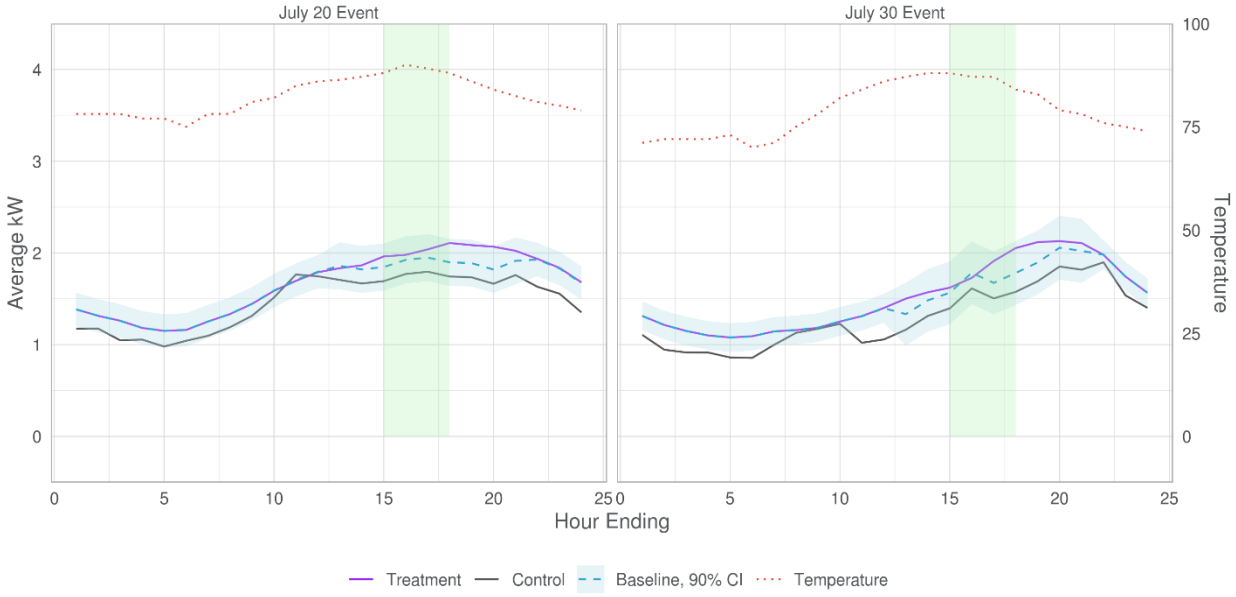
The figures in this section show the average usage for the treatment and quasi-control groups during the 2019 events for Level 1 Passive customers for the model specifications not already included in the body of the report. The baseline is shown in blue with a shaded 90% confidence interval.

Figure 9. All Weather Matches Predicted Baseline for 2019 Events*



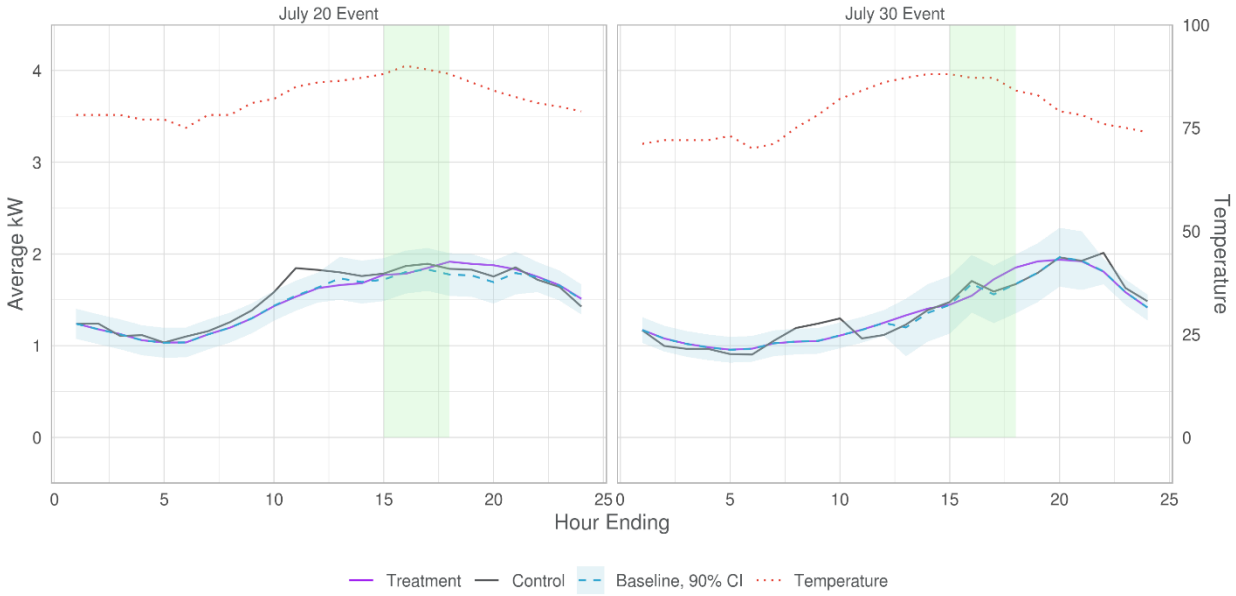
* Level 1 Passive customers, with the top five matched non-event days for each event.
Source: Guidehouse analysis

Figure 10. All Weather Matches, All Quasi-Controls Predicted Baseline for 2019 Events*



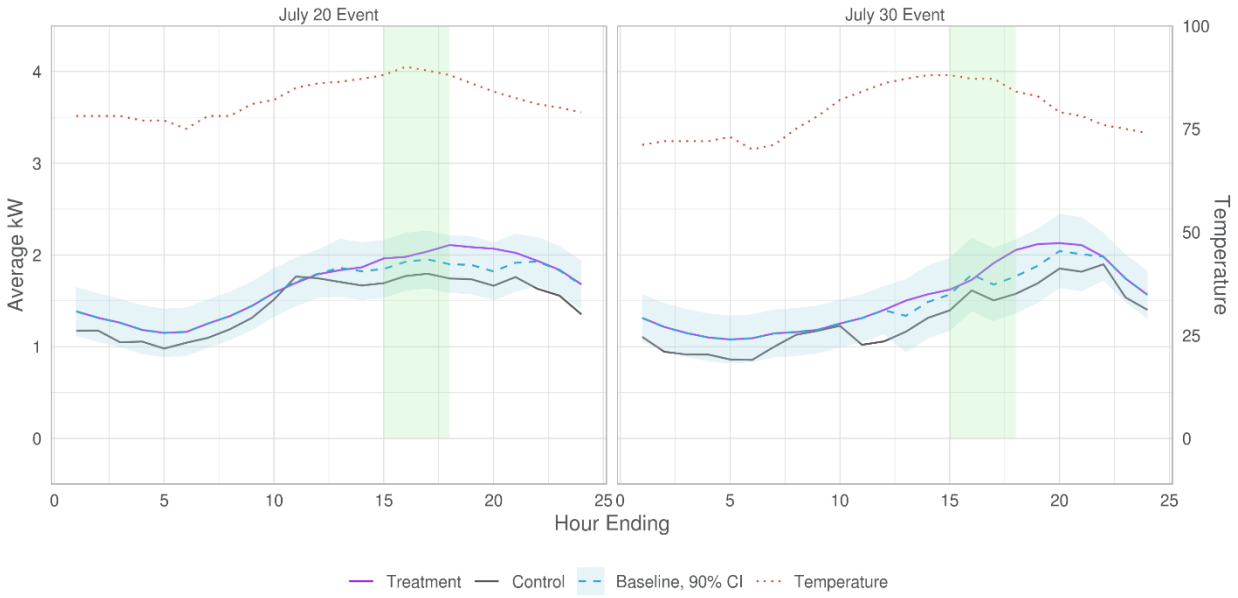
* Level 1 Passive customers, with the top five matched non-event days for each event. Includes all of the treatment and control groups.
Source: Guidehouse analysis

Figure 11. All Weather Matches, Top Quasi-Controls Predicted Baseline for 2019 Events*



* Level 1 Passive customers, with the top five matched non-event days for each event. Excludes the top 3% of treatment group and bottom 3% of control group.
Source: Guidehouse analysis

Figure 12. Top Weather Matches, All Quasi-Controls Predicted Baseline for 2019 Events*



* Level 1 Passive customers, with the single best matched non-event day for each event. Includes all of the treatment and control groups.
Source: Guidehouse analysis