Take the Load Off: Time and Technology as Determinants of Electricity Demand Response*

Megan Bailey† David P. Brown‡ Blake Shaffer§
Frank A. Wolak¶

May 2025

Abstract

As electricity systems transition toward more variable renewable energy, flexible demand has emerged as a critical tool for grid management. Yet a fundamental question remains: are emerging smart technologies sufficient to unlock demand response, or does human behavior remain the critical barrier? Our field experiment examines this question through a novel approach that individually randomizes peak event timing for each participating household, allowing us to leverage both within-subject and between-subject variation. We compare the response to "peak events" on electricity consumption for households equipped with three distinct demand response programs: a fully automated system requiring no action; app-enabled smart devices requiring minimal effort; and traditional manual adjustments. The results are striking—households with passive, automated responses reduced consumption five times more than those required to take any action at all, even when the burden is greatly reduced via smart technology. The provision of enabling technologies alone made no difference in households' responsiveness, as compared to a fully manual setting, when active participation was still required. These findings reveal that the opportunity cost of time and effort—not technology limitations—may be the fundamental obstacle to unlocking electricity demand flexibility. To achieve its full potential, "smart home" technologies need to incorporate these behavioral realities as barriers to responsiveness.

*We thank Meredith Fowlie, Katrina Jessoe, Robert Metcalfe, and attendees at the 2022 Electricity Camp in the Rockies, 2022 Berkeley Energy Camp, 2022 Canadian Resource and Environmental Economics Annual conference, 2023 ASSA sessions, the University of Alberta, 2023 Berkeley POWER conference, 2023 NBER Spring Environment and Energy Economics Meeting, and 2023 AERE Summer Conference for useful feedback. We are grateful to our partner electric utility company for sponsoring and managing the field experiment on which this study is based. This work is supported by excellent research assistance from Mallika Sharma. We acknowledge funding from our utility partner, the University of Calgary, the Canada First Research Excellence Fund as part of the University of Alberta's Future Energy Systems research initiative, the University of Calgary's Global Research Initiative, and the Social Sciences and Humanities Research Council Insight Development Grant. This experiment is registered in the AEA RCT Registry under ID #AEARCTR-0008680 and the UCalgary Conjoint Faculties Ethics Board (REB21-0161).

†Haskayne School of Business and the School of Public Policy, University of Calgary megan.bailey@ucalgary.ca

‡Department of Economics, University of Alberta dpbrown@ualberta.ca

§Department of Economics, University of Calgary blake.shaffer@ucalgary.ca

¶Program and Energy and Sustainable Development and Department of Economics, Stanford University and NBER wolak@zia.stanford.edu

1 Introduction

"While the electricity system of the past involved forecasting demand and dispatching supply, going forward grid operators will increasingly find themselves forecasting supply and dispatching demand."

–Jeff Dagle, Pacific Northwest National Laboratory

The energy transition is rapidly changing electricity systems around the world as supply decarbonizes and new sources of demand are electrified (e.g. heating, transportation). However, the growth of variable renewable generation sources, such as wind and solar, has raised reliability challenges for power grids, where supply and demand must balance at every instant. These market dynamics highlight the potential for flexible demand to play a critical role in facilitating the transition to low-carbon electricity supply by managing an increasingly variable grid. Despite its promise, questions remain as to how to elicit such flexibility from notoriously inelastic electricity consumers.

One solution, long advocated by economists, is to send dynamic price signals to reflect the time-varying nature of marginal electricity system costs (Boiteux, 1960; Kahn, 1970; Joskow and Wolfram, 2012). However, inattention and other barriers may make such a first-best solution unrealistic (Schneider and Sunstein, 2017). In addition to the onus of having to gather (and understand) their electricity price and usage, consumers face an even more fundamental obstacle: The reward for adjusting consumption may simply not be worth their opportunity cost of time (Becker, 1965).

Many utility demand response programs amount to consumers "picking up pennies" in a series of irregular and relatively low stakes opportunities where consumers earn rewards, or savings, for reducing consumption during "peak events" (Harding and Sexton, 2017). Consumers can respond to these events, but to do so requires allocating time to actively participate and exert effort. Consumers may be understandably reluctant to allocate their scarce time when the individual private gains are small—despite potentially large social gains in the aggregate. This can lead to low price responsiveness and is reflected in inertia in consumer decisions.[1]

In this paper, we examine the role of time allocation and associated effort as barriers to flexible residential electricity demand. In particular, we investigate the role that

[1]Other examples where the opportunity cost of time or "hassle costs" act as a barrier to responsiveness include monthly subscription renewals (Einav et al., Forthcoming) and health insurance enrollment (Shepard and Wagner, 2025).

emerging technologies and active versus passive response requirements can play in reducing these barriers. Partnering with a large electric utility in Canada, we recruited approximately 1,000 households to a demand response experiment that lasted 17 months. Participants received individually randomized notifications of "peak events", roughly 3 to 4 per month, whereby they received financial rewards for reducing electricity consumption during the 3-hour window of each event.

Participants were assigned to one of the three demand response programs that differed in the provision of enabling technology and whether active versus passive responses to events were required. Households in the most basic **Manual program** received information on their real-time consumption as well as notifications of peak events via the Utility's App and had to *actively* respond by manually adjusting their consumption, as the name suggests. Households in the **Tech program** received the same information but also had app-enabled load controllers installed on their baseboard thermostats, hot water heaters, and electric vehicle (EV) chargers, allowing them to *actively* respond to events remotely via pushing a button in an App. Finally, households in the **Central program** had the same information provision and load control technology installed, but with the key difference that the Utility automatically reduced their consumption in response to an event, i.e. their response was *passive*. Households could override the automatic reductions, but, in contrast to the Tech program, they needed to actively make an effort to *not* reduce consumption by pushing a button on their App.

The Tech and Manual programs constitute what we term *decentralized* demand response in that households in these programs need to actively respond to a peak event notification by taking an action. In the case of the Tech program, the household could simply adjust connected devices via the push of a button in an app, easing the burden of effort but still requiring an active action. Households in the Central program constitute our *centralized* demand response, in that that their response was utility-controlled.[2] Even if they exert no effort, or pay no attention to the event, households in this program will have their consumption reduced and can earn rewards.

Our treatment of interest is the effect of peak events and, importantly, how that differs across participants in different programs. Accordingly, we randomize our treatment events, i.e. peak event notifications, at the level of household-day. That is, each participant receives a unique randomized schedule of peak events over the 17-month

[2] There are a number of recently developed programs that include utility-managed load-control for various appliances including hot water heaters (Wattersaver, 2023), thermostats (PG&E, 2023), electric vehicles (DTE Energy, 2022), and solar-plus-storage systems (Spector, 2020).

period of the study.[3] This granularity of randomization differs from typical "between-subject" designs, where randomization occurs at the level of group allocation, with participants subsequently receiving common treatments. Our design incorporates what is often called a "panel experiment" or "within-subject" design (Charness et al., 2012; Bojinov et al., 2021; List, 2025). Within-subject designs are far less common than between-subject designs, in part because of the complexity involved in implementing an experiment with personalized randomization.

There are several benefits of leveraging within-subject identification. First, we are able to directly estimate the average treatment effect from peak events, separately by program, for the population of households that accepted the demand response program offers—a group that is of interest to utility planners. Since our interest is in the effect of randomized peak events, we can directly estimate this effect due to 100% compliance of receipt of treatment events, i.e. peak event notifications, by participants in the study. Second, the variation provided by the individualized randomization schedules increases the statistical power of our estimates. Third, the within-subject design allows us to estimate household-specific treatment effects, providing estimates for the full distribution of responses to peak events for each demand response program. We analyze these household-level estimates to better understand the mechanisms driving our results.

Our results provide three key insights. First, we find that people do respond to financial incentives, but time allocation/effort is a major barrier to responsiveness. Using both within- and between-subject variation, we find participants in the Central program reduced consumption, on average, by 26.3% during events, as compared to 4.8% and 5.3% in the decentralized Tech and Manual programs, respectively. Having to take an action to reduce consumption, even one as small as pushing a button on an app to respond to an event, results in one-fifth of the effect. Second, technology alone is not sufficient. Simply providing consumers with smart home technology to monitor and control their energy use is insufficient to drive significant behavioral change. The Tech program, despite offering remote control capabilities, performed similarly to the Manual program, highlighting that technology adoption alone does not guarantee success. Third, the magnitude of the Central program's response reinforces the considerable potential of demand flexibility as a resource for electricity grids. The results make a strong case for the value of centralized/automated demand response.

[3] The schedule is blind to the participant prior to the notification that occurs 21 and 2 hours preceding an event.

We find the distributions of household treatment effects, estimated solely based on within-subject variation, differ significantly across the programs. Participants in the Central program are normally distributed around a mean reduction of 24%. Whereas, those in the Tech and Manual programs have mean average reductions of approximately 5%, consisting of a large mass around 0% and a small subset of high performers. Our within-subject design also allows us to examine the consistency of individual household treatment effects by exploring the variance of responsiveness across events within each household. We find that the standard errors are similar across the three programs, with slightly smaller average values for the Central program. This suggests that households' responses to events are broadly consistent. That is, the Central program households are consistently large responders, while the Tech and Manual households are in large part consistently low responders. This finding has important policy implications as the Central program can be viewed as a reliable source of demand-side flexibility.

We explore potential mechanisms driving our results by combining the household estimated treatment effects with detailed data on participants' interactions with the utility's App. We provide evidence that differences in performance are driven by differences in time allocation and attention. We find that when households do not interact with their App on peak event days, the average household-level treatment effect is roughly 3% for the Tech and Manual programs and 24% for the Central program. When households do interact with their App, the reductions increase to 8.5% for the Tech and Manual programs, and 27% for the Central program. These results emphasize that while all programs increased responsiveness when households interacted with the App, the Central program's "headstart" from removing its requirement to have to take an action to respond is associated with the largest effects.

The App interaction data provides further insights into the intensity to which households allocated time to monitor and provide demand flexibility and how this correlates to their performance in responding to peak events. We find that the small subsample of "high achievers" in the Tech and Manual programs interacted with their App very frequently (61% and 51% of days on average), which suggests that their performance relied on considerable time devoted to their electricity consumption decisions and effort in reducing it during events. In contrast, the high achievers in the Central program interacted with the App significantly less (31% of days on average).

We also investigate how price affects the degree of responsiveness by randomizing peak event incentives into regular and "high" reward events, where the rewards for consumption reductions are roughly double in the latter. We find no significant evidence that higher prices motivate greater consumption reductions.[4] This further suggests a story of responsiveness where time allocation/effort acts as the key barrier, rather than a smooth response to price signals.

Finally, we use data from an exit survey to understand the relationship between our estimated household-level treatment effects and measures of an individual's opportunity cost of time. This analysis provides further support for our key identified barrier to demand response. Households with lower income and/or a higher perceived personal net benefit for providing demand flexibility were associated with larger responses to peak events. The Central demand response program is able to overcome these time allocation-related barriers through the use of enabling technologies and utility-controlled/automated responses to events.

Our paper builds on several strands of the literature. First, we add to the rich set of empirical research estimating household responsiveness to time-varying pricing in electricity.[5] Our experiment is most similar to the critical peak pricing (CPP) strand of this literature. The results from our Manual program with no load control/automation technology are broadly in line with those observed in prior studies.

Second, our paper contributes to a growing literature that explores automation options for consumers to overcome barriers to demand response. There is evidence that automation of smart thermostats and electric vehicles (EVs) can assist in facilitating short-run demand responsiveness when combined with pricing (Harding and Lamarche, 2016; Bollinger and Hartmann, 2020; Burkhardt et al., 2023; Blonz et al., 2025; Bailey et al., 2025). Our work builds on this work by providing a detailed decomposition of the role of prices, technology, and time/effort barriers, leveraging automation in a broader range of smart home technologies. In particular, we are able to uniquely look at the relative impact of automation compared to a program that has the exact same technology provided without automation; previous estimates may conflate the impact of automation with automation-providing technology. Recent work finds that consumers may override important settings with such technology, reducing the anticipated benefits (Brandon et al., 2022). Consistent with the latter, we find our Tech program performs no better than the Manual program on average.

[4] This finding is similar to that of Prest (2020) who estimates the price-responsiveness of households facing time-of-use prices in Ireland (without automated technologies) and find that households are responsive to TOU price signals, but the magnitude of the price does not matter.

[5] See Faruqui and Sergici (2010), Harding and Sexton (2017), and Yan et al. (2018) for surveys of this literature.

That is, given the ability to remotely control large appliances as well as automate some aspects of their electricity usage (e.g. thermostat settings), they fare no better than consumers who require a more manual action. The key takeaway here is that in order to achieve its full potential, technology needs to incorporate the behavioral realities of the opportunity costs of time and effort as a barrier to responsiveness.

Third, our paper relates to existing work on default effects that consider a range of settings, including retirement savings (Chetty et al., 2014; Bernheim et al., 2015), organ donations (Abadie and Gay, 2006), and retail electricity plans (Fowlie et al., 2021). Broadly speaking, this literature shows that default effects can significantly impact outcomes. The work most closely related to ours is Fowlie et al. (2021) who look at opt-in vs opt-out default effects at the extensive margin of selecting time-varying retail electricity pricing plans. Our paper complements this work by focusing on default effects at the intensive margin reflecting the decision to alter consumption decisions in response to financial incentives. Our key contribution beyond the existing literature is the finding of significantly greater responsiveness when consumption reductions are made the default, or passive, action in response to demand response events. Requiring customers to take action—even with the provision of technology that makes the associated cost as minimal as remote control with a mobile phone app—is no match for the power of demand response that is managed on the consumer's behalf. This speaks to the importance of recognizing the opportunity cost of time/cost of effort in settings where inattention is high and individual event rewards are relatively small, despite potentially large social gains in aggregate.

Our analysis proceeds as follows. In Section 2, we begin by presenting a conceptual framework for how time/effort cost can inhibit responsiveness based on Becker (1965). Section 3 describes our experimental design and data. We start the analysis with a descriptive analysis in Section 4. This is followed by our formal estimation framework in Section 5 and estimation results in Section 6. Section 7 presents empirical evidence to understand the mechanisms behind our results. Section 8 concludes.

2 Conceptual Framework

This section describes a model to serve as a conceptual framework for our field experiment. The model is motivated by Becker's (1965) model of household production, where a household allocates time and/or purchases market goods to produce goods and services according to a household production function. In our setting, this includes allocating time to provide electricity demand flexibility. In our field experiment, we provided households with the production technology they can use to provide demand flexibility. This model illustrates how the properties of the production technology can impact the household's provision of demand flexibility.

The household consumes three goods and services (Z_1, Z_2, Z_3) and earns utility $U(Z_1, Z_2, Z_3)$.[6] Z_1 is a composite good representing all non-energy household goods and services, Z_2 is energy services, and Z_3 is energy flexibility. The utility from energy services arises from the use of electricity for heating and cooling, powering lights, etc. Energy flexibility provides utility to the household through the "warm glow" of providing flexible electricity demand and other non-pecuniary benefits. By modeling these as two separate services, energy flexibility aims to capture the household's ability to shift consumption inter-temporally in the short run, rather than an aggregate change in the total electricity used. The household receives a financial payment for providing energy flexibility, denoted by $\rho \geq 0$.

The household's decision problem is represented by the following:

$$\begin{aligned}
\max_{\mathbf{Z},\mathbf{t},\mathbf{X}} \quad & U(Z_1, Z_2, Z_3) \\
\text{subject to} \quad & Z_i = f_i(X_i, t_i | R) \quad \text{for} \quad i = 1, 2; && (1)\\
& Z_3 = f_3(t_3 | R); && (2)\\
& \sum_{i=1}^{3} t_i + t_w = \overline{T}; && (3)\\
& p_1 X_1 + p_2 X_2 = t_w \omega + \rho Z_3. && (4)
\end{aligned}$$

For goods $i = 1, 2$, the production function for good Z_i depends on the input market good X_i, which comes at a price p_i, and time used in household production, t_i. Market goods do not yield utility directly, but rather serve as an input to the final household goods. For energy services (Z_2), the market good X_2 can reflect the input electricity, for example. For demand flexibility (Z_3), the production technology depends on the amount of time t_3 the household allocates to providing this service. This could represent, for example, the amount of time spent turning off devices during an event and/or learning how to use technology to provide flexibility. We assume that there is no choice of the market input for demand flexibility (i.e., there is no choice of X_3). We are taking the set of technologies used to facilitate demand flexibility as

[6] We assume that the utility function is continuous, differentiable, and strictly quasi-concave. Appendix A provides derivations and illustrates the model results using Cobb-Douglas utility.

fixed. The household is randomly assigned to the Manual, Tech, or Central demand response programs and provided different energy flexibility enabling technologies. We parameterize these technologies by R in their production functions.[7]

We follow Becker (1965) and assume the following production functions:[8]

$$X_i = a_i(R)\, Z_i \quad \text{for} \quad i = 1, 2 \quad \text{and} \quad t_i = b_i(R)\, Z_i \quad \text{for} \quad i = 1, 2, 3. \tag{5}$$

This production technology implies that for each unit of Z_i, it takes $a_i(R)$ market goods and $b_i(R)$ units of time for $i = 1, 2$. For energy flexibility, an increase in $b_3(R)$ means the household needs to allocate more time to produce the same level of energy flexibility. In the context of our experiment, all households in demand response programs are allocated an App to monitor consumption. A subset of these households receive technology that can either be actively controlled (i.e., Tech) or passively controlled (i.e., Central) by the household to respond to events. This illustrative model considers a single representative household, but with (randomly) provided household energy technologies that impact the household's choice of demand flexibility via its impact on $b_3(R)$.

The household has a finite amount of time it can allocate, $\overline{T}$. In addition to allocating time to producing household goods and services, the household can spend time t_w working and earning a wage of ω. The time constraint is represented by equation (3). Finally, the household faces the budget constraint in equation (4) that depends on the amount spent on market goods, the amount earned from working, and compensation for providing demand flexibility. Facing these two constraints, the household has a finite amount of time and financial resources it can allocate to producing goods and services, and it faces an opportunity cost of time as it foregoes the opportunity to earn income.

In this framework, the household's decision problem can be rewritten as a standard

[7] In our experiment, households could install flexible energy technologies (e.g., smart thermostats) themselves. However, we observe minimal adoption of such technologies that can be linked to the Utility's App to observe and adjust device-specific consumption. The model could be expanded to include the choice of X_3 on top of the randomly allocated technology set parameterized by R. The key qualitative conclusions, summarized below, will persist with additional notational complexity.

[8] The results of this model hold under more general production technology assumptions (e.g., allowing time and market goods to be substitutes, multiple market goods for each final household product). Pollak and Wachter (1975) show that if the production technology has constant returns to scale and there is no joint production, then the cost of providing a household good does not depend on the level of the household good or service. This property is leveraged below to collapse the model down to a standard utility-maximization problem.

utility maximization problem:

$$\begin{aligned}
\max_{\mathbf{Z}} \quad & U(Z_1, Z_2, Z_3) \\
\text{subject to} \quad & \sum_{i=1}^{3} \pi_i(p, \omega, \rho|R)\, Z_i \; = \overline{T}\,\omega \quad \text{for} \quad i = 1, 2, 3;
\end{aligned} \tag{6}$$

where π_i represents the traditional role of price for household good Z_i and is defined as follows:

$$\begin{aligned}
\pi_i(\,\cdot\,) \;&=\; p_i\, a_i(R) + \omega\, b_i(R) \quad \text{for} \quad i = 1, 2, \quad \text{and} \\
\pi_3(\,\cdot\,) \;&=\; \omega\, b_3(R) - \rho.
\end{aligned} \tag{7}$$

The solution is achieved by equating the marginal rate of substitution with the price ratio of any two household goods. Consequently, a decrease in the price of Z_3 (π_3) increases the provision of demand flexibility. The household only has a finite amount of time. A reduction in $b_3(R)$, through an improved demand-flexibility inducing production technology, effectively reduces the price of providing demand flexibility. Alternatively, if the wage rate ω increases, the household's opportunity cost of time increases, raising the price of providing demand flexibility.

In the context of our experiment, this illustrative modeling framework emphasizes that the flexibility-enabling technology has the potential to impact the provision of demand flexibility by reducing the cost of providing this service, given the presence of the household's time and budget constraints. The model does not take a stand on the precise impact of each technology in our experiment as this is an empirical question that will be explored and discussed in detail below. Rather, it illustrates the mechanism through which technology has the potential to increase demand flexibility through the household's rational (utility-maximizing) decisions.

3 Experimental Design and Data

3.1 Overview

We partnered with a large Canadian electric utility (hereafter referred to as the "Utility") to create three demand response programs that vary in terms of the enabling technology provided to customers, as well as household versus utility-initiated electricity demand changes on specific devices. We are primarily interested in how customers in each program adjust their electricity consumption in response to "peak events", times during which the Utility asks consumers to reduce consumption and rewards them financially for doing so.

A unique feature of our study is the fact that we randomized the timing of peak events at the household level. We leverage the randomization of these events (as well as the richness of our data) to estimate household-specific treatment effects. These features allow us to look at the distribution of effects within each program to better understand what is driving average program-specific results.

3.2 Treatment Events

Customers in each of the demand response programs received notifications of peak events through an electricity consumption management phone App offered by the Utility. Peak events had the possibility of occurring at one of two time periods: morning (7am to 10am) or evening (5pm to 8pm). The schedule and timing (morning or evening) of events were unknown ex-ante to the customer, each receiving a unique, randomized schedule of events over the course of the experiment. Consequently, households could not predict the day or event time when they would receive a peak event. Households received event notifications 21 and 2 hours before the event that included an offer for households to receive financial rewards for reducing electricity consumption during the peak event period, relative to their household-specific baseline.[9] Further, because events were randomized in time for each household, they are not correlated with other drivers of household electricity consumption.

"Event types" were also randomized and were one of two pricing levels, "normal" and "high", with rewards increasing in the latter for large reductions. High peak events were only possible during evening periods. During normal events, households could receive \$1 for a 10% reduction, \$2 for a 30% reduction, or \$3 for a 50% reduction. During high peak events, households could receive \$1 for 10%, \$3 for 30%, or \$6 for 50% reductions.[10] By randomizing the pricing levels, we are able to estimate the effect of greater price incentives on household consumption behavior. The incentive amounts translate to payments ranging from approximately \$1.11 to \$2.22 per kWh

[9] Baselines were calculated based on a household's average consumption during the relevant event time window over the last five weekdays prior to the event, excluding days where events occurred. Customers did not know how the baseline was calculated to avoid the potential for customers to "game" their baseline 21 hours was selected as the longest notification period to avoid consumers having knowledge of a pending event while their baseline was still being set.

[10] Report dollars are in Canadian dollars. CAD\$1 equals approximately USD\$0.75 or EUR€0.68 as of December 2023.

of electricity reduced, for the average household.[11] These incentives are in the range of wholesale price caps that are used to limit electricity scarcity pricing in a number of jurisdictions in North America.[12]

Events randomly occurred on weekdays, excluding holidays. Households typically experienced two "normal" and one "high" event per month. This schedule was altered in the summer months of July and August when the likelihood of peak events is lower in Canada. During these months, households experienced no "high peak" events. Events started on February 22, 2022, and continued until June 30, 2023 resulting in 30,502 household-event days in our experiment.

Event notifications provided information on the time of the event and the financial incentives for the different demand reduction levels. Once consumers received the 21-hour notifications, they could also see event details in the App itself. See Appendix C.1 for examples of the notification and in-App event messages. Households' rewards for consumption reductions during events were displayed in the App at a two- to three-day lag. The App also gave households a summary of their total rewards to date. See Appendix C.2 for details on each program's in-App experience.

3.3 Demand Response Programs

Our initial sample consists of all households that first downloaded a utility-facilitated App. The App provides individuals with household-level hourly consumption posted at a one-day lag. From the pool of approximately 9,000 households that joined the App, we identified 1,661 eligible households based on factors such as being located near a major metropolitan city and household characteristics.[13] Eligible households were randomized into one of three demand response programs or two never-treated groups. Email invitations were sent to eligible households randomized into the demand response programs, with several reminders. See Appendix B.1 for a complete

[11] The average household consumes 1.8 kWh in each hour in our sample. A 10%, 30%, and 50% reduction translates to a 0.54, 1.62, and 2.7 kWh reduction over the three-hour event, respectively. Consequently, for a 50% reduction during a normal peak event, we compensated households $\frac{\$3}{2.7} = \$1.11$ per kWh. For a 50% reduction during a high peak event, compensation was $\frac{\$6}{2.7} = \$2.22$ per kWh. The other percent reductions lie between these two cases.

[12] Examples include the wholesale price cap of CAD\$1.00/kWh in Alberta (Brown and Olmstead, 2017), USD\$3.50/kWh in the Mid Continent Independent System Operator that operates in the Midwest United States (IRC, 2017), and USD\$5/kWh in Texas (Smith, 2022).

[13] Eligible households were those that were in or near a major metropolitan city for which it was feasible for utility-partnered electricians to install load control equipment; homeowners; not in condos or apartments; had at least one month of historical electricity consumption data as of September 2021; and had at least one controllable electrical device (level 2 electric vehicle charger, electric baseboard heater as the primary heat source, and/or an electric hot water tank).

description of the recruitment and assignment process.

Table 1. Summary of Household Programs

Programs	DR Control	Load Controller	Price Incentive	Real-Time Usage Info
Central	Utility	✓	✓	✓
Tech	Household	✓	✓	✓
Manual	Household		✓	✓
Info	Household			✓
Control	Household			

Notes. DR Control represents whether demand response to events is controlled entirely by the household (decentralized) or by the Utility for the load-controlled devices (centralized). Load Controller denotes whether the household has load controller equipment installed. Price Incentive reflects if households receive peak events and rewards for reduced demand during events. Real-Time Usage Info denotes whether households receive real-time household-level consumption information. ✓indicates categories that are applicable to each program.

Table 1 summarizes our demand response programs and never-treated groups. Households in the **Manual** program earned financial rewards for demand reductions during events but did not have any load controller equipment installed by the Utility to manage their consumption via the App. They had to respond to events manually.[14] Households were sent a device that allowed them to monitor their real-time electricity usage information in the Utility's App.

The **Tech** program differs from the Manual program in that the Utility installed load controller equipment on one or more of the household's electric hot water heaters, baseboard thermostats, and level 2 EV chargers, to enable remote electricity consumption ("load") reductions. This equipment allows households to see device-specific electricity consumption and turn on and off devices remotely via the App. Critically, while the Tech program is equipped with load control technology to ease their effort in responding to events, they still must take active action to do so via their phone's App. See Appendix C.2 for detail on the experience that participants had with the App, by program. Both the Manual and Tech programs allow us to test the efficacy of a *decentralized* approach to demand response—one without and one with enabling smart technology, respectively.

[14]We can observe if a household independently installs its own equipment and links it to the Utility's App. Only 3 (out of 242) Manual households installed equipment independently. These 3 all installed smart thermostats that allow the monitoring and remote control of the household's electric baseboard heaters.

The **Central** program received the same equipment installed in their homes as the Tech program. However, during an event, the passive setting for Central program participants was for the Utility to manage their load-controlled devices by reducing electricity consumption. That is, without any active response, the Central program participants would reduce consumption via demand management initiated by the Utility. For example, during events, their thermostat temperature is automatically reduced, EV charging is delayed, and/or the water heater is turned off. Central program households needed to actively choose *not* to respond to an event (i.e., select out of utility management) by pushing a button on their App. The Central program allows us to examine the efficacy of centralized demand management and, as compared to the Tech program, the difference between having to take active action during an event versus passively responding by conceding control to a third party.[15] It is important to note that the Central, Manual, and Tech programs had symmetric information about their real-time usage which could be observed within the App.

Finally, we have two groups of households that serve as *never-treated* baselines throughout the study. One is a **Control** group that receives no intervention or messaging regarding the experiment after joining the utility's App. Another is an **Info** group that is identical to the Control group, but these households have access to real-time consumption information for their home via the App after they accepted and installed a device provided by the Utility. Both of these never-treated groups do not receive peak events or financial incentives. We passively monitor their consumption.

While households outside of the Control group needed to accept being in each program, the final sample of participating households are representative of those that utility companies are interested in until such time that demand response programs serve as the default.

3.4 Acceptance

Table 2 summarizes the number of households invited and the acceptance rates for each program offer. Acceptance rates among all programs were high. In particular, the acceptance rate for the Central offer was 42%, and only marginally statistically different than the acceptance rate for the Tech offer (48%). A difference in means test

[15] Customers in the Central program received advanced notifications of events just as other program participants, and the language of the notifications reminded them that the Utility would manage their devices. See Appendix C.1 for event notifications for each program. They had the option to opt-out of utility management for all devices ahead of an event or for individual devices during an event. These options are described in Appendix C.2.

between these two values yielded a p-value of 0.072. Compared to Tech and Central, the Manual program had a statistically significantly higher acceptance rate of 59%, followed by the Info-only group at 68%.[16] Finally, Control had 100% acceptance because their participation was not subject to an offer. The acceptance rates of the Central and Tech programs were lower than the others due in part to the need for load controllers to be successfully installed in households that accepted these offers.[17]

Table 2. Program Acceptance by Program

	Central	Tech	Manual	Info	Control
Invited	423	382	409	259	188
Accepted	177	184	242	177	188
Pct. Accepted	(42%)	(48%)	(59%)	(68%)	(100%)

Notes. "Invited" reflects the number of households invited to participate in the experiment, by program. "Accepted" is the number of households that accepted our offer and made it through the equipment installation process (as applicable, by program). "Pct. Accepted" displays acceptance rates relative to the number of households invited.

We take the similarity among final acceptance rates between the Central and Tech programs as the first set of evidence that we can compare our estimated treatment effects between these programs. While the Manual program had a higher final acceptance rate, concerns that the Manual program participants systematically differ from those of the other two programs are mitigated based on a comparison of observable characteristics across programs in our final sample, as are concerns about differences in the composition of households invited to each program. See Appendices B.2 and B.3 for a detailed discussion of balance using pre-treatment data.

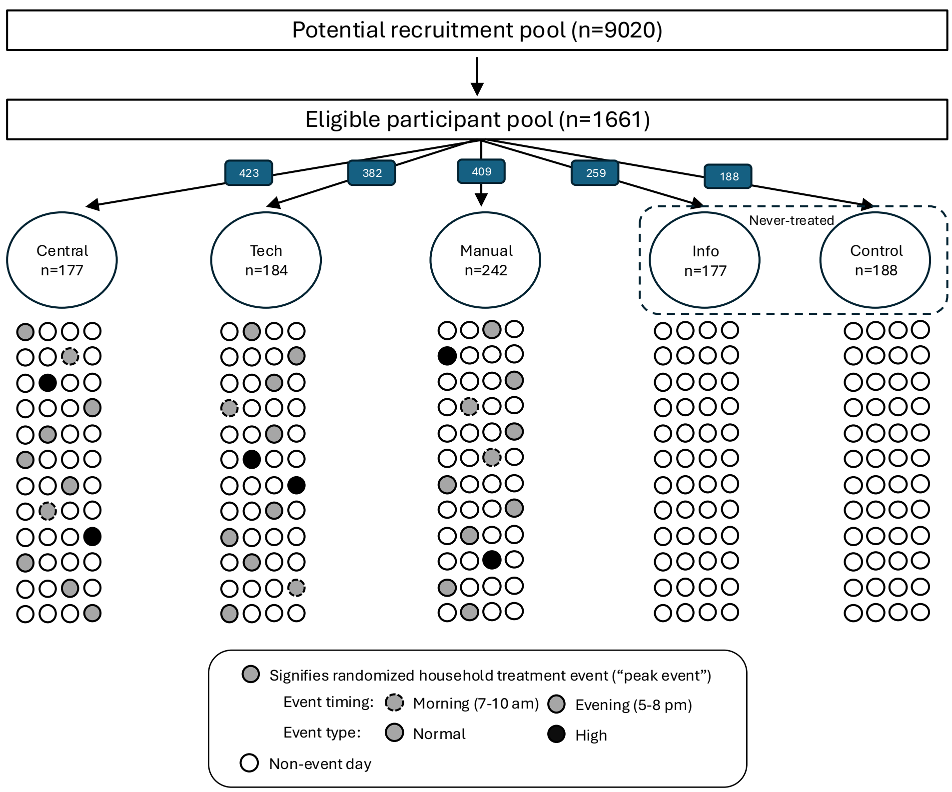
Figure 1 provides a summary of our experimental design, starting with our initial recruitment and eligible participant pools, and subsequent allocation and acceptance into our demand response programs and never-treated groups. The dots represent each day of our sample, with shaded dots capturing the randomly allocated peak events that can occur on different days across individuals. We leverage this within individual randomization in our identification strategy outlined in Section 5 to identify the average treatment effect of peak events on the population of households that

[16] Like the Manual program, the info group required actively accepting the offer to join the experiment and installing a device (called the "Hub") that facilitates the monitoring and reporting of real-time consumption.

[17] We observed unsuccessful installation at households that initially accepted these offers due to, for example, households never responding to subsequent inquiries to receive and install equipment or households not being in compliance with local electrical codes.

accepted participation in our demand response programs. We evaluate if responses vary by peak event type (i.e., morning, normal evening, high evening) by program.

Figure 1. Experimental Design Summary



Notes: Dots represent individual household-days. Treatment events are indicated by filled dots, which are randomized at the household level, i.e. running vertically in this diagram. Dashed contours on treatment events indicate morning events, solid contours indicate evening events. Grey fill indicates "normal" event types and darker fill indicates "high type" (i.e. larger incentive) events.

3.5 Data Description

For all households in our experiment, we track hourly household-level consumption (in kWh) from October 1, 2020 until June 30, 2023. We also have information on a number of household characteristics, such as household appliances, that were provided through survey responses as a necessary condition to enter the first phase of our recruitment process. In addition, the Utility provided supplementary household information, including the type of household (e.g., single-family/duplex, row home) and an approximate geographical location. We are also provided time-stamped information on household interactions with the Utility's App at the daily level.

We complement the detailed household-level data with demographic information from the 2016 Canadian Census (Statistics Canada, 2021). We are provided a household's Census Dissemination Area (CDA) identifier; the CDA is the most granular geographical unit for which all Census information is provided publicly. We collect hourly weather information to control for environmental factors that impact electricity consumption, including temperature and humidity at three stations that are geographically representative of the households located in our study.[18] These data were accessed at Environment and Climate Change Canada.

In the last month of the experiment (mid-June 2023), we conducted an additional survey that contained questions on participants' experience with their respective demand response program. We provide a subset of questions from the survey in Appendix E.1. We use these survey responses in our analysis in Section 7.

4 Descriptive Results

We begin our analysis with descriptive evidence that participating households reduce their electricity consumption during peak events and show how this response differs across demand response programs. Figure 2 provides average hourly household-level consumption for the Central, Tech, and Manual programs for the entire sample period during non-event (solid lines) and event (dashed lines) days. The shaded regions reflect the relevant event hours for each event type.

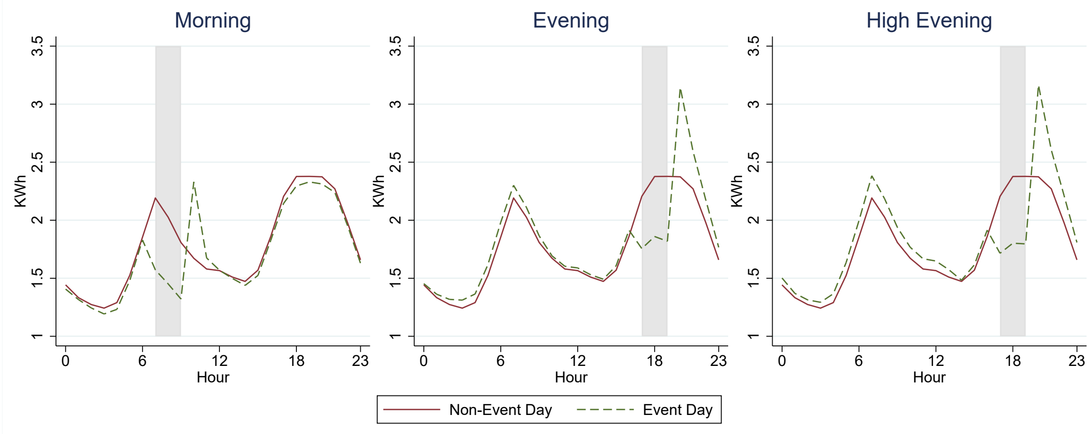
Figure 2a demonstrates that the Central program had a large reduction in average consumption during events regardless of the event type. After each event, we observe a large spike in consumption. This "snap-back" is consistent with the devices turning on immediately after the event (e.g., to reheat the water tank and/or home, or restart EV charging).[19] Comparing High Evening to Evening event consumption, we see no discernible difference in response to this higher reward.

Figures 2b and 2c demonstrate that the Tech and Manual programs show negligible

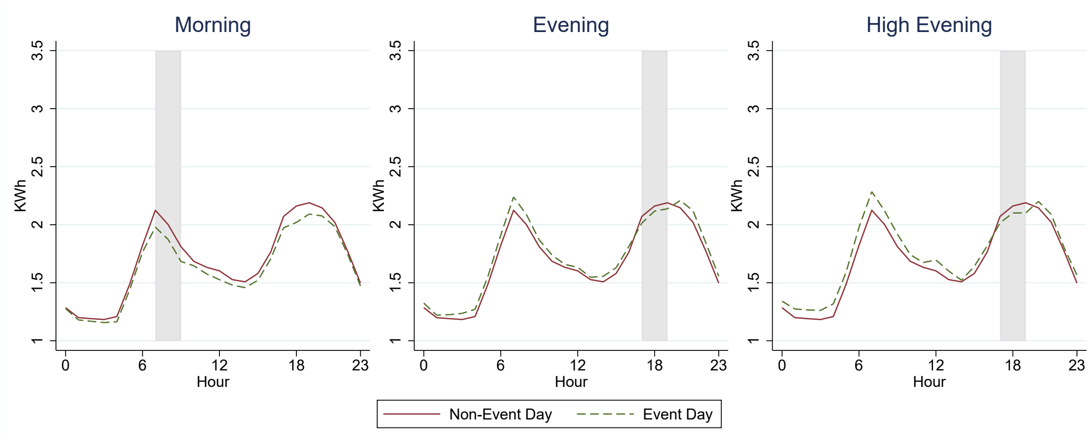
[18] We match the households in our sample with their closest weather station.

[19] Because we are interested in event-induced demand flexibility among programs, not energy efficiency, we are unconcerned with demand being shifted to another time. However, based on conversations with the Utility, the new problematic peaks in demand created by the observed snap-back could be mitigated by the Utility staggering the beginning and/or end of the load-controlled event across households or only partially adjusting the demand levels on controllable devices. Managing the "snap-back" or "shadow peak" from demand response is an important area for further research.

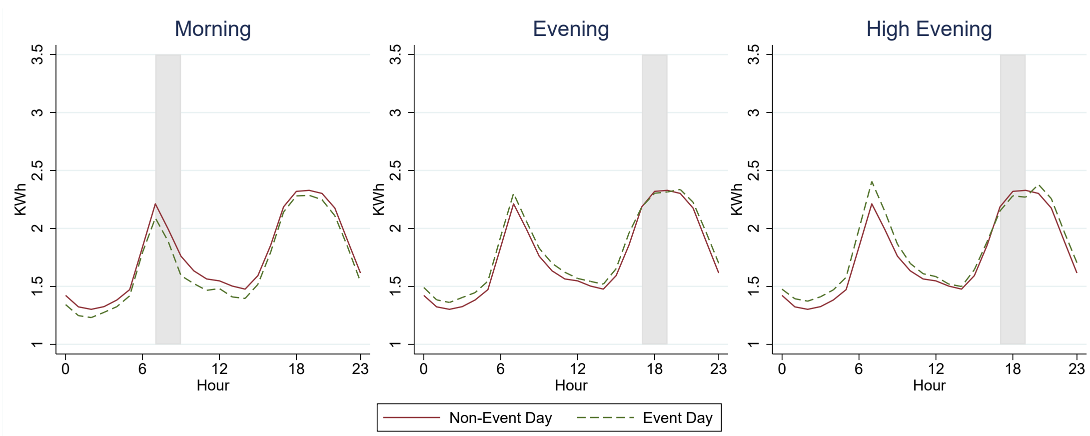
Figure 2. Average Household Consumption



(a) Central Program



(b) Tech Program



(c) Manual Program

Notes: Figure plots mean household consumption by hour and demand response program on weekdays, on event days and non-event days over the period February 1, 2022 - June 30, 2023. Event days are separated by type: Morning, Evening, and High-Evening. The shaded area represents the relevant 3-hour event period.

changes in consumption patterns during events. This limited observable response for the Tech program arises despite the fact that this program has access to the same equipment as the Central households. However, unlike the Central program where changes are automated, the Tech households must actively engage with the App or device to turn off the same appliances during an event.

Taken together, these descriptive results suggest that the Central program has a considerably larger response to each event type. Further, we see no visual evidence of greater performance for greater financial rewards. Rather, the largest difference appears to be whether the household is in a passive (centralized) versus active (decentralized) program. In the sections that follow, we undertake a formal empirical analysis to quantify these effects and control for potentially confounding factors.

5 Empirical Framework

5.1 Program-Level Regressions

We estimate the average treatment effect of peak events by comparing electricity consumption for households facing a randomized peak event to those who, contemporaneously, are not, using the following model at the household i and hour t level:

$$ln(c_{it}) = \sum_{j \in \{C,T,M\}} \beta_j \, \text{Program}_{ji} \cdot E_{it} + \alpha_i + \tau_t + \delta X_{it} + \varepsilon_{it} \qquad (8)$$

in which $ln(c_{it})$ is the log of household electricity consumption, E_{it} is the household-specific event indicator that equals one if the household is (randomly) assigned an event in hour t, and Program_{ji} is a categorical variable for which demand response program the household i is enrolled in (e.g., Central [C], Tech [T], or Manual [M]). A key advantage of this model is that it allows us to readily test for differences in responsiveness to events across the three demand response programs. We use the log of household electricity consumption on the left-hand side to account for the right-skewed nature of consumption.[20]

We include α_i, household fixed effects, which control for time-invariant household characteristics. We also include τ_t, an hour-of-sample fixed effect, which controls for time-varying factors that impact consumption. Household electricity consumption and consumer responses to events may vary with local weather conditions (especially

[20] Our results are robust to functional form; we observe similar results with a linear-linear specification.

due to thermostat settings). To control for this, we include X_t, a vector of hourly weather controls that include the relative humidity and cooling degrees and heating degrees above and below 65° F (18.33° C). Since these may vary in weather conditions in a nonlinear way, we include a flexible functional form with a polynomial up to the third degree for each weather-related covariate. ε_{it} is the error term. We cluster standard errors at the household level.

We also consider a version of this regression specification where the event indicator variable, E_{it}, is adjusted to be a categorical variable for the three potential event-types: Morning, Evening, and High-Evening. This analysis allows us to evaluate if households' responses to events differ by time-of-day and the financial reward for responding, in the case of Evening compared to High Evening events.

Our parameters of interest are β_j for $j \in \{C, T, M\}$, which measure the change in household-level electricity consumption during peak events for each of the Central, Tech, and Manual demand response programs. Because of our log-linear specification, we transform our estimates to report the percentage change in hourly consumption during an event via $100 \times (exp(\widehat{\beta}_j) - 1)$.

Our empirical framework relies on three identifying assumptions to recover the effect of events on household-level consumption. First, events are not correlated with other drivers of household electricity consumption. This is met via our randomization of events. We include weather controls to ensure that our estimated treatment effects can be interpreted as weather-agnostic.

Second, our analysis falls within the literature on experiments that leverage "within-subject" variation, which have been referred to as "panel experiment" designs (Charness et al., 2012; Bojinov et al., 2021; List, 2025). Panel experiments involve the treatment of interest (e.g., peak events) that vary in time. This differs from designs that use a one-time treatment and between-subject variation to identify treatment effects. A key identification assumption needed for estimating event-level treatment effects in the presence of within-subject variation is that our random treatment events do not "carryover" to a persistent change in behavior in similar hours on non-event days. This could occur if, for example, experiencing an event led to a household persistently scheduling consumption reductions during the event hours on all days going forward. Or, conversely, a savvy participant might suspect a financial benefit of "gaming" the baseline by purposely increasing consumption during event hours on non-event days.[21] We test the validity of the "no carryover" assumption using a DID

[21] We mitigated this latter effect by providing no information on how the baseline consumption

regression to evaluate whether households in each demand response program adjusted their consumption during the event windows on non-event days in the post-treatment period, relative to the never-treated households. Details of the empirical approach are provided in Appendix D.2.

Third, as noted above, for each program, Equation (8) compares event time consumption to non-event time consumption of households in the same demand response program, other demand response programs, and never-treated households. This relies on the assumption that the households in the never-treated groups and other demand response programs provide a valid counterfactual on non-event days. While observed characteristics are similar across the demand response programs (see Appendix B.3), one may be concerned that the programs are differentially selected, and this impacts our ability to compare event-time to non-event-time behavior across programs, even after including our various control variables.

To address this potential concern of between program comparison validity, we perform a robustness check whereby we vary the set of comparators. We start with the most restrictive set: limiting the analysis to be solely within each demand response program and excluding the never-treated households. This restrictive specification ensures we are only comparing behavior across participating households within the same program and guards against concerns of comparing consumption between participating and never-treated households. We run three separate regressions, one for each program, to separately estimate β_C, β_T, and β_M. The treatment effect is identified within each program by comparing consumption when households receive peak events versus not. Next, we augment the comparison group to include never-treated households, i.e. those in the Control and Info groups, as non-treated comparators, but again run three separate regressions to estimate treatment effects for each demand response program separately. Finally, we pool all participants and never-treated households to run a regression comparing those receiving a peak event to all households not receiving a peak event at the same time (i.e., Equation (8)).To the extent that our results are robust to these alternative specifications, this will alleviate concerns that potential differential selection into demand response programs impacts our key conclusions.

Our empirical framework provides estimated average treatment effects to peak events by program for households that accepted the demand response program offers. In our setting, the treatment of interest is randomized household-specific peak events that all households in demand response programs received. As a result, there was no

that was used to determine the household's rewards was calculated.

non-compliance to our treatment of interest. One might consider estimating an Intent-to-Treat (ITT) and a Local Average Treatment effect (LATE) measure for the impact of program participation on event-time consumption (i.e., including households that did not accept the invitation to participate). However, households that did not join a demand response program did not subsequently receive the (randomized) peak events. Because of this, non-participating households will have (noisy) null event responses. In this setting, an ITT approach yields an estimate for each demand response program that is approximately equal to our main estimated event treatment effects multiplied by the share of households that joined the program.[22]

5.2 Household-Level Regressions

A unique feature of our setting is our ability to estimate household-level treatment effects solely from *within-subject* variation by leveraging the randomized event timing. This allows us to examine heterogeneity and distribution of event responsiveness across households, and to explore factors associated with household-level responsiveness that speak to the role of attention and effort/time allocated to responding to events.

Separately for each household i in the demand response programs, we estimate a household-specific treatment effect using the following model:

$$ln(c_{it}) \;=\; \gamma_i + \beta_i\, E_{it} + T_t + \delta_i X_{it} + \eta_{it} \tag{9}$$

where, analogous to above, c_{it} is consumption, E_{it} equals 1 when household i has an event and zero otherwise, and X_t includes the same set of temperature controls as the specification in Equation (8). In this specification, T_t is a set of time fixed effects that includes day-of-week, hour-of-day, and year-month to capture time-varying factors that impact consumption.[23] η_{it} is the heteroskedastic-robust error term.

The regression in Equation (9) estimates a separate $\widehat{\beta}_i$ for each household in our

[22]We carried out an analysis to verify these statements by randomly assigning "synthetic events" to households that did not join our demand response programs. Further, because under plausible conditions, the LATE equals the ITT divided by the share of compliers, we find that a LATE-like estimate equals a noisy estimate of our main demand response program-specific treatment effects. Details are available upon request.

[23]We cannot include an hour-of-sample fixed effect in the household-level regression because it would absorb the variation we are using for identification in this specification. We include several calendar fixed effects to absorb seasonal, day-of-week, and hour-of-day factors that impact consumption. We estimated our program-level specification detailed in Equation (8) using this set of fixed effects. The results closely reflect the estimates reported below.

demand response programs. The identification strategy of this household-level regression compares consumption behavior during event hours to non-event hours within the same household, conditional on time-based fixed effects and weather variables. Similar to the approach in our program-level estimation, we also consider a specification that allows for differential responses by event type.

The ability to estimate household-level treatment effects provides us with the opportunity to understand the potential mechanisms driving our results. In particular, we have data on when a household has interacted with the App on a given day. App interactions are indicative of the time and attention that households expended to respond to events.[24] We leverage this to estimate separate event treatment effects, by household, for when households do and do not interact with the App.

We run a specification of Equation (9) that interacts E_{it} with an indicator variable App Interact_{it} that equals one if the household has interacted with the Utility's App on the relevant day and zero otherwise. A key benefit of this approach is that it allows us to quantify how a specific household's estimated treatment effect varies by whether or not they interacted with the App on an event day. This helps overcome the sample selection challenge that would arise by running an analogous regression using all households within a given demand response program. With such a regression, it would not be possible to disentangle whether the different treatment effects arise because the household interacted with the App on an event day or whether the households that interact with the App are unique in the way they respond to events.

6 Empirical Results

This section presents the results of our demand response program-level econometric analyses. In particular, we provide the average treatment effect of events for each program, across all events and then separated by event type.

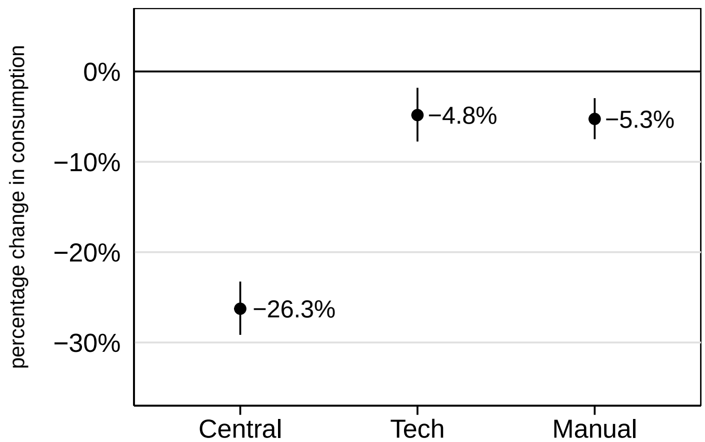
6.1 Program-Level Treatment Effects

Figure 3 provides the estimated average response to events by program as a percentage change in household-level consumption using the specification in Equation (8)

[24] Recall that households can use the App to monitor their hourly household consumption and observe the timing and rewards for upcoming events 21 hours in advance. Households with installed devices can also monitor their device-level consumption in the App. Tech households can adjust their connected devices by pushing a button in the App (e.g., to turn off their use during events). Finally, Central households can adjust their connected devices and opt out of centralized load management during an event.

including the full set of comparator households. We observe an average 26% reduction in consumption during events for the Central program. In contrast, the Tech and Manual programs reduced demand by approximately 5% on average during events. Both of these effects are significantly different from zero. Even though the Tech program had the same equipment as the Central program, it demonstrated a significantly lower response to events. Additionally, the average response for the Tech and Manual program are not significantly different.

Figure 3. Average Estimated Treatment Effects of Participants by Program



Notes: The reported results are program-specific marginal effects calculated from estimating $\widehat{\beta}_j$ in (8) for $j \in \{C, M, T\}$. We present the marginal effects to be a percentage change in consumption using the transformation $100 \times (exp(\widehat{\beta}_j) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

Recall that the Central program has the ability to opt-out of events using the App. We observe an opt-out rate of only 4% at the event-connected device level from the participants in the Central program. When taken together with the results for the Tech and Manual programs, this low opt-out rate suggests that the large reductions for the Central program are primarily attributable to consumers allowing utility management of their devices during events.[25]

These results align with the descriptive data presented in Section 4 that suggest that households in the Tech program did not use the load controller equipment to the same extent as the Central program. These results indicate that the installation of technology in the Tech program that enables remote control of household devices

[25]Notably, when consumers opted out of central management, they generally did so for their thermostats: 90% of device opt-outs occurred by households adjusting thermostats during events.

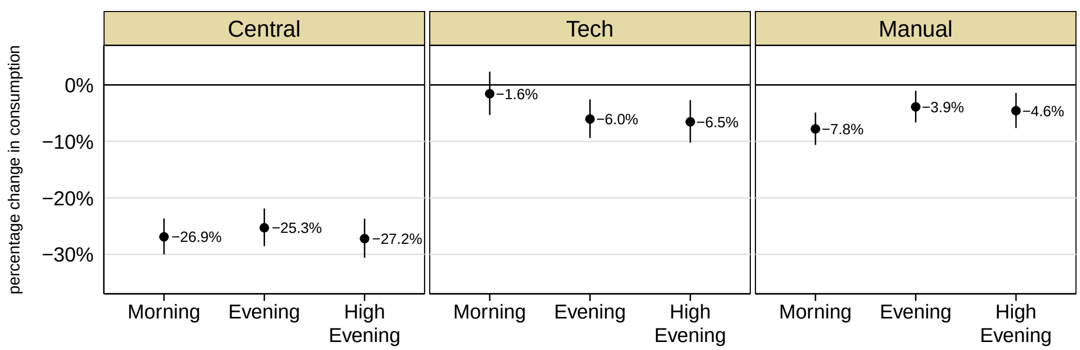
is insufficient to resolve the time allocation/effort barrier that limits the provision of demand flexibility (recall the model in Section 2). In contrast, centrally managed demand response that flips the active versus passive response requirement, thus requiring minimal-to-no effort, results in large demand reductions.

As described in Section 5.1, we undertake additional analyses to evaluate the validity of our identification strategy. Appendix Table C1 provides the results of our regression analysis when we vary the comparison group during non-event hours to estimate the event treatment effects. These results demonstrate that our estimated treatment effects are highly robust to varying the comparison groups included in the regressions. Appendix Table C3 provides the results for our test of the validity of our no carryover assumption. In this analysis, we find no evidence of changes in behaviour during the event window on non-event days for any of our demand response programs.

6.2 Program-Level Treatment Effects by Event Type

In addition to randomized event timing, we also randomize event types, varying both the time of the event and the financial reward for reductions. This allows us to estimate how consumers respond to different price incentives and event times.

Figure 4. Average Treatment Effect of Participants by Program and Event Type



Notes: The reported results are program- and event type-specific marginal effects calculated from estimating $\widehat{\beta}_j$ in (8) for $j \in \{C, M, T\}$, adjusted to allow for event-type interactions with the program indicator variables D_i. We present the marginal effects to be a percentage change in consumption using the transformation $100 \times (exp(\widehat{\beta}_j) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

Figure 4 presents the estimated response to events allowing for differential responses by event type. For the Central program, we see a large demand reduction for all event types, with an approximate 27% reduction during morning events, 25%

during evening events, and a 27% average reduction during high evening events. This indicates that the Central households allowed central management of demand during both morning and evening times. It also indicates that they were not distinctly more responsive to the greater incentives offered during the High Evening events.

During the Evening and High Evening events, the Tech program reduced its demand by approximately 6%, while the Manual program had a 4% estimated reduction in demand during these event types. These effects are statistically different from zero. The Evening and High Evening Tech and Manual program effects are not significantly different from each other, when compared within each event type.

The Tech program had a response to Morning events that are not statistically different from zero. This differs (statistically significantly) from the Manual program, which had an average estimated reduction of 8% during the Morning events. This is a counter-intuitive result, as the Tech program had all the same information, incentives, and abilities as the Manual program in making electricity consumption reductions during events, with the added ability to remotely control thermostats, EV chargers, and hot water heaters on which they have load controllers installed.

For all three programs, the change in consumption during High Evening events does not statistically significantly differ from their responses to regular Evening events. This suggests that the increased financial incentives does not motivate participants to undertake additional effort to make greater reductions in usage. This result suggests the barrier to demand responsiveness may have less to do with the scale of financial rewards and more to do with the hurdle of the opportunity costs of time/effort.[26]

Similar to the discussion in the previous section, Appendix Table C2 provides the results of our regression analysis when we vary the comparison group used to estimate event treatment effects. These results continue to demonstrate that our estimated treatment effects are largely robust to varying the comparison groups included in the regressions. The estimated response to morning events for the Tech and Manual programs varies with the relevant comparison group, with a smaller response for the Manual and a larger response for the Tech than our main specification. However, our key conclusions persist. The Central program is considerably more responsive to all

[26]That said, participants were only eligible to receive $1 more for achieving a 30% reduction in electricity use during a High Evening vs. Evening event. The reward for achieving a 50% reduction was doubled ($6 vs. $3). It is possible that a larger scaling of incentives could induce a greater response. However, given our rewards fall in the range of wholesale price caps observed in practice, it is unlikely that incentives provided in a real-world setting would be considerably larger than the amounts we provided.

event types. Further, no program shows a distinct response to the elevated incentives during the high evening events.

Finally, Appendix Table C4 provides the results for our test of the validity of our no carryover assumption, allowing for differential estimates for the morning and evening event windows. In this analysis, we find no evidence of changes in behavior during either the morning or evening event windows on non-event days in the post-treatment period for any of our demand response programs.

7 Time, Effort, and Attention

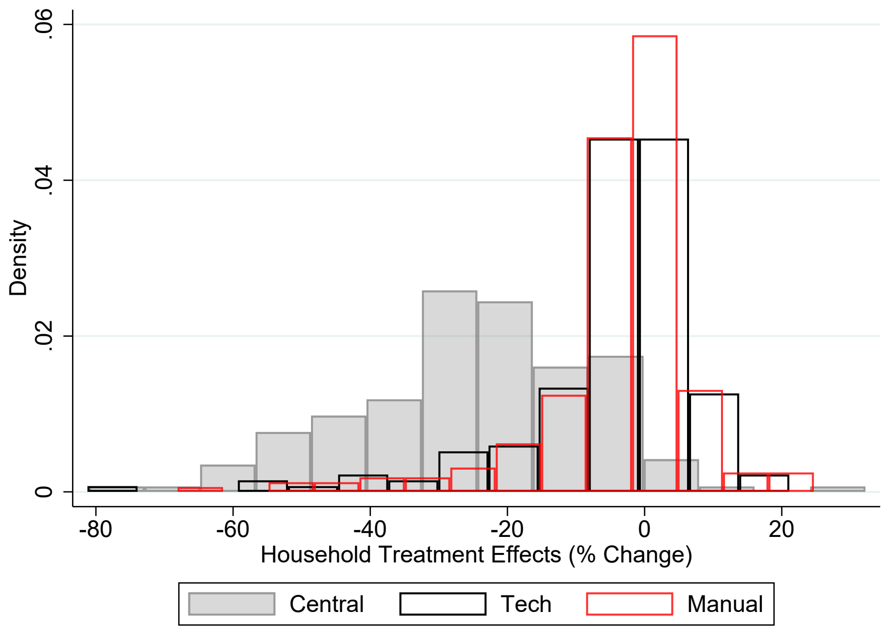
Our results have so far focused on average estimated treatment effects by demand response programs. Overall, our main results are in line with the model presented in Section 2. Central program participants, for whom the least time and effort was required to respond to events, display the highest reductions in electricity usage during events. Participants in the Tech and Manual programs reduced electricity much less on average during events, with similar behavior across the two programs.

A natural and important economic question is: What drives these differences across the three programs? In particular, what behavior underlies the differences between the responses to events of the Central compared to the Tech programs? In this section, we estimate household-level treatment effects and investigate the extent to which there is heterogeneity in event responsiveness. We then look at potential drivers of this heterogeneity. We use App interaction data during events to analyze the extent to which time allocation/effort relates to the household-specific estimated treatment effects. Additionally, we use survey data on household income and the stated value of event participation to look at the relationship between the households' opportunity cost of time and response to events.

7.1 Household-Level Treatment Effects

Figure 5 presents the distributions of household-level treatment effects by program, estimated as per Equation (9). We observe that the Tech and Manual programs' treatment effects are tightly distributed near zero. On average, households in the Tech and Manual programs reduce their consumption during events by 4.6% and 4%, respectively. This corresponds closely to the estimated program-level treatment effect in the previous section. Only 20% and 18% of the household-level estimated treatment effects are negative and significant at the 5% level for the Tech and Manual programs, respectively.

Figure 5. Household-Level Estimated Treatment Effect Distributions by Program

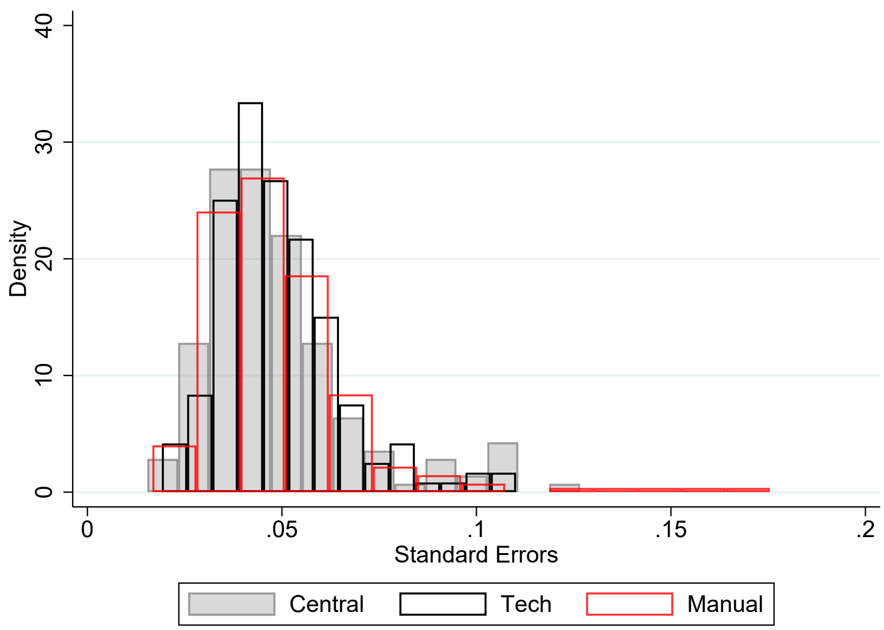


Notes. The reported results summarize the distribution of estimated household-level treatment effects obtained from estimating specification (9). We present the marginal effects to be a percentage change in consumption using the transformation $100 \times (exp(\widehat{\beta}_i) - 1)$.

Despite the fact that the Tech program has enabling devices, the *distribution* of their estimated treatment effects closely resembles that of the Manual households. In both the Tech and Manual programs, we observe a long left tail, indicating that there is a subset of households exhibiting large responsiveness. This is striking, as even though the Tech and Manual programs have similar mean responses when looking across all households, we might expect that the Tech program would contain a subset of households with higher household-specific event responses given their ability to remotely control their hot water heaters, EV chargers, and/or thermostats during events. There is only modest evidence to support this; the Manual households' estimated effects are more tightly distributed near zero.

Figure 5 also re-confirms the larger response from households in the Central program. The average household-level treatment effect is a reduction of 24% during events for households in the Central program, again closely reflecting the estimated effects in the program-level regressions above. There is a broad range of household estimates in the Central program, consistent with the fact that there is variation in controllable household devices. However, only the largest household estimates in the Manual and Tech programs achieve reductions that are in the range that is typical in Central. In contrast to these programs, 80% of the Central household-level estimated treatment effects are negative and significant.[27]

Figure 6. Standard Errors of the Household-Level Estimated Treatment Effects by Program



Notes. The reported results plot the standard errors of the estimated household-level treatment effects from specification (9).

Figure 6 plots the distribution of the standard errors of the estimated household-level treatment effects. These results demonstrate that while the Central program has a slightly lower standard error on average, the distributions are similar across all three programs. The similarity in the precision of the estimates, combined with the average treatment effects noted above, suggest that the Central households are consistently large responders to events, while the Manual and Tech households are consistently low responders.[28] These findings provide important policy implications because they

[27] Appendix D.3 estimates household-level treatment effects by program and event type. We continue to find no evidence of a larger response to the elevated incentives during high peak events.

[28] There are several large standard error estimates for the Manual program. This small subset of households has an average consumption reduction of 8.4% during events, more than double the average household treatment effect for the Manual program of approximately 4%. These large standard errors suggest that these households are relatively inconsistently large responders to events.

suggest that in addition to having larger estimated reductions in consumption during events, the Central program provides a reliable source of demand-side flexibility. Appendix D.4 provides additional empirical evidence demonstrating the consistency in the responses to events by demand response program throughout our experiment.

7.2 App Interactions

Despite the lower average response to events for the Tech and Manual households, Figure 5 demonstrates there is a small subset of high performers in the Tech and Manual programs that have large negative estimated treatment effects. In this section, we leverage App interaction data to evaluate if these high-performers differ in their use of the App during events. While households could be aware of an upcoming event without opening the App through push notifications sent to their phones, App interactions are a strong indication that the household is aware of the demand response event and serves as a proxy for time spent/effort to respond.[29]

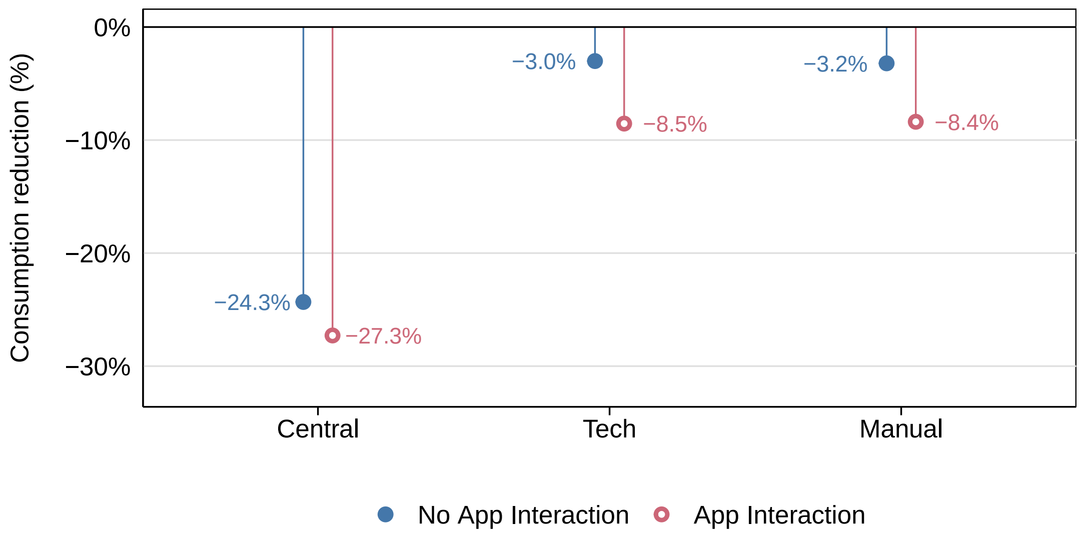
The App data tell us when users are interacting with the App on a given day, as well as more details on which features of the App they are accessing. For the Central and Tech households, interacting with the App allows them to control connected devices. For the Central program, the App can be used to opt out of automatic load control before or during events. In all programs, the App allows the household to observe the details of upcoming events 21 hours in advance, detailed information about household consumption in real-time, and performance in previous events.

Figure 7 reports the average estimated household-specific treatment effects, allowing for heterogenous treatment effects by whether or not the household interacted with the App on a given day. For the Central program, the average estimated household-level treatment effect is approximately -24% when the household does not interact with the App, increasing in magnitude to -27% when they do interact with the App during an event day. In contract, for the Manual and Tech programs, the average household-level treatment effect is -3% when they do not interact with the App, increasing in magnitude to roughly -8.5% when they interact with the App.

The results shown in Figure 7 provide two key findings. First, there is a positive relationship between App interactions and demand reductions. Households who were more attentive to the App on event days achieved higher reductions. This differential effect is slightly larger for the Tech and Manual programs, where the change in demand

[29] In addition to actively opening the App on their phone, the App opens and logs an interaction if a user clicks on the event push notification sent to their phone.

Figure 7. Household-Level Average Treatment Effects by App Interaction



Notes. The reported results are average household-level treatment effects by whether or not the household interacted with the App on an event day. This represents the specification in equation (9), adjusted to interact the event indicator (E_{it}) with an App Indicator$_{it}$ variable that equals 1 when the household interacts with the App on the event day and zero otherwise. All specifications include fixed effects at the year-month, day-of-week, and hourly levels.

reductions between interacting and not interacting with the App was approximately 5% as compared to about 3% for the Central program. Second, there is a massive difference in the "no interaction" estimates between Central program participants and those in the other programs (about -24% vs. -3%). This indicates that the Central program received a roughly 21% "headstart" over the other programs by not having to take action during the events to respond to them. This "headstart" is key to the overall finding of greater demand response by the Central program participants.

As noted above, we also have detailed data on the types of pages/tabs the households interact with in the App (i.e., General Interactions, Energy Usage Dial, Devices Tab, and Advisor Tab). This provides insight into the types of actions households may have taken to respond to events. Table 3 summarizes the average daily App interaction frequency for each demand response program separated by the performance quartile. The performance quartiles are determined by ranking household-level treatment effects over all demand response programs. This separates households into categories of whether or not they are high or low-performing households during our experiment. We summarize the count of the households and the percentage of households enrolled in each demand response program that fall within a specific performance quartile.

Table 3. Average Daily App Interaction Frequency by Program and Performance Quartiles

Program	Performance Quartile	Household Count	General Interactions	Energy Usage Dial	Devices Tab	Advisor Tab
Central	1	112 (63%)	0.32	0.30	0.18	0.26
	2	42 (24%)	0.20	0.18	0.12	0.16
	3	16 (9%)	0.20	0.19	0.14	0.15
	4	7 (4%)	0.15	0.15	0.11	0.11
Tech	1	20 (11%)	0.54	0.49	0.32	0.42
	2	45 (25%)	0.30	0.27	0.17	0.23
	3	52 (28%)	0.23	0.22	0.11	0.17
	4	67 (36%)	0.13	0.12	0.07	0.09
Manual	1	19 (8%)	0.60	0.59	0.06	0.50
	2	64 (26%)	0.18	0.18	0.03	0.15
	3	83 (34%)	0.15	0.15	0.03	0.12
	4	76 (31%)	0.14	0.14	0.02	0.11

Notes. The reported results provide the daily frequency of App Interactions by program and performance quartile. Performance quartiles are determined using the household-specific estimated treatment effects obtained from estimating specification (9). Household Count represents the number of households that fall within each performance quartile. The percentages report the percentage of households within a program that falls within each quartile. General Interactions reflect any interactions with the App. Energy Usage Dial displays the energy dial in the App that provides data on real-time usage. The Devices Tab displays a household's connected devices and allows households to adjust the use of the installed devices. The Advisor Tab reports information on upcoming events and historical performance on past events.

Table 3 demonstrates that the majority of Central households (63%) fall within the top quartile of performance. In contrast, only 11% and 8% of Tech and Manual households are in the top quartile. This is consistent with our results above that the Central program has a significantly greater demand response than the Tech and Manual programs, which show similar results.

Within each program, we see a reduction in the frequency of daily General Interactions as we move down the performance quartiles. This suggests that higher-performing households are more attentive and/or spend more time to monitor and react to events. In fact, across all columns, the frequency of App interactions in the top quartile is significantly greater (at the 5% level) than the frequency in the second-highest quartile.

We see a considerably higher frequency of App interactions among the Tech and Manual households that achieve the top quartile of performance, interacting with the App on 54% and 60% of days on average, respectively. In contrast, Central households in the top quartile only interact with the App on 32% of the days, suggesting that achieving this high threshold of performance required less time and attention. The top quartile of the Central program interacts significantly less with the App than the top quartile of the Tech and Manual programs (at the 5% level). This is true for each App interaction column in the table.

Looking across the App categories, households interacted with the Energy Usage Dial the most, followed by the Advisor Tab. This suggests that when households used the App, they often monitored their real-time consumption and the details of upcoming events and/or their past performance in the Advisor Tab. The top quartile performers in the Tech and Manual programs stand out in both of these categories, with the frequency of their interactions with both the Energy Usage Dial and Advisor Tab being roughly 2 to 4 times those of the bottom 3 quartiles.

Tech households in the highest-performing quartile interacted with the Devices Tab in the App at a significantly higher frequency than all other households. This suggests that a (small) subset of households in this program were using the installed devices to achieve larger demand reductions. However, looking at the lower performance quartiles in the Tech and Central programs, households in the Tech program interacted with the Devices Tab about as much as those in the Central program.[30]

Overall, these results suggest that households in the Tech and Manual programs had to allocate a higher degree of their time/effort to achieve relatively high demand reductions and the associated rewards. In contrast, the Central program participants had to invest less time in responding to events than others to generate the same scale of demand reductions.

7.3 Opportunity Cost of Time Preferences

In this section, we use data collected from a survey at the end of our experiment (in June 2023) to further understand the mechanisms driving our results. Specifically, participants in our three demand response programs were asked questions to better understand their opportunity costs of time. We correlate households' responses to these survey questions with their estimated household-level treatment effects, and evaluate if these results are consistent with the economic incentives outlined in the

[30]In addition to the installed devices from our experiment, a subset of households had other devices linked to the App. These were primarily smart plugs linked to lighting in the home. This helps explain why we observe interactions with the Devices Tab for the Manual program. Only 3 Manual households installed and linked devices to the App during our experiment, all three were thermostats for electric baseboard heaters.

conceptual framework in Section 2.

We observe a high response rate to the survey, with 75%, 71%, and 69% of households in the Central, Tech, and Manual programs completing the survey. Respondents were paid $20 upon completion of the survey. The results presented below focus only on the households that completed the survey.[31]

There were two key survey questions to assess households' opportunity costs of time. First, we asked participants their annual household income level and gave them several income categories (Less than $50k, $50-99k, $100-149k, $150-200k, and over $200k). Second, we asked a stated-preference question to assess the extent to which participants felt that participating in events was worth their time: "For the events you noticed, how often was it worth your time to participate by attempting to reduce your electricity consumption?" Respondents choose one of the following: (1) Never, (2) Sometimes, (3) About half the time, (4) Most of the time, or (5) Always. Appendix E.1 provides a more detailed summary of the exit survey.

Using the survey responses and household-level estimated treatment effects, we estimate the following equation:

$$Y_i = \beta_0 + \beta_1 I_i + \beta_2 Z_i + \beta_3 G_i + \gamma X_i + \epsilon_i \tag{10}$$

in which Y_i is the estimated treatment effect for household i (see Section 7.1), I_i is a household's reported income, and Z_i is the household's response to the Worth Time question described above. We include several control variables to control for a household's appliances and allocation to different demand response programs. G_i is an indicator variable for each demand response program and X_i is a vector of variables controlling for whether the household has an electric hot water heater, a categorical variable for whether a household has an electric vehicle and their corresponding charger type (No, Level 1, Level 2), a categorical variable for air conditioning in the home (No, Window Unit, Central Air), a categorical variable for electric baseboard heating (No, 1 – 3 Units, 4 or more units), and a house/duplex dummy variable (1 for house/duplex, 0 for row home). We report results with heteroskedastic robust standard errors.

The regression results are reported in Table 4. Focusing on the impact of household

[31] Appendix E.2 compares the characteristics of households that responded to the survey with those that did not using pre-treatment data. While the households are similar on a number of characteristics, responders consumed less electricity on average. To the extent that consumption is correlated with factors that relate to the opportunity cost of time, this could downward weight households with a higher opportunity cost of time in the end of experiment survey.

Table 4. Household Treatment Effect on Worth Time and Income

	Coefficient	Std. Error	P-Value
Worth Time			
Sometimes	-3.49	2.13	0.10
Half of the Time	-6.30	2.39	0.01
Most of the Time	-9.03	2.35	0.00
Always	-15.05	2.86	0.00
Income			
50 - 99k	5.53	4.56	0.23
100 - 149k	7.32	4.55	0.11
150 - 200k	9.89	4.55	0.03
>200 k	9.74	4.50	0.03
Program Indicators			
Central	-17.34	1.74	0.00
Tech	-1.37	1.57	0.38
Controls			
Electric Hot Water Heater	-6.33	1.51	0.00
Electric Vehicle			
Yes, Level 1	-1.07	2.48	0.67
Yes, Level 2	-3.65	2.12	0.09
Air Conditioning			
Yes, Window Unit	1.69	1.74	0.33
Yes, Central Air	0.93	1.85	0.61
Baseboard Heating			
Yes, 1 - 3 Units	3.01	2.23	0.18
Yes, 4 or more	1.75	1.76	0.32
Home/Duplex	3.70	2.25	0.10

Notes. The reported results present the estimates for Equation (10).

income, we see that higher income is associated with larger (positive) treatment effects, or smaller electricity consumption reductions during events. The magnitude of the coefficients are increasing in income, with less than \$50,000 per year being the excluded category. Coefficients for income brackets of \$150,000 per year or more are statistically different than the excluded lowest income bracket. These results are consistent with the framework in Section 2 that households with a higher opportunity cost of allocating time (e.g., the wage rate in the model) will spend less time providing demand flexibility (translating to a smaller estimated treatment effect).

For a given level of income, demand response program, and stock of household appliances, households may have heterogeneous preferences in their willingness to allocate time to provide demand flexibility. Table 4 finds that all of the Worth Time variable categories are negative and significantly different than the excluded "Never" category (with "Sometimes" being marginally significant). Additionally, the coefficients are more negative as we move down the spectrum of the Worth Time measure. These results suggest that the participants' perceived net benefits of allocating time to participate in events are correlated with greater reductions during events.

While these results focus on a stated-preference survey, they provide further support for the conceptual framework outlined in Section 2 to explain why we observe large differences in responses to peak events across our demand response programs. Factors that lead to a higher opportunity cost of time either financially or preference-based are associated with lower estimated household-level treatment effects. The Central demand response program, through the use of enabling technologies and utility-controlled/automated default responses to events, is able to overcome these time allocation and effort barriers.

8 Conclusion

The flexibility of electricity demand is becoming more valuable as electricity supply evolves to include a growing share of variable renewable sources. Moreover, there is a growing expectation that emerging technologies such as smart thermostats and electric vehicles will provide greater opportunities for flexible electricity demand. However, limited consumer responsiveness to dynamic electricity prices has long posed a problem for flexible demand to be meaningful.

Suspecting that inattention to dynamic pricing is rational—that consumers' rewards for paying attention to dynamic electricity prices and learning about how to respond to them are not worth the associated costs—we run a large-scale field experiment that tests the efficacy of utility-managed ("centralized") electricity demand on consumer responses to dynamic prices. Centralized demand management has the potential to take the burden of actively responding to price signals off of consumers' shoulders while allowing them (as well as other consumers and grid operators, in critical conditions) to reap the rewards of adjusting the timing of consumption with changing electricity system conditions.

We find that customers participating in a centralized demand management program, the Central program, reduced consumption by 26% on average during critical "peak events". In contrast, participants in the Tech program, who had the same smart technology as those in the Central program to remotely control baseboard thermostats, hot water heaters, and electric vehicle chargers, but had to initiate reductions themselves, only reduced consumption by 5% during events. This difference indicates that centralized electricity demand management has large potential to help consumers overcome barriers to respond to electricity prices. We find that the take-up rates between the Central and Tech programs are not appreciably different, suggesting that centralized electricity management is not as unpalatable as one might expect.

Somewhat surprisingly, we find that participants in the Tech program reduce consumption during events no more than those in our Manual program, who do not have smart, remote device adjustment capability. This indicates that smart home energy technology was not sufficient on its own to induce demand flexibility; overcoming the key barrier of effort/attention requires switching the default response, as per the Central program. The key takeaway here is that to achieve its full potential, technology needs to incorporate the behavioral realities of human effort and the opportunity costs of time as a barrier to responsiveness. Financial incentives motivate consumers to take some action, but consumers still face barriers that a centralized demand response program can resolve.

Interestingly, consumers in our context were not motivated to reduce consumption more when rewards were increased during periodic "high peak events". It is possible that consumers would need more than what we offered them to overcome the barriers that the Central program does. However, since our rewards were in line with peak electricity system prices, larger offers would not likely be economically efficient.

We are able to estimate household-specific treatment effects to events, leveraging our experimental design that has randomized household-specific event schedules. The distributions of household treatment effects across programs reveal that households in the Central program have a symmetrically distributed set of treatment effects, with the central mass of effects less than zero. In contrast, the Tech and Manual program participants display a distribution of household-level treatment effects that are centered around zero with a long left tail. This suggests that "high achievers" in these programs drove average treatment effects to events. Additionally, it suggests there is something about the Central program that facilitates the average household to respond to events by reducing consumption.

To understand the mechanisms behind our results, we use evidence from data on participant interaction with the experiment electricity management phone App. Across all programs, App interaction is correlated with larger household-level treatment effects. Average household-level treatment effects when households do not interact with the App are about 3% for the Tech and Manual programs and 24% for the Central program. When households *do* interact with their App, these numbers increase to about 8.5% for the Tech and Manual programs, and 27% for the Central program. This highlights the Central program's "headstart", whereby its participants achieve consumption reduction even in the absence of App interaction. We find that "high achievers" in the Tech and Manual programs, who drive the average consumption reductions during events for these programs, interact with their App on 60% and 54% of days on average during the experiment, whereas the high achievers in the Central program interacted with the App on average significantly less (32% of days on average). This suggests that high achievers in the Tech and Manual program devoted a lot of attention to their electricity consumption and time/effort in reducing it during events. Taken together, this evidence points to attention and effort (in the form of app interaction) being an important component of responsiveness to events, and that the Central program relieved participants of needing to devote such time to electricity management to achieve large consumption reductions during events.

Given our results, we surmise that programs and policies that relieve consumers of cognitive, time, and other burdens that contribute to rational inattention will have large potential to lead to welfare improvements. In the case of residential electricity, we see centralized demand as one such program, having the potential to both save money for consumers and facilitate flexible demand to meet emerging needs of electricity markets.

References

Abadie, Alberto and Sebastien Gay, "The impact of presumed consent legislation on cadaveric organ donation: a cross-country study," *Journal of health economics*, 2006, *25* (4), 599–620.

Bailey, Megan, David Brown, Erica Myers, Blake Shaffer, and Frank Wolak, " Electric Vehicles and the Energy Transition: Unintended Consequences of Time-of-Use Pricing ," *American Economic Review: Insights*, 2025.

Becker, Gary S, "A Theory of the Allocation of Time," *The Economic Journal*, 1965, *75* (299), 493–517.

Bernheim, B Douglas, Andrey Fradkin, and Igor Popov, "The welfare economics of default options in 401 (k) plans," *American Economic Review*, 2015, *105* (9), 2798–2837.

Blonz, Joshua, Karen Palmer, Casey J Wichman, and Derek C Wietelman, "Smart thermostats, automation, and time-varying prices," *American Economic Journal: Applied Economics*, 2025, *17* (1), 90–125.

Boiteux, Marcel, "Peak-Load Pricing," *The Journal of Business*, 1960, *33* (2), 157–179.

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard, "Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective," *Quantitative Economics*, 2021, *12* (4), 1171–1196.

Bollinger, Bryan K and Wesley R Hartmann, "Information vs. Automation and Implications for Dynamic Pricing," *Management Science*, 2020, *66* (1), 290–314.

Brandon, Alec, Christopher M Clapp, John A List, Robert D Metcalfe, and Michael Price, "The Human Perils of Scaling Smart Technologies: Evidence from Field Experiments," Technical Report, National Bureau of Economic Research 2022.

Brown, David P and Derek EH Olmstead, "Measuring Market Power and the Efficiency of Alberta's Restructured Electricity Market: An Energy-Only Market Design," *Canadian Journal of Economics*, 2017, *50* (3), 838–870.

Burkhardt, Jesse, Kenneth T Gillingham, and Praveen K Kopalle, "Field experimental evidence on the effect of pricing on residential electricity conservation," *Management Science*, 2023, *69* (12), 7784–7798.

Charness, Gary, Uri Gneezy, and Michael A Kuhn, "Experimental methods: Between-subject and within-subject design," *Journal of economic behavior & organization*, 2012, *81* (1), 1–8.

Chetty, Raj, John N Friedman, Søren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen, "Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark," *The Quarterly Journal of Economics*, 2014, *129* (3), 1141–1219.

DTE Energy, "DTE Energy Smart Charge Program," 2022.

Einav, Liran, Benjamin Klopack, and Neale Mahoney, "Selling Subscriptions," *American Economic Review*, Forthcoming.

Faruqui, Ahmad and Sanem Sergici, "Household Response to Dynamic Pricing of Electricity: A Survey of 15 Experiments," *Journal of Regulatory Economics*, 2010, *38* (2), 193–225.

Fowlie, Meredith, Catherine Wolfram, Patrick Baylis, C Anna Spurlock, Annika Todd-Blick, and Peter Cappers, "Default Effects and Follow-on Behaviour: Evidence From an Electricity Pricing Program," *The Review of Economic Studies*, 2021, *88* (6), 2886–2934.

Harding, Matthew and Carlos Lamarche, "Empowering Consumers Through Data and Smart Technology: Experimental Evidence on the Consequences of Time-of-Use Electricity Pricing Policies," *Journal of Policy Analysis and Management*, 2016, *35* (4), 906–931.

_ **and Steven Sexton**, "Household response to time-varying electricity prices," *Annual Review of Resource Economics*, 2017, *9* (1), 337–359.

IRC, "Market Design Executive Summary: 2017 IRC Markets Committee," Technical Report, ISO/RTO Council 2017.

Joskow, Paul L and Catherine D Wolfram, "Dynamic Pricing of Electricity," *American Economic Review*, 2012, *102* (3), 381–385.

Kahn, Alfred E., *The Economics of Regulation: Principles and Institutions: Volume 1: Economic Principles*, John Wiley, 1970.

List, John A, "The Experimentalist Looks Within: Toward an Understanding of Within-Subject Experimental Designs," Technical Report, National Bureau of Economic Research 2025.

PG&E, "SmartAC Switches," 2023. https://www.pge.com/en_US/residential/save-energy-money/savings-solutions-and-rebates/smart-ac/smartac-switch.page. Accessed: 2-27-2023.

Pollak, Robert A and Michael L Wachter, "The relevance of the household production function and its implications for the allocation of time," *Journal of Political Economy*, 1975, *83* (2), 255–278.

Prest, Brian C, "Peaking interest: How awareness drives the effectiveness of time-of-use electricity pricing," *Journal of the Association of Environmental and Resource Economists*, 2020, *7* (1), 103–143.

Schneider, Ian and Cass R Sunstein, "Behavioral Considerations for Effective Time-Varying Electricity Prices," *Behavioural Public Policy*, 2017, *1* (2), 219–251.

Shepard, Mark and Myles Wagner, "Do ordeals work for selection markets? Evidence from health insurance auto-enrollment," *American Economic Review*, 2025, *115* (3), 772–822.

Smith, Jordan, "Texas Regulators Lower Electric Price Cap to $5,000 per Megawatt Hour," 2022. https://www.chooseenergy.com/news/article/texas-regulators-lower-electric-price-cap/.

Spector, Julia, "Portland General Electric Jumps Into Virtual Power Plant Business," 2020. https://www.greentechmedia.com/articles/read/portland-general-electric-launches-525-home-battery-fleet-virtual-power-plant.

Statistics Canada, "Census of Population. Data Products, 2016 Census," 2021. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm, Accessed: 2021-07-15.

Wattersaver, "WatterSaver Program," 2023. https://www.watter-saver.com/. Accessed: 2-27-2023.

Yan, Xing, Yusuf Ozturk, Zechun Hu, and Yonghua Song, "A Review on Price-Driven Residential Demand Response," *Renewable and Sustainable Energy Reviews*, 2018, *96*, 411–419.

A Conceptual Model Derivations

Using (1) - (3) and (5), (4) can be rewritten as:

$$
\begin{aligned}
p_1 X_1 + p_2 X_2 &= \left(\overline{T} - \sum_{i=1}^{3} t_i \right) \omega + \rho Z_3 \\
\Rightarrow \quad p_1 a_1(R) Z_1 + p_2 a_2(R) Z_2 &= \left(\overline{T} - \sum_{i=1}^{3} b_i(R) Z_i \right) \omega + \rho Z_3 \\
\Rightarrow \quad \sum_{i=1}^{3} \pi_i(R) Z_i &= \overline{T} \omega \qquad (11)
\end{aligned}
$$

where $\pi_i(R)$ are defined in (7).

(1) - (4), (7), and (11) implies that the utility maximization problem can be rewritten as (6) and (7), reflecting a standard utility maximization problem where $\pi_i(R)$ represents the price of household good Z_i. Consequently, the solution is achieved by equating the marginal rate of substitution with the price ratio of any two household goods.

To illustrate, suppose that the household's preferences are represented by a Cobb-Douglas utility function:

$$
U(Z_1, Z_2, Z_3) = Z_1^{\alpha_1} Z_2^{\alpha_2} Z_3^{1-\alpha_1-\alpha_2} \qquad (12)
$$

where α_1 and α_2 are positive constants. Define λ to be the lagrangian multiplier on the constraint. Using (6), (7), and (12), the solution to the household problem is characterized by the following conditions:

$$
\begin{aligned}
\alpha_1 Z_1^{\alpha_1-1} Z_2^{\alpha_2} Z_3^{1-\alpha_1-\alpha_2} &= \lambda \pi_1(R); \qquad (13) \\
\alpha_2 Z_1^{\alpha_1} Z_2^{\alpha_2-1} Z_3^{1-\alpha_1-\alpha_2} &= \lambda \pi_2(R); \text{ and} \qquad (14) \\
(1-\alpha_1-\alpha_2) Z_1^{\alpha_1} Z_2^{\alpha_2} Z_3^{-\alpha_1-\alpha_2} &= \lambda \pi_3(R) \qquad (15)
\end{aligned}
$$

(13) and (15) imply:

$$
\left(\frac{1}{\pi_1(R)} \right) \left[\alpha_1 Z_1^{\alpha_1-1} Z_2^{\alpha_2} Z_3^{1-\alpha_1-\alpha_2} \right] = \left(\frac{1}{\pi_3(R)} \right) \left[(1-\alpha_1-\alpha_2) Z_1^{\alpha_1} Z_2^{\alpha_2} Z_3^{-\alpha_1-\alpha_2} \right]
$$

$$\Rightarrow \quad \frac{\pi_3(R)}{\pi_1(R)} = \frac{(1-\alpha_1-\alpha_2)\, Z_1^{\alpha_1}\, Z_2^{\alpha_2}\, Z_3^{-\alpha_1-\alpha_2}}{\alpha_1\, Z_1^{\alpha_1-1}\, Z_2^{\alpha_2}\, Z_3^{1-\alpha_1-\alpha_2}}$$

$$\Rightarrow \quad Z_1 = \left(\frac{\alpha_1}{1-\alpha_1-\alpha_2}\right)\left(\frac{\pi_3(R)}{\pi_1(R)}\right) Z_3\,. \tag{16}$$

(14) and (15) imply:

$$\left(\frac{1}{\pi_2(R)}\right)\left[\alpha_2\, Z_1^{\alpha_1}\, Z_2^{\alpha_2-1}\, Z_3^{1-\alpha_1-\alpha_2}\right] = \left(\frac{1}{\pi_3(R)}\right)\left[(1-\alpha_1-\alpha_2)\, Z_1^{\alpha_1}\, Z_2^{\alpha_2}\, Z_3^{-\alpha_1-\alpha_2}\right]$$

$$\Rightarrow \quad \frac{\pi_3(R)}{\pi_2(R)} = \frac{(1-\alpha_1-\alpha_2)\, Z_1^{\alpha_1}\, Z_2^{\alpha_2}\, Z_3^{-\alpha_1-\alpha_2}}{\alpha_2\, Z_1^{\alpha_1}\, Z_2^{\alpha_2-1}\, Z_3^{1-\alpha_1-\alpha_2}}$$

$$\Rightarrow \quad Z_2 = \left(\frac{\alpha_2}{1-\alpha_1-\alpha_2}\right)\left(\frac{\pi_3(R)}{\pi_2(R)}\right) Z_3. \tag{17}$$

Using, (11), (16), and (17), we can solve for the utility-maximizing level of Z_3 (demand flexibility):

$$\pi_1(R)\left(\frac{\alpha_1}{1-\alpha_1-\alpha_2}\right)\left(\frac{\pi_3(R)}{\pi_1(R)}\right) Z_3 + \pi_2(R)\left(\frac{\alpha_2}{1-\alpha_1-\alpha_2}\right)\left(\frac{\pi_3(R)}{\pi_2(R)}\right) Z_3 + \pi_3(R)\, Z_3 = \overline{T}\,\omega$$

$$\Rightarrow \quad \pi_3(R)\, Z_3 \left\{\left(\frac{\alpha_1}{1-\alpha_1-\alpha_2}\right) + \left(\frac{\alpha_2}{1-\alpha_1-\alpha_2}\right) + 1\right\} = \overline{T}\,\omega$$

$$\Rightarrow \quad \frac{\pi_3(R)}{1-\alpha_1-\alpha_2}\, Z_3 \left\{\alpha_1 + \alpha_2 + 1 - \alpha_1 - \alpha_2\right\} = \overline{T}\,\omega$$

$$\Rightarrow \quad Z_3 = \frac{(1-\alpha_1-\alpha_2)\, \overline{T}\,\omega}{\pi_3(R)}. \tag{18}$$

(7) and (18) imply that Z_3^* varies with $b_3(R)$, ρ, and ω as follows:

$$\frac{\partial Z_3^*}{\partial b_3(R)} = -\,(1-\alpha_1-\alpha_2)\, \overline{T}\,\omega^2\, (\pi_3(R))^{-2} < 0;$$

$$\frac{\partial Z_3^*}{\partial \rho} = (1-\alpha_1-\alpha_2)\, \overline{T}\,\omega\, (\pi_3(R))^{-2} > 0;$$

$$\frac{\partial Z_3^*}{\partial \omega} = \frac{(1-\alpha_1-\alpha_2)\, \overline{T}\, \pi_3(R) - (1-\alpha_1-\alpha_2)\, \overline{T}\,\omega\, b_3(R)}{[\pi_3(R)]^2}$$

$$\stackrel{s}{=} \pi_3(R) - \omega\, b_3(R) = -\rho < 0. \quad \blacksquare$$

B Supplementary Experimental Framework Material

B.1 Recruitment and Assignment

The study sample was drawn from the population of residential customers in the Utility's service territory in and near a large metropolitan city in Canada. We employed a two-step recruitment strategy. In Phase 1, starting in August 2021, the Utility invited households to join an App operated by a third-party company in partnership with the Utility. The App provides households with household-level hourly consumption posted at a one-day lag. The App can be coupled with other devices to provide more detailed information on real-time usage and device control. Households were recruited to the App using several marketing strategies, including advertisements on the Utility's website, social media posts, the Utility's newsletter, website notifications when users logged into their Utility accounts, and emails sent to approximately 306,000 residential households.

The recruitment onto the Utility's App provided us with a pool of 9,020 households to draw from. When households signed up to join the App, they were required to answer a six-question survey. The survey asked households about their motivation for joining the App and whether the household rents or owns their home. It asked about devices eligible for load control in our experiment, including whether the household has an electric hot water tank, an electric vehicle (EV), and electric baseboard heaters as the primary heat source. Households with EVs were asked what type of charger (level 1 or 2) they use. It also asked whether households have air conditioning, a major source of demand flexibility.

We applied several selection criteria to this pool of households. Customers had to be in or near a large metropolitan city in the province for which it was feasible for Utility-partnered electricians to install load control equipment, as needed. Only homeowners were permitted to participate. Condos and apartments were removed, leaving primarily single-family homes, duplexes, and row homes as eligible. Households must have at least one month of historical consumption data as of September 2021, and the customers must have at least one controllable electric device. Recall, the set of controllable electric devices includes a level 2 electric vehicle charger, electric baseboard heaters used as the primary heat source, and an electric hot water heater tank. This left us with a sample of 1,661 potential households that we used for our randomized assignment to experimental programs.

In Phase 2 of recruitment, we randomized the eligible households into our treatment programs and never-treated groups.[32] Starting in October 2021, we sent program-specific emails to households inviting them to join a new "Trial" program. These emails provided details about the specific experiences households would face in the program to which they were being invited, including a summary of the expected rewards they could earn over the course of the Trial, equipment they would receive and its estimated value, and future peak events. Households were also randomly offered a small sign-on incentive of the amounts \$10 or \$20, or no incentive. All households faced a yes/no decision regarding accepting our program-specific offer. The never-treated Control group that received no equipment, price incentives, or real-time usage information (recall Table 1) received no further communication beyond joining the App in the first phase of recruitment.

Households had to accept the invitation to join the relevant experimental program actively. After selecting to join, households were mailed a device called the "Hub" that facilitates monitoring real-time energy usage via the App. Installers contacted households in the Central and Tech programs to install the load controller equipment.

This two-phase recruitment process occurred over the months of August 2021 - February 2022. The second phase of recruitment occurred in five waves starting in October 2021. As additional households joined the App, we collected the survey responses, identified eligible households, randomized households into programs, and sent the second-phase recruitment emails. This process was used to facilitate the time required to schedule and install the load controllers, as well as to achieve the targeted sample size.

Finally, during the invitations to join each program, we randomized the upfront incentive. While we find a higher rate of initial acceptance with higher upfront incentive payments, the differences are small and not significantly different.[33]

[32] Specifically, we used a randomization procedure designed to balance important observable characteristics over programs and groups. We first used the machine learning algorithm "kmeans" to group households based on observable characteristics. These included cumulative household electricity consumption (in kWh) and load factor by season (Fall, Spring, Winter, and Summer), variables that indicate if a household has an electric vehicle, electric baseboard heating, or air conditioning, and census data on median household income. Load factor is the average electricity consumption divided by maximum consumption over a specific time period; it is a way to capture the relative utilization rate of consumption at the household level. We then randomized program assignments so households within a cluster were balanced across programs.

[33] Households that received a \$0, \$10, and \$20 upfront incentive accepted the initial invitation with a 63%, 67%, and 68% probability, respectively.

B.2 Comparison of Household Characteristics Upon Randomization

We evaluate if there are differences in pre-treatment characteristics across our various programs to assess the quality of our randomization. Table A1 provides summary statistics by program for a number of variables, including those used in the clustering procedure during randomization (recall the discussion in Footnote 32). The sample presented in this Table represents all 1,661 households invited to participate in the experiment. For all variables, we report the p-values from a one-way ANOVA test to evaluate if there are statistical differences in means across the programs.[34]

Table A1 shows that we do not find significant differences in key characteristics pre-treatment across our programs. These results indicate that our randomization approach effectively achieved balance on observables pre-treatment. In addition, Table A1 demonstrates that the majority of households in our sample have electric hot water heating and use baseboard heating as the primary heat source. In contrast, electric vehicles are less common, representing approximately 30% of households. The majority of households are single-family homes or duplexes, with the remainder being primarily row homes. The households consume considerably more electricity during the winter, with the lowest consumption arising in summer. This demonstrates the potential for larger opportunities for load shifting during these months.

B.3 Comparison of Household Characteristics After Acceptance

We compare the pre-treatment means in observable characteristics by program, including only the households that accepted our invitation to join each program. Large differences in observable characteristics would raise questions about the comparability of our estimated treatment effects from the main specifications.

Table A2 shows observables across programs for the final set of households included in each program. We observe limited differences in these characteristics across programs. The exceptions are that we find a statistically significant difference in the proportion of households that live in single-family homes/duplexes. There is a larger proportion of households in this building type in the Manual program than in other programs, in particular. We also observe a difference across programs in the proportion of households that have EVs, but this difference is only marginally statistically

[34] The seasonal cumulative consumption and load factor data only contain households with a full year's worth of historical consumption. We computed analogous statistics for the entire sample of households using only data from September 2021, the month in which all households have complete pre-treatment consumption data. We find no evidence of statistically significant differences in means across the programs using this data.

Table A1. Comparison of Means by Programs - Initial Randomization

	Central	Tech	Manual	Info	Control	ANOVA (p-value)
Cumul. kWh						
Winter	5,279	5,268	5,442	4,859	5,265	0.27
	(2,694)	(3,032)	(3,076)	(2,748)	(2,950)	
Spring	3,760	3,773	3,818	3,503	3,712	0.48
	(1,924)	(2,112)	(1,911)	(2,116)	(1,974)	
Summer	2,845	2,836	2,708	2,614	2,729	0.54
	(1,742)	(1,872)	(1,539)	(1,861)	(1,710)	
Fall	3,633	3,670	3,700	3,458	3,623	0.66
	(1,663)	(1,945)	(1,974)	(1,796)	(1,860)	
Load Factor						
Winter	24.66	24.98	25.41	24.73	24.67	0.81
	(8.20)	(8.15)	(8.80)	(8.29)	(8.63)	
Spring	19.52	20.12	20.01	19.28	19.91	0.65
	(7.25)	(6.97)	(6.70)	(7.73)	(7.41)	
Summer	16.82	16.55	16.73	16.12	16.32	0.82
	(7.89)	(6.30)	(5.93)	(8.11)	(8.29)	
Fall	18.56	18.90	19.34	18.42	19.06	0.42
	(5.89)	(6.23)	(6.00)	(6.48)	(6.50)	
Electric Vehicle	0.27	0.27	0.27	0.33	0.27	0.41
	(0.44)	(0.45)	(0.45)	(0.47)	(0.45)	
Baseboard Heating	0.61	0.64	0.61	0.63	0.63	0.95
	(0.49)	(0.48)	(0.49)	(0.48)	(0.48)	
Air Conditioning	0.52	0.51	0.50	0.51	0.54	0.95
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Electric Hot Water	0.70	0.66	0.70	0.66	0.72	0.38
	(0.46)	(0.47)	(0.46)	(0.47)	(0.45)	
House/Duplex	0.77	0.76	0.81	0.78	0.84	0.17
	(0.42)	(0.43)	(0.39)	(0.42)	(0.37)	
Median Income	86,376	88,291	85,931	87,470	85,948	0.48
	(19,503)	(22,227)	(19,255)	(21,574)	(21,541)	
Households	423	382	409	259	188	

Notes. This table compares pre-treatment average values across the five different programs. Parentheses contain the standard deviations. Cumul. kWh and Load Factor represent the cumulative household-level consumption and load factor by season. The seasonal cumulative consumption and load factor data only contain households with a full year's worth of historical consumption. Electric Vehicle, Baseboard Heating, Air Conditioning, and Electric Hot Water are indicator variables denoting the presence of each device. House/Duplex is an indicator variable that equals one if the home type is a single-family home or duplex and zero otherwise. Median Income reports the median household-level income of the Census Dissemination Area where the household is located. ANOVA reports the p-value from one-way ANOVA tests for differences in means across programs. Statistical significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

significant. Overall, these results suggest that the balance on observables that arose due to the initial randomization largely remains in the final sample.

Table A2. Comparison of Means by Program - Final Accepted Households

	Central	Tech	Manual	Info	Control	ANOVA (p-value)
Cumul. kWh						
Winter	5,507	5,302	5,422	5,037	5,265	0.71
	(2,706)	(2,737)	(3,240)	(2,768)	(2,950)	
Spring	3,900	3,739	3,797	3,642	3,712	0.85
	(1,934)	(1,791)	(1,939)	(2,159)	(1,974)	
Summer	2,851	2,672	2,766	2,702	2,729	0.93
	(1,869)	(1,759)	(1,547)	(1,849)	(1,710)	
Fall	3,754	3,550	3,677	3,547	3,623	0.86
	(1,733)	(1,659)	(1,992)	(1,788)	(1,860)	
Load Factor						
Winter	24.62	25.56	24.93	24.93	24.67	0.90
	(8.68)	(8.21)	(9.04)	(7.76)	(8.63)	
Spring	19.33	20.48	19.80	19.59	19.91	0.72
	(7.43)	(6.30)	(6.45)	(7.05)	(7.41)	
Summer	16.33	16.87	16.95	16.61	16.32	0.91
	(8.54)	(6.02)	(5.95)	(7.80)	(8.29)	
Fall	18.17	18.97	19.11	18.53	19.06	0.65
	(6.27)	(5.78)	(6.29)	(6.11)	(6.50)	
Electric Vehicle	0.25	0.21	0.30	0.34	0.27	0.07^{*}
	(0.43)	(0.41)	(0.46)	(0.47)	(0.45)	
Baseboard Heating	0.68	0.70	0.60	0.59	0.63	0.12
	(0.47)	(0.46)	(0.49)	(0.49)	(0.48)	
Air Conditioning	0.46	0.46	0.51	0.51	0.54	0.41
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Electric Hot Water	0.75	0.74	0.68	0.65	0.72	0.16
	(0.43)	(0.44)	(0.47)	(0.48)	(0.45)	
House/Duplex	0.82	0.77	0.89	0.84	0.84	0.02^{**}
	(0.39)	(0.42)	(0.32)	(0.37)	(0.37)	
Median Income	84,978	88,274	86,718	89,504	85,948	0.23
	(19,647)	(20,432)	(19,494)	(21,079)	(21,541)	
Households	177	184	242	177	188	

Notes. This table compares pre-treatment average values across the five different programs for households that were in our final programs. Parentheses contain the standard deviations. Cumul. kWh and Load Factor represents the cumulative household-level consumption and load factor by season. Electric Vehicle, Baseboard Heating, Air Conditioning, and Electric Hot Water are indicator variables denoting the presence of each device. House/Duplex is a indicator variable if the home type is a single-family home or duplex. Median Income reports the median household-level income of the Census Dissemination Area where the household is located. ANOVA reports the p-value from one-way ANOVA tests for differences in means across programs. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

C Treatment Details

C.1 Program-Specific Event Notifications

Each treatment program experienced event notifications tailored to their treatment. Each program received a notification 21 and 2 hours before an event. All participants were shown a short notification according to their device and in-app notification settings. If participants touched and pressed the notification, they were shown the long notification specific to their program, featured below, with event incentive details.

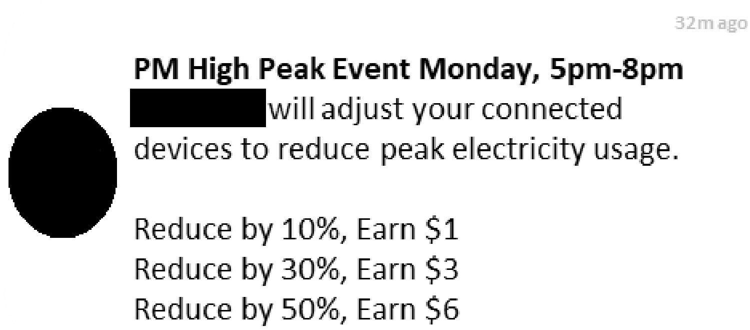


Figure B1. Long Notification for Central program

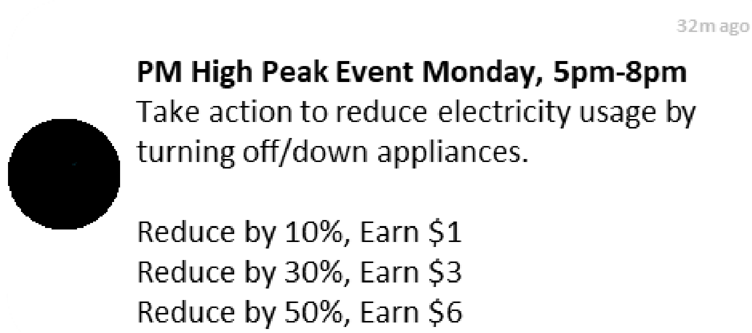


Figure B2. Long Notification for Tech program

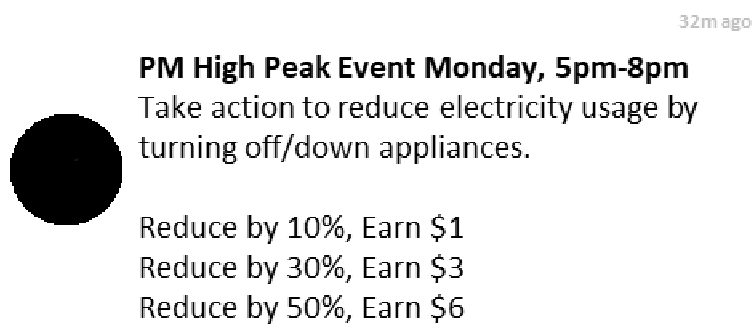


Figure B3. Long Notification for Manual program

Note that all program participants in the three programs were able to locate event details in the "Advisor" tab of the App, a centralized location for information from the App company, once they received an event notification. The "Learn More" button at the bottom right of this information card took participants to the "FAQs" section of the program-specific experiment website.

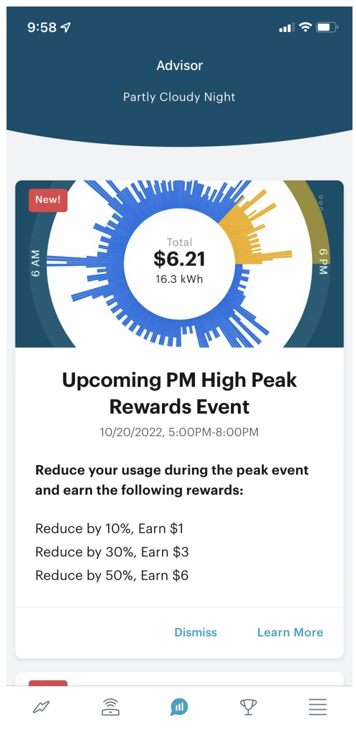


Figure B4. Event info in App

C.2 Treatment Program-Specific App Functionality

Each program in our experiment had an App experience and functionality that differed according to their program assignment. We detail that here and walk through how participants in each program could have responded to peak events, given the options in the App.

C.2.1 Central Program

The Central program participants receive 21-hr and 2-hr notifications regarding upcoming events, as described in Appendix C.1. These notifications allow them to see the timing of the event and the magnitude of rewards for electricity consumption reductions. They also remind participants that their devices with load controllers would be altered by the Utility to reduce consumption, unless they opted-out of the event.

There are several ways that Central program participants can opt-out of events. Before an event starts, they can push an "Opt-out" button in the "My Devices" tab of the App (Figure B5). (This tab is a central App location that allows App users to remotely control devices that have load controllers and see the individual electricity consumption of those devices.) This button removes the participant from the event globally by removing all of their load-controlled devices from the event.

If they do not opt-out in this way, they see a series of screens in the "My Devices" tab. These indicate the progression of the event to the participant and signal when their devices' electricity consumption is being controlled by the utility, via the icons above the text "You are opted in", "Event", and "Complete" (Figure B6).

During an event, participants can cancel Utility device control in a device-specific way. For EV chargers and hot water heaters, they can remotely opt-out their device from being controlled, or they can physically turn off the load controller at the device itself. For thermostats, participants can opt-out of load control by adjusting them physically or remotely through the App, during an event.

Note that the Central program has remote and manual control of all devices with load controllers, just like the Tech program. Central program households can also change anything else in the house to alter their electricity consumption during events.

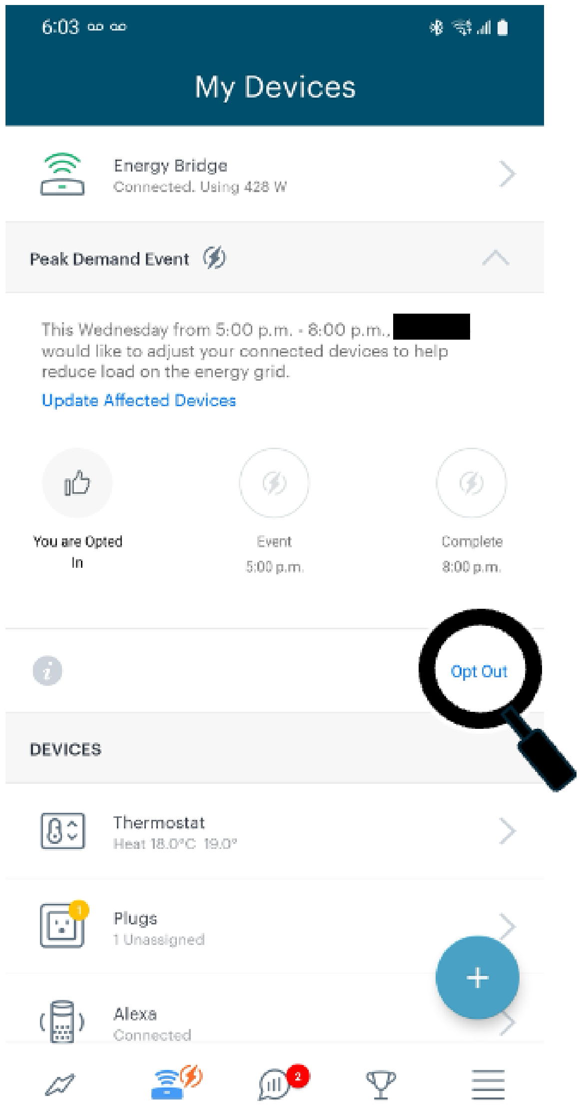


Figure B5. Central Program Opt-Out Functionality

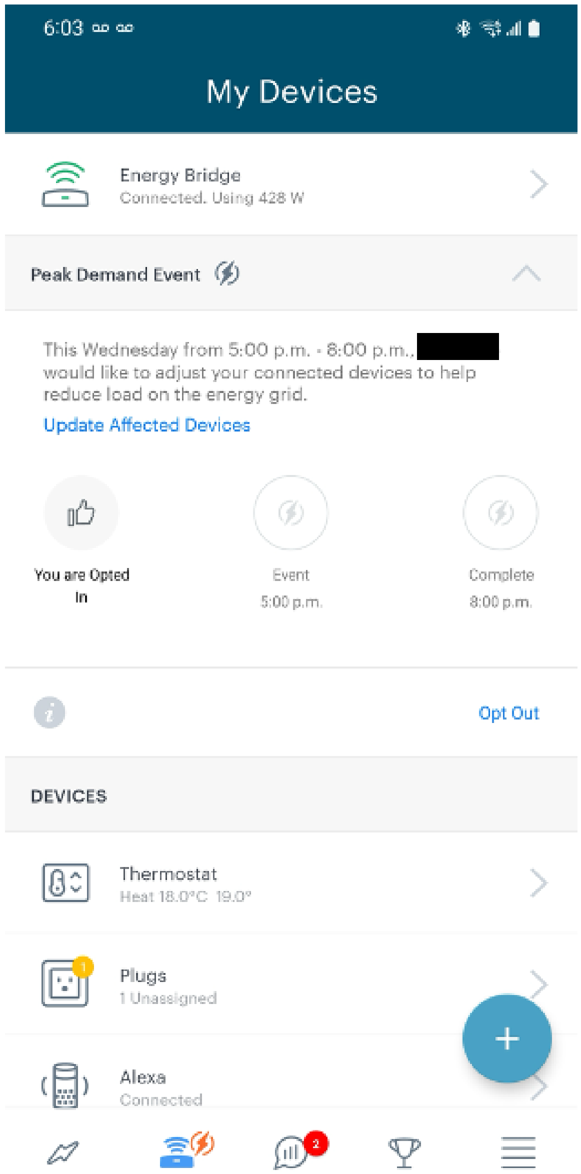


Figure B6. Central Program Event Experience

C.2.2 Tech Program

The Tech program participants receive 21-hr and 2-hr notifications regarding upcoming events, as described in Appendix C.1. These notifications allow them to see the timing of the event and the magnitude of rewards for electricity consumption reductions. They also remind participants that they need to “take action” to make consumption changes to receive the rewards offered.

The Tech program can remotely control any device that has an installed load controller through the App. For EV chargers and hot water heaters, they can turn them off via two clicks from the My Devices section of the App. (See Figure B7 below for the instructions sent to participants that explain these actions.) Tech program participants cannot make a schedule to turn off these devices before events start and must turn them off before or during events to reduce consumption this way. (They must also remember to turn them on unless they set up a turn-on schedule.)

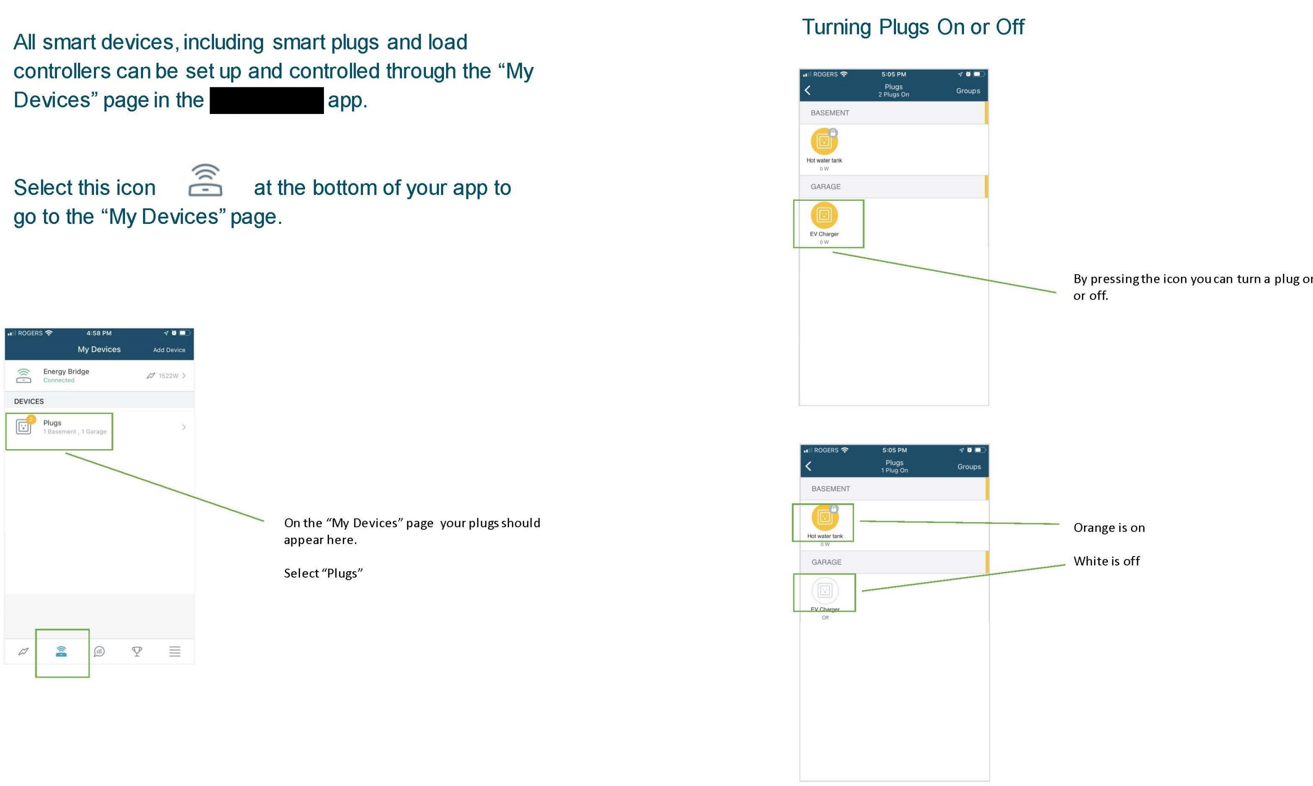


Figure B7. Controller Guide for Tech Program

For thermostats, the Tech program can set up schedule for their thermostat setpoint before events, using the App. They can also adjust their thermostats remotely during events with the App.

C.2.3 Manual Program

The Manual program participants receive 21-hr and 2-hr notifications regarding upcoming events, as described in Appendix C.1. These notifications allow them to see the timing of the event and the magnitude of rewards for electricity consumption reductions. They also remind participants that they need to "take action" to make consumption changes to receive the rewards offered.

Manual program participants do not load controllers given to them as part of this experiment or Utility control of any devices. They therefore only observe these notifications as well their aggregate, real-time household consumption through the App. If Manual program participants install their own smart home devices, they may be able to link them to the smart electricity consumption technology ecosystem used in this experiment. If so, they may have the capabilities of the Tech program to observe the real-time consumption of those devices/devices individually and adjust them remotely through the App. (Only three households in the Manual program installed their own smart thermostats over our sample period.)

C.2.4 Central, Tech, and Manual Programs

After each event, all three of the Central, Tech, and Manual programs receive a result on their performance, as depicted below. This appears in the "Advisor" tab of the App, a central location for information from the App company. This result card reminds participants of the event type (reward magnitudes being "high" or not) and the day and time of the event. It shows the incremental reward the participant earned from the event as well as their cumulative rewards throughout the entire experiment, including the reward from the prior event. The text below the reward for the last event is variable and depends on whether a participant met one of the reward tiers. The rewards screen with one of these text options is shown below in Figure B8.

From this rewards screen, participants can select "Event History" and see their recent history of event rewards, as shown in Figure B9.[35]

[35] Figure B9 was created for illustrative purposes using a series of simulated events. As a result, the event times differ from the event times considered in our study (i.e., 7:00 AM - 10:00 AM and 5:00 PM - 8:00 PM).

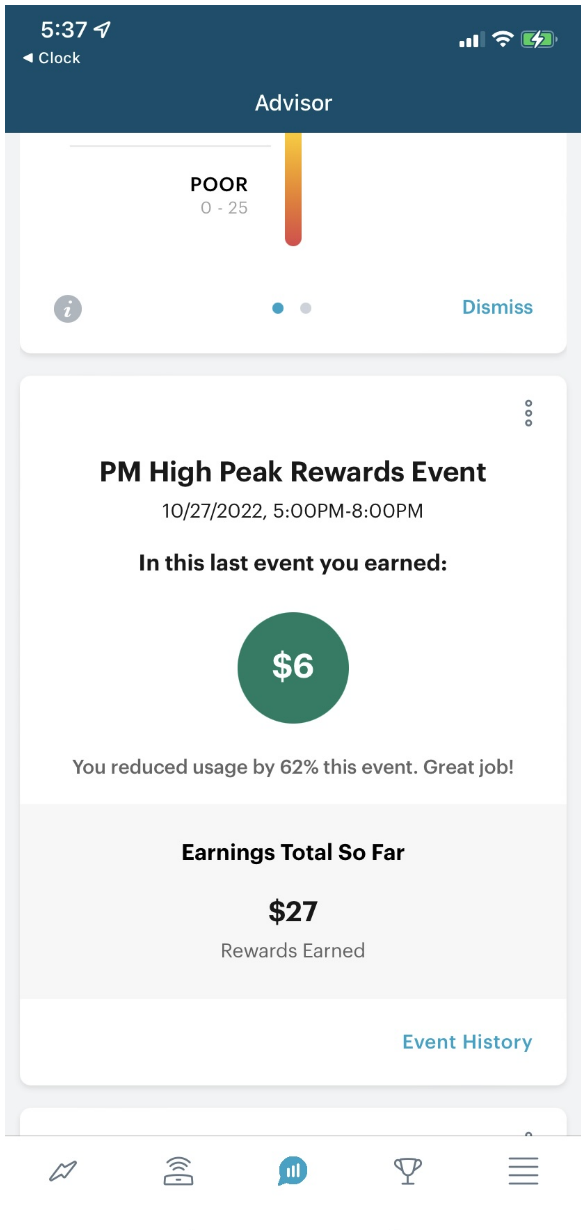


Figure B8. Rewards Screen

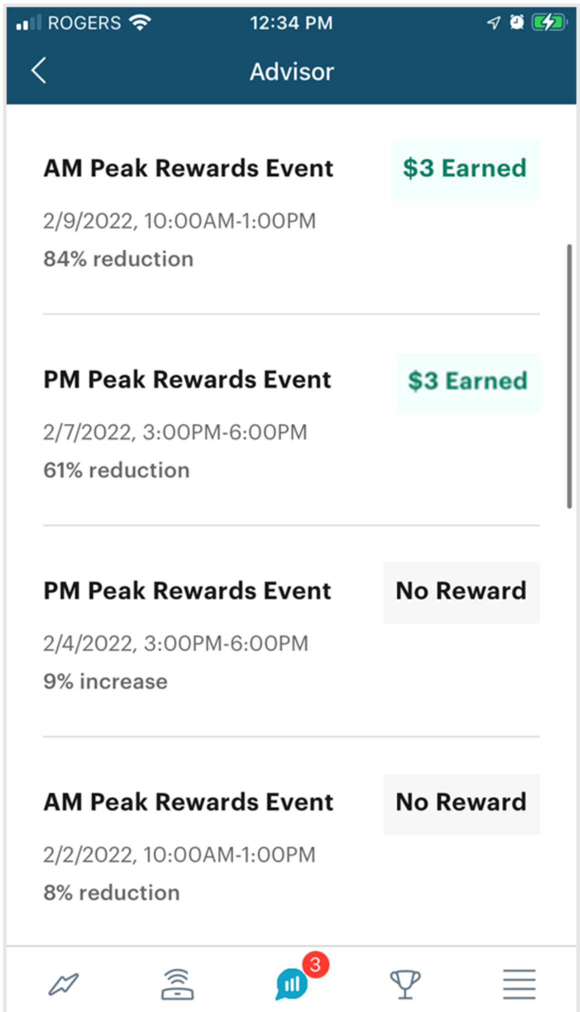


Figure B9. Event History

D Extensions and Robustness

D.1 Comparison Group in Program-Level Regression

In our main specification in Equation (8), our analysis includes all three demand response groups and never-treated households. As discussed in Section 5.1, this approach compares event time consumption to non-event time consumption of households in the same demand response program, other demand response programs, and never-treated households. Tables C1 and C2 present our estimated treatment effects of participants by program to events and separated by event type, respectively, allowing for regressions only including households in the same demand response program (Column (1)), same demand response program and the never-treated (Column (2)), and the results from our main specification in Column (3) for comparison purposes. Our results are consistent across all three specifications.

Table C1. Treatment Effects of Participants by Program

Program	(1)	(2)	(3)
Central	-0.3151***	-0.3007***	-0.3047***
	(0.0206)	(0.0204)	(0.0204)
Tech	-0.0661***	-0.0475***	-0.0495***
	(0.0132)	(0.0155)	(0.0159)
Manual	-0.0507***	-0.0459***	-0.0540***
	(0.0092)	(0.0117)	(0.0122)
Comparisons			
Own Program	Y	Y	Y
Other Treated			Y
Never Treated		Y	Y

Notes. The reported results are program-specific treatment effect coefficients. Standard errors are reported in the parentheses and clustered at the household level. Column (1) reports the regression results including only within demand response program comparisons, column (2) includes both within demand response program and the never-treated (Info and Control) groups, and column (3) reports results when all programs/groups are included. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table C2. Treatment Effects of Participants by Program and Event-Type

Program	(1)	(2)	(3)
Central			
Morning	-0.3198***	-0.3160***	-0.3133***
	(0.0221)	(0.0224)	(0.0223)
Evening	-0.3040***	-0.2850***	-0.2916***
	(0.0218)	(0.0226)	(0.0227)
High Evening	-0.3291***	-0.3105***	-0.3177***
	(0.0237)	(0.0241)	(0.0243)
Tech			
Morning	-0.0495***	-0.0273	-0.0158
	(0.0128)	(0.0188)	(0.0199)
Evening	-0.0713***	-0.0545***	-0.0623***
	(0.0155)	(0.0181)	(0.0185)
High Evening	-0.0786***	-0.0604***	-0.0675***
	(0.0166)	(0.0199)	(0.0205)
Manual			
Morning	-0.0488***	-0.0689***	-0.0812***
	(0.0097)	(0.0150)	(0.0159)
Evening	-0.0488***	-0.0332**	-0.0396***
	(0.0108)	(0.0143)	(0.0149)
High Evening	-0.0567***	-0.0403**	-0.0467***
	(0.0126)	(0.0159)	(0.0166)
Comparisons			
Own Program	Y	Y	Y
Other Treated			Y
Never Treated		Y	Y

Notes. The reported results are program-specific treatment effect coefficients by event type. Standard errors are reported in the parentheses and clustered at the household level. Column (1) reports the regression results including only within demand response program comparisons, column (2) includes both within demand response program and the never-treated (Info and Control) groups, and column (3) reports results when all programs/groups are included. Each regression is adjusted to include an event-type-specific categorical variable. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

D.2 No Carryover Assumption

As discussed in Section 5.1, an assumption in our identification strategy is that our randomized events do not impact (or "carryover" to have a treatment effect on) persistent changes in behavior in the event hours on non-event days. We test the validity of this assumption by estimating a DID regression. Separately for each demand response program, we run the following regression that excludes event days in the post-treatment period and includes the never-treated households:

$$ln(c_{it}) \;=\; \beta\, D_i \cdot \text{EventWindow}_{it} + \alpha_i + \tau_t + \delta X_{it} + \varepsilon_{it} \tag{19}$$

where EventWindow_{it} equals 1 in the post-treatment period for hours where morning or evening events occur and 0 otherwise. D_i equals 1 if the household is in a demand-response group (i.e., C, T, or M) and 0 otherwise. All other features of the regression analysis are identical to those in Equation (8). This analysis evaluates whether households in each demand response program adjusted their consumption during the event windows on non-event days in the post-treatment period, relative to the never-treated households. If households treated with events did not systematically alter their behavior on non-event days in response to being exposed to events, β should be statistically indistinguishable from zero. In addition, we consider a specification that estimates separate effects for the morning and evening event windows on non-event days.

Table C3 presents the results of our no carryover assumption DID test, described in Section 5.1. Table C4 presents the results when we allow for differential effects across the morning and evening event windows on non-event days. In both specifications, we find no evidence of changes to non-event day consumption during the event windows.

Table C3. Carry Over DID Estimates by Program

	Central	Tech	Manual
Event Window	0.0241	0.0302	-0.0014
	(0.0169)	(0.0181)	(0.0155)

Notes. The reported results are the program-specific Event Window coefficients from equation (19). For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, and $^{***}\,p < 0.01$.

Table C4. Carry Over DID Estimates by Program and Event Type

	Central	Tech	Manual
Morning Event Window	0.0188	0.0353	-0.0226
	(0.0206)	(0.0235)	(0.0201)
Evening Event Window	0.0295	0.0251	0.0196
	(0.0212)	(0.0210)	(0.0185)

Notes. The reported results are the program-specific Event Window coefficients from equation (19), allowing for differential effects during the morning and evening event windows. For each demand response program, the sample includes households from their own treatment program and the never-treated groups. Standard errors are reported in the parentheses and clustered at the household level. All specifications include fixed effects at the household and hour-of-sample levels. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

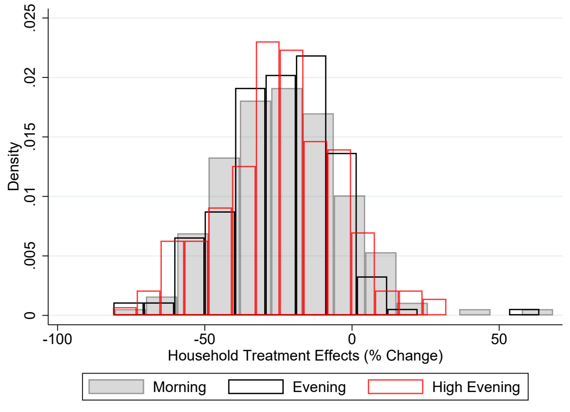
D.3 Household Treatment Effects by Event Type

Figure B10 presents the distributions of our household-level treatment effects by program and event type. More specifically, we estimate the specification detailed in (9) but permit different treatment effects by event type: morning, evening, and high evening.

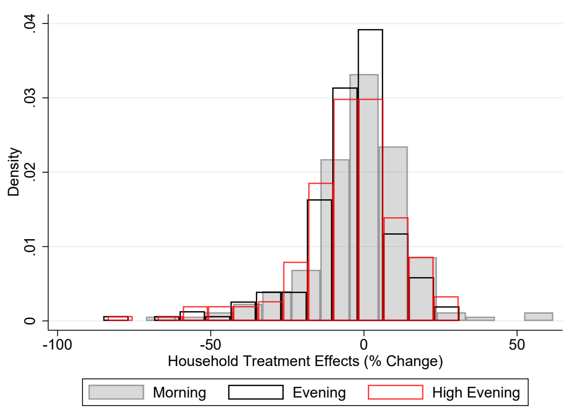
For the Central program, looking across all households, the average percentage reduction in consumption ranges from 23% during morning events to 24% in evening and high evening events. For Tech, this ranges from 1% during morning events to 5% during evening and high evening events. Finally, for Manual, the average reduction is 3% during morning events and 4% during evening and high evenings. While the precise values vary slightly, the average household-level treatment effects are in a similar range to those found in the program-level regressions summarized in Section 6.2. Importantly, we do not find any evidence of a larger average reduction in consumption during high peak evening events compared to evening events.

A benefit of estimating household-level treatments effects is that we can observe the full distributions for evening and high evening events. Figure B10 shows that the Central program's distributions are quite similar by event type. For the Tech and Manual programs, there is some evidence to suggest that the response to evening events are more tightly distributed around zero compared to the high evening events. However, this does not translate into differences in average treatment effects when looking across all households, as documented above.

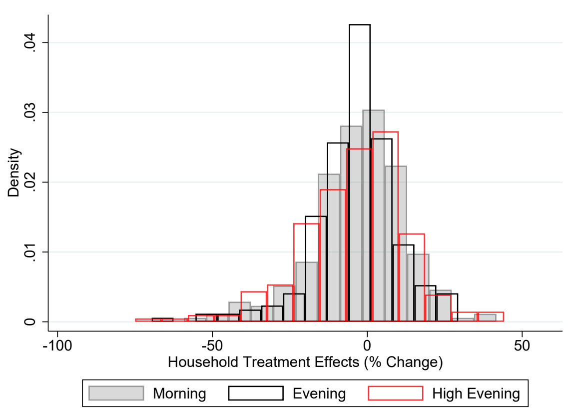
Figure B10. Household-Level Estimated Treatment Effect Distributions by Program and Event-Type



(a) Central Program



(b) Tech Program



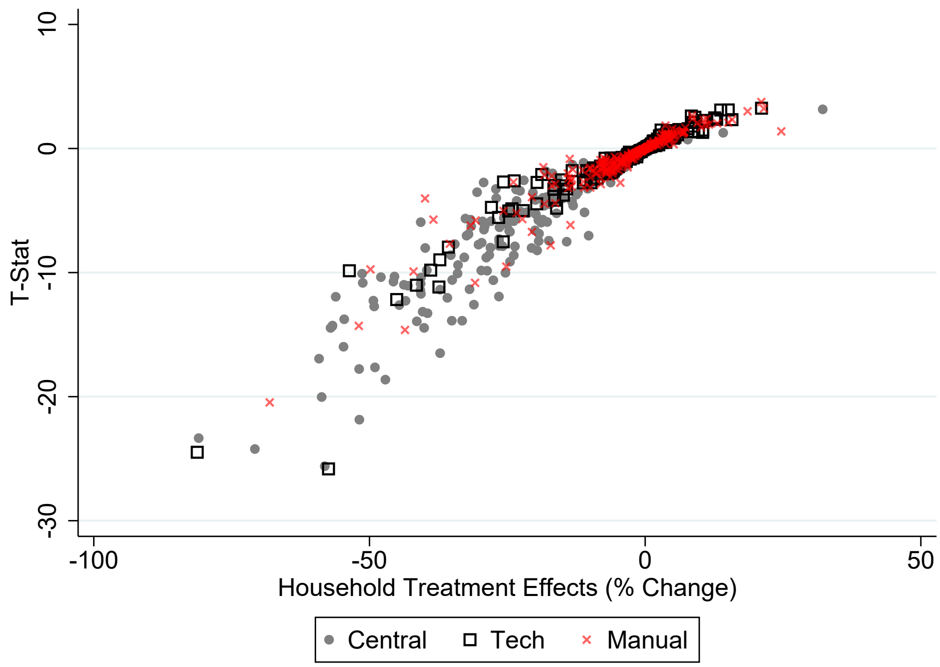
(c) Manual Program

Notes: The reported results summarize the distribution of estimated household-level treatment effects obtained from estimating specification (9), permitting different treatment effects by event type: morning, evening, and high evening. We present the marginal effects to be a percentage change in consumption using the transformation $100 \times (exp(\widehat{\beta}_i) - 1)$.

D.4 Consistency of Demand Response

In this section, we expand upon the household-level results to investigate the consistency of responses to events by program. This builds off of the discussion in Section 7.1 and Figure 6 in particular.

Figure B11. Household-Level Estimated Treatment Effects and T-Statistics by Program



Notes. The reported results plot the estimated household-level treatment effects and corresponding t-statistic obtained from estimating specification (9). We present the marginal effects to be a percentage change in consumption using the transformation $100 \times (exp(\widehat{\beta}_i) - 1)$.

Figure B11 plots the estimated household event responses and the corresponding t-statistics to better understand the significance of the estimates. The majority of the Manual and Tech values are closely distributed near zero in both the estimated event response and t-statistics. Only 20% and 18% of the household-level estimated treatment effects are negative and significant at the 5% level for the Tech and Manual programs, respectively.[36] This suggests that only a small subset of high-performing households in these programs were consistent responders to peak events over our 17-

[36] There is a small subset of households with positive estimated treatment effects. These estimated effects are systematically statistically insignificant, with only 3% being positive and statistically significant.

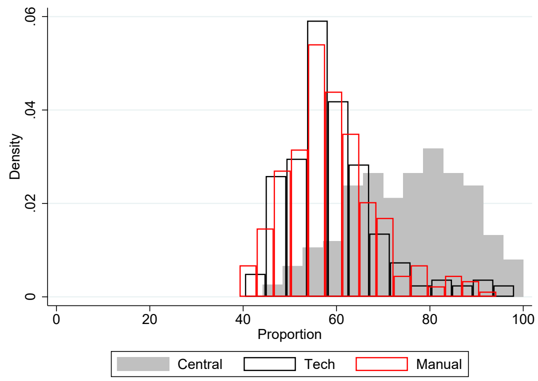
month treatment period. Figure B11 reinforces our finding that the Central program has large and precisely estimated responses to peak events. 80% of the Central household-level estimated treatment effects are negative and significant, indicating that the majority of Central households provide large and consistent reductions in consumption during peak events.

In addition to having estimates for household-level treatment effects, we observe each household's consumption relative to their baseline for *each* event. As described in Section 3.2, the baselines were calculated based on a household's average consumption during the relevant event time window over the last five weekdays prior to the event, excluding event days. For each household, we calculate the proportion of events where their observed percentage reduction in consumption relative to their baseline was greater than 0%, 10%, 20%, and 30%. It is important to recognize that this measure will have some noise. This is particularly relevant for the comparison of consumption relative to the baseline being greater than 0%. Even if a household did not adjust its behavior to respond to the event, natural variation would induce consumption to fall below the baseline for a certain proportion of events. If this randomness is normally distributed around the baseline, the proportion would be 50%. While this makes interpreting the results for this measure more nuanced, as we will describe below, it is still valuable because we can compare the relative performance across programs.

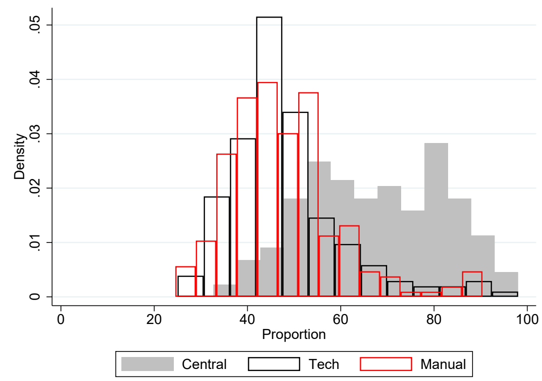
Figure B12 presents, at the household-level, the proportion of events where the observed percentage reduction in consumption relative to the baseline was greater than 0%, 10%, 20%, and 30% by program. Focusing initially on the results for greater than 0% threshold, the Tech and Manual groups have similar distributions and are slightly rightward shifted above 50%. As noted above, noise in this measure could result in values near 50% being viewed as having a minimal reduction relative to the baseline for the greater than 0% threshold. The fact that the Tech and Manual programs have distributions situated just to the right of 50%, with a small subset of households exceeding 70%, is consistent with our household-level regression results. In contrast, the majority of Central households have proportions in excess of 70% of event days being greater than 0%. This continues to suggest that Central households consistently reduce their consumption during events.

Figure B12 shows that as the threshold increases to greater than 10%, 20%, and 30% reductions relative to the baseline, the proportion of events where households achieve reductions above this threshold declines across all programs. However, the proportion of events where Tech and Manual achieve reductions above this threshold

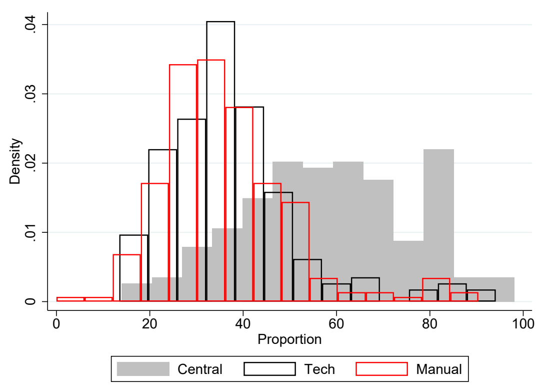
Figure B12. Household-Level Proportion of Responses with Reductions Relative to Baseline Greater Than 0%, 10%, 20%, and 30%



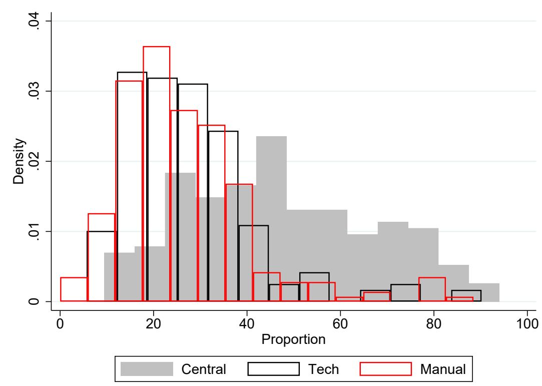
(a) Reduction greater than 0



(b) Reduction greater than 10%



(c) Reduction greater than 20%



(d) Reduction greater than 30%

Notes: Figure plots at the household-level the proportion of events where their observed percentage reduction in consumption relative to their baseline was greater than 0%, 10%, 20%, and 30% by program over the period February 1, 2022 - June 30, 2023.

declines at a much faster rate than Central. For example, the majority of Central households achieve reductions greater than 20% in at least 50% of events. In contrast, this number is in the range of 20% - 30% of events for the majority of Tech and Manual households, with a small subset of higher performers who consistently reduce their consumption above this threshold in the majority of events.

Taken together, these results suggest that the Central program was able to achieve a high and consistent demand reductions for the majority of events over the 17-months they were exposed to peak events. This differs from the Tech and Manual programs that achieved more modest demand reductions for a smaller proportion of events.

E End of Experiment Survey

E.1 Survey Details

The following is text from a voluntary, end-of-experiment survey sent out to participants in the Central, Tech, and Manual programs via email, in mid-June, 2023.

Survey instructions:
"This short survey is designed to hear about your experience in the Peak Rewards Trial through [APP NAME]. All homes had a different experience, and we want to hear about yours.

We appreciate the time and thought you put into this survey.

Properly completed surveys will be rewarded with $20 on bill credit as a token of our gratitude."

Survey questions used in our analysis:

"What is your approximate household income?

- Less than $50k per year
- $50-99k per year
- $100-149k per year
- $150-200k per year
- Over $200k per year
- Don't know/Rather not say"

"For the events you noticed, how often was it worth your time to participate by attempting to reduce your electricity consumption?"

- Never
- Sometimes
- About half the time
- Most of the time
- Always
- Don't know/Not Applicable"

E.2 Survey Response

In this section, we compare the observable characteristics of participants who filled out the end-of-experiment survey to those who did not. The Table below recreates our balance Table A2 (using pre-treatment data), but separates the results by whether or not the household responded to the exit survey. The p-value corresponds to a difference in means test.

Table C5 demonstrates that the non-respondents had larger cumulative consumption during the pre-treatment period. Non-respondents also were more likely to have an electric vehicle. All other characteristics are similar across the two groups.

Table C5. Balance by Exit Survey Response (Pre-Treatment Data)

	Yes	No	p-value
Cumul. kWh			
Winter	5,229	5,892	0.04**
	(2,810)	(3,199)	
Spring	3,675	4,167	0.02**
	(1,788)	(2,104)	
Summer	2,614	3,155	0.01***
	(1,492)	(2,142)	
Fall	3,528	4,012	0.02**
	(1,721)	(2,016)	
Load Factor			
Winter	24.96	25.21	0.78
	(8.61)	(8.89)	
Spring	19.82	19.98	0.83
	(6.64)	(6.93)	
Summer	16.56	17.21	0.37
	(6.87)	(6.76)	
Fall	18.66	19.12	0.51
	(5.91)	(6.72)	
Electric Vehicle	0.24	0.32	0.05**
	(0.42)	(0.47)	
BaseBoard Heating	0.66	0.64	0.71
	(0.47)	(0.48)	
Air Conditioning	0.47	0.51	0.32
	(0.50)	(0.50)	
Electric Hot Water	0.72	0.70	0.56
	(0.45)	(0.46)	
House Duplex	0.83	0.82	0.71
	(0.37)	(0.38)	
Median Income	86,377	87,434	0.55
	(19,853)	(19,835)	
Observations	429	174	

Notes. This table compares pre-treatment average values by whether or not the household participated in the exit survey. Parentheses contain the standard deviations. Cumul. kWh and Load Factor represents the cumulative household-level consumption and load factor by season. Electric Vehicle, Baseboard Heating, Air Conditioning, and Electric Hot Water are indicator variables denoting the presence of each device. House/Duplex is a indicator variable if the home type is a single-family home or duplex. Median Income reports the median household-level income of the Census Dissemination Area where the household is located. p-value from a difference in means test across the two groups. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.