CrowdFlex report: Utilisation Trial

Summer 2025

Submitted: December 2025

NESO
National Energy
System Operator

Executive Summary

CrowdFlex is a NESO led innovation project, funded by Ofgem's Strategic Innovation Fund (SIF), which is investigating the potential of domestic flexibility to help operate the grid. CrowdFlex is aiming to establish domestic flexibility as a reliable energy and grid management resource by identifying the technology capability, understanding the statistical nature of flexibility and aligning NESO and DSO requirements. Through large-scale randomised control consumer trials, CrowdFlex is collecting data to develop demand and consumer flexibility prediction models using common APIs.

NESO is delivering CrowdFlex with a consortium of industry partners: OVO, Ohme, Centre for Net Zero, ERM, AWS, National Grid Electricity Distribution, Scottish and Southern Electricity Networks, and supported by Smart Grid Consultancy, CGI, Smith Institute and Centre for Sustainable Energy.

The *Summer 2025 CrowdFlex Utilisation Payments Trial* examined how different incentive structures and notice periods influenced household participation and demand response during flexibility events. Conducted between July and October 2025, the trial involved more than 70,000 OVO Energy customers across 60 events and provided one of the largest evidence bases to date on residential flexibility in Great Britain.

Building on earlier *CrowdFlex* trials, this study showed that a range of incentives can shift household electricity use. It also highlighted that the way incentives are framed and communicated - and how much notice households receive - affects participation and response. The findings offer practical lessons for designing effective flexibility schemes that support the UK's transition to a low-carbon energy system.

Trial design

Participants were randomly allocated to one of five treatment groups and to one of three notice periods before each event, or to a randomised control group. This randomised controlled trial design ensured that differences in behaviour could be attributed directly to the interventions tested. The trial primarily focused on turn-up events, given its timing during the summer and the importance of addressing demand turn-up as a key evidence gap for the final CrowdFlex utilisation payments trial. However, we also gathered evidence on turn-down events – specifically, on how turn-down performance varied with notice period – from the second treatment group listed below, which experienced one turn-down event per week in addition to the four weekly turn-up events common to all treatment groups.

Incentive-based treatment groups:

- **£0.25/kWh (Direct Payment):** a cash reward for every kilowatt-hour of increased or reduced consumption during events (given the unit rates at the time of the trial, this reward made incremental electricity consumption nearly free).
- **£0.25/kWh + Down Events:** the same direct reward, with an additional weekly "turn-down" event to test whether exposure to both directions of flexibility affected engagement or performance.

- **£0.20/kWh + Community Pot:** a slightly lower individual payment, combined with contributions to a shared community reward fund distributed to top-performing postcode groups.
- **Free Electricity (Volumetric):** a reward of free electricity hours proportional to the volume of additional demand provided during events.
- **Free Electricity (Consistency):** a reward of free hours of electricity based on consistent participation across events, regardless of the amount shifted.

Notice periods: Participants were notified of events either day-ahead (at 18:00 the previous day), five hours before, or two hours before. From mid-trial, randomly selected customers also received SMS notifications and reminders, allowing assessment of how real-time communications influence engagement.

Key findings

1. **All incentive structures effectively shifted demand.**
   Every treatment group showed a statistically significant change in consumption during events compared with the control group. Direct monetary rewards generated the largest increase in electricity use during turn-up events (around 20% higher than the control), but alternative incentives performed nearly as well.

2. **In-kind rewards delivered similar flexibility at lower cost.**
   Although designed to offer comparable headline value, the realised cost of in-kind incentives was much lower. Free-electricity groups achieved comparable demand response to the £0.25/kWh cash group but at roughly one-third of the cost per unit of flexibility delivered. This suggests that participants may have responded as much to the perceived opportunity of a reward as to its direct monetary value or incentive mechanism. Community-based schemes were also effective in shifting demand but did not significantly change engagement.

3. **Opt-in reduced but did not eliminate baselining inaccuracy.**
   There are many potential baselining methods for remunerating customers and settling flexibility between system operators and demand response providers. We found that the simplest baseline produced, on average, similar estimates of delivered flexibility to those obtained by comparing treatment groups with our randomised control group. "Clipping" wrong-direction responses for individual customers reduced the accuracy of measured flexibility. However, this inaccuracy was substantially reduced when responses were aggregated only across customers who had actively opted in, rather than across the full treatment group.

4. **Longer notice periods increased response.**
   The length of notice had a clear influence on turn-up and turn-down response. For turn-down, day-ahead notified customers had statistically significant response, while the other groups did not; for turn-up, the day-ahead and 5-hour notice period groups had greater turn-up than the 2-hour notice group. Day-ahead notifications produced the highest opt-in rates (36% for turn-up and 41% for turn-down), compared with 25–30% under two-hour notice.

5. **SMS reminders increased response.**
SMS notifications and reminders proved effective: messages increased participation by approximately 3% and consumption during turn-up events by a further 3-4 percentage points (a 15-20% increase on top of the 16-17% turn-up in the absence of notifications). The impact was perhaps most pronounced for shorter-notice events, suggesting that timely reminders can make up for less notice.

6. **Frequent or mixed event exposure slightly reduced performance.**
Customers who experienced both turn-up and turn-down events showed a small decline in turn-up compared with those facing turn-up events alone. This modest effect - about 2–3% lower - may have reflected mild behavioural fatigue, given that participants were asked to respond to an extra event per week, particularly one that was in a different direction (turn-down) to the other four turn-up events.

7. **Flexibility was widespread but strongest among households with low-carbon technologies.**
The interventions generated consistent positive responses across demographic and customer subgroups, indicating broad appeal. The largest relative increases were observed among households with low-carbon technologies - such as electric vehicles, heat pumps, or batteries - whose response was nearly 50% higher than that of non-LCT households. This highlights significant potential for linking domestic flexibility with wider electrification and automation initiatives to deliver scalable demand response.

The Summer 2025 trial demonstrated that households provided meaningful flexibility across a range of incentive models. As responses were similar across incentive groups, we hypothesise that households were motivated more by the chance to take part and the perceived value of the reward than by the exact monetary value or the incentive structure. The findings suggest that combining in-kind incentives with clear, timely communications and intuitive event design can deliver cost-effective, scalable flexibility.

Executive summary tables

Executive summary table 1: Percentage difference in demand shift by trial arm

Trial arm	% difference (95% confidence interval)	
	Turn-up	**Turn-down**
£0.25/kWh	19.55%*** (17.88%, 21.22%)	
£0.25/kWh + down events	16.29%*** (14.66%, 17.92%)	-1.42%* (-2.68%, -0.16%)
£0.20/kWh + community pot	18.08%*** (16.44%, 19.72%)	
Free hours (volumetric)	16.39%*** (14.92%, 17.85%)	
Free hours (consistency)	17.98%*** (16.45%, 19.51%)	

Executive summary table 2: Percentage difference in demand shift by notice period

Trial arm	% difference (95% confidence interval)	
Control	Turn-up	Turn-down
2 hours	14.94%*** (13.63% ,16.24%)	-0.23% (-1.97%, 1.51%)
5 hours	19.49%*** (18.11%, 20.87%)	-1.08% (-2.92%, 0.76%)
Day ahead	18.55%*** (17.21%, 19.90%)	-2.95%** (-4.73%, -1.17%)

Note: *The Percentage difference refers to the coefficient (treatment effect) vs. the Control group. Significance levels: * P<0.05, ** P<0.01, *** P<0.001.*

1. Introduction

1.1 Background and rationale

As the UK moves towards a low-carbon energy system, increased renewable generation, such as wind and solar, brings greater variability in supply. Managing this variability requires new ways to align electricity supply and demand to manage grid constraints. Flexible demand from residential customers offers one way to meet this challenge, supporting efficient grid operation, reducing system costs, and contributing to decarbonisation objectives.

CrowdFlex is an innovation project about how to unlock that potential. It is led by the National Energy System Operator (NESO) in partnership with OVO Energy, Ohme, Centre for Net Zero, ERM, AWS, National Grid Electricity Distribution, and Scottish and Southern Electricity Networks, with support from Smart Grid Consultancy, CGI, Smith Institute, and the Centre for Sustainable Energy. Funded by the Ofgem Strategic Innovation Fund, the programme combines model development with rigorous randomised and matched controlled trials to test interventions that enhance demand flexibility.

Previous research, including the CrowdFlex's summer 2024 and winter 2024-25 utilisation payments trials, demonstrated that financial incentives can influence household electricity consumption patterns. However, questions remained about how the delivery of incentives and notice periods shaped response, particularly for turn-up events, where further evidence was needed to build upon initial winter findings.

The Summer 2025 CrowdFlex Utilisation Trial, involving over 70,000 participants, was designed to address these evidence gaps. It systematically tested how different incentive structures - ranging from direct payments to community rewards and in-kind "free electricity offers" - and varying notice periods affect participation and demand response. This report presents findings from the trial, which ran 60 events between from 14 July 2025 through 5 October 2025, and offers novel empirical evidence to inform the design of future consumer flexibility programmes in Great Britain.

1.2 Objectives

The primary aim of the trial was to generate robust evidence regarding how households respond to different configurations of demand-side flexibility events. Specifically, the trial examined:

- Whether different types of incentives – such as direct payments per kilowatt-hour shifted, community reward structures, or in-kind benefits like free electricity – resulted in varying levels of electricity demand change during flexibility events.
- Whether the length of advance notice affected households' ability to participate in, and ability to respond to, trial events.

1.3 Research questions and hypotheses

The trial investigated several research questions.

1.3.1 Primary research questions:

1. How does the type and mechanism of delivering incentives influence electricity consumption during turn-up events?
2. How does the length of the notice period affect electricity consumption during both turn-down and turn-up events?

1.3.2 Primary hypotheses:

1. Direct financial incentives (£ per kilowatt-hour) will generate a greater shift in consumption during turn-up events compared to alternative incentive mechanisms.
2. Alternative incentive structures, such as community pots or free electricity offers, will produce measurable but likely smaller shifts in consumption than direct price-based incentives.
3. Shorter notice periods will result in lower participation and smaller consumption shifts compared to longer notice periods.

1.3.3 Secondary research question

As above, but for participation (opt-in rate) instead of performance.

1.3.4 Secondary research hypotheses:

As above, but for participation instead of performance.

1.3.5 Exploratory research questions

Exploratory analysis examined:

- Consumption during hours before events (after event notifications have been received) and after them, to explore demand creation/destruction vs displacement.
- Subgroup analysis.
- Impact of SMS reminders on electricity demand during turn-up and turn-down events.
- Accuracy of P376 baseline.

2. Methodology

2.1 Trial design

Participants were randomly assigned to one of several trial arms designed to test different combinations of incentives and event types. The primary trial arms are detailed in Table 1 and Table 2.

Table 1: Treatment group description

Treatment group	Core mechanism	Key customer experience/reward
Direct £0.25/kW	Direct £/kWh payment for increased and decreased demand was offered.	Customers earned a specific amount of money (£0.25) for every unit (kWh) of electricity demand they shifted up.
Direct £0.25/kWh + Down events	Direct £/kWh payment for increased demand, plus specific turn-down events were included.	Customers earned a specific amount of money (£0.25) for every unit (kWh) of electricity demand they shifted up, and participated in turn-down events.
£/kWh + Community Pot	Participants contributed a small portion of their reward to a shared pot.	Customers earned a direct reward (£0.20/kWh) plus the chance to win a large bonus if their community was a top performer.
Free Electricity (Volumetric)	Reward was based on the volume (kWh) of energy flexibility delivered.	Customers earned free hours of electricity for every kWh of demand they shifted during events.
Free Electricity (Consistency)	Reward was based on the consistency of participation in events.	Customers earned free hours of electricity based on how consistently they participated in events, regardless of the amount of shift.

Table 2: Summer utilisation trial incentive-type trial arms

	Direct £0.25/kW	Direct £0.25/kWh + Down events	£/kWh + Community Pot	Free Electricity (Volumetric)	Free Electricity (Consistency	Control
Turn-up incentive	£0.25/kWh	£0.25/kWh	£0.20/kWh + ~£0.05/kWh put into community pot	Free hours based on kWh delivered	Free hours based on consistency of response	N/A
Turn-down incentive	N/A	£0.50/kWh	N/A	N/A	N/A	

The trial primarily focused on turn-up events, given its timing during the summer and the importance of addressing demand turn-up as a key evidence gap for the final CrowdFlex utilisation payments trial. However, we also gathered some evidence on turn-down events – specifically, on how turn-down performance varied with notice period – from the second

treatment group, which experienced one turn-down event per week in addition to the four weekly turn-up events common to all treatment groups.

For both free electricity groups (volumetric and consistency), customers were able to redeem their earned free electricity between 12:00 and 14:00 on the Sunday immediately following the email confirming their successful participation. This timing was designed to complement OVO's existing Free Electricity proposition for 'Beyond' customers (an OVO rewards programme) without creating conflicts. The goal was to provide an overall average reward for turn-up similar in cost to the turn-up rewards for the other three treatment groups (i.e., approximately £0.25/kWh). This ended up not coming to pass – as we will discuss, the free electricity groups had much lower total remuneration than the direct payment group, despite having similar turn-up. This discrepancy likely reflects lower-than-expected electricity use during the Sunday redemption window, combined with lower overall consistency in event participation than was assumed in the original reward cost modelling. More detail on the treatment groups is provided in the appendix.

2.2 Randomisation

A strength of this trial lies in its randomised controlled design. This approach means that differences observed between participant groups can be attributed to the interventions under study, rather than to pre-existing differences among participants or external influences confounding identification of impacts. By randomly allocating participants to trial arms, we created groups that were similar on both measured and unmeasured characteristics, providing a robust foundation for drawing conclusions about the effectiveness of different incentives and notice periods in influencing electricity consumption behaviour.

Randomisation in the trial was conducted in two stages, with stratification by Grid Supply Point (GSP) applied throughout. Stratification by GSP ensured balanced representation of geographic areas across trial arms, reflecting potential regional variations in grid conditions and customer characteristics.

2.2.1 Assignment to incentive groups

All eligible participants, including newly recruited customers and individuals who had participated in the preceding Winter utilisation trial, were randomly assigned to one of six trial arms: the control group or one of five incentive treatment groups.

Stratification by GSP was applied during this allocation to help achieve comparable groups across different regions. For participants returning from previous trials, randomisation was used to maintain balanced representation across groups, except for customers who had been in the control group during the previous Winter trial. These individuals were intentionally excluded from the new control group and randomised into a treatment group to ensure they experienced an active intervention in the current trial.

2.2.2 Assignment to notice periods

Within each incentive group, participants were further randomised to one of three possible notice periods for each event:

- Short notice (two hours in advance)
- Medium notice (five hours in advance)
- Long notice (day-ahead notice at 18:00 the previous day)

Stratification by GSP was again applied at this stage to distribute notice periods evenly across different regions.

All participants were required to opt in individually for each event, allowing the trial to capture deliberate engagement decisions and providing further insight into customers' willingness to participate under varying conditions.

2.3 Balance checks and trial arm comparability

Before analysing results from the summer utilisation trial, OVO performed detailed balance checks to ensure that customers assigned to different trial groups were comparable. These checks help confirm that any differences in outcomes are due to the trial itself, rather than differences in the types of customers in each group.

The checks assessed a range of customer characteristics. These included region, measured as the electricity Grid Supply Point (GSP) area, and tenure with OVO, recorded in years as a continuous variable. OVO also examined Estimated Annual Consumption (EAC), which reflects a customer's expected electricity use over a year in kilowatt-hours. Additionally, peak electricity consumption was checked, defined as the total amount of electricity used during peak hours between May and July 2025, and non-peak electricity consumption, recorded as the total electricity used during off-peak hours over the same period.

The analysis used pairwise t-tests. No significant differences were found between trial groups for GSP regions, peak consumption, or non-peak consumption. There were some differences observed in tenure and in estimated annual consumption (EAC) when comparing specific trial groups to the control group, though these differences became smaller and mostly non-significant after adjusting for multiple testing. The slight differences identified were mainly attributed to the winter control group, which was reassigned into the summer trial's treatment groups for this phase of the project.

Overall, the balance checks showed that the trial groups were broadly similar across the variables measured. This provides confidence that the results of the trial can be reliably interpreted as the effect of the interventions tested, rather than underlying differences between customer groups.

Balance check results are located in the appendix.

2.4 Outcome measures

The trial's primary outcome was the total electricity consumption measured in kilowatt-hours (kWh) per half-hour during each event, providing a direct measure of demand shifted in response to the trial interventions. All analysis results are then multiplied by 2 to yield results in kW (kWh per hour) for improve interpretability.
Secondary outcomes included the rate at which participants opted into events, serving as a proxy for customer engagement and willingness to participate in demand-side flexibility.

2.5 Implementation and operation

The trial was conducted between 14 July 2025 and 5 October 2025. A total of 60 events were scheduled during this period, comprising 48 turn-up events and 12 turn-down events. Events were scheduled to begin between 12:00 and 21:00 to align with periods of potential grid stress or surplus generation, while avoiding unsociable hours. Each event lasted one hour.

The scheduling of events in the trial was designed to align with periods of operational significance for the electricity system, reflecting times when flexibility could deliver the greatest value in supporting grid management and integrating renewable generation.

The 48 turn-up events were distributed across five start times between 12:00 and 20:00. The highest number of events were scheduled at 12:00 (14 events) and 13:00 (12 events), with fewer events taking place later in the afternoon and evening, including three events starting at 20:00. This pattern reflects a deliberate emphasis on testing flexibility during daylight hours, when solar generation is typically high, as well as exploring the feasibility of engaging additional demand during later periods when network conditions may vary.

The 12 turn-down events were scheduled across a broader range of times. The largest concentration occurred at 18:00 (four events) and 17:00 (three events), aligning with periods when electricity demand on the grid often increases in the early evening. Single turn-down events were also held at 12:00, 14:00, 19:00, and 20:00.

This scheduling approach was intended to ensure that the trial captured a representative range of grid conditions, allowing assessment of how time of day interacts with notice periods and incentive types in influencing the delivery of demand-side flexibility. The event schedule was also constructed to avoid overlap with existing flexibility schemes or conflicting grid requirements.

For the event on 17 September 2025, notifications were intentionally not dispatched to customers in the two Scottish GSPs. This was a planned test designed to simulate the process of cancelling an event notification from a Distribution Network Operator (DNO) perspective. The aim was to assess the technical feasibility and operational readiness of withholding or cancelling notifications in response to real-time grid requirements.

2.6 Trial opt-in mechanism

In this trial, participation in each flexibility event was contingent on customers explicitly opting in on an event-by-event basis. Unlike the previous CrowdFlex utilisation payments trials, where participation was assumed by default, the current trial design required active customer opt-in for each individual event.

The opt-in was communicated via the event notification email, which contained a single opt-in button. Clicking this button confirmed a customer's opt-in to the event.
Customers were permitted to opt in up to 60 seconds after the scheduled event start time. This brief extension accounted for potential system delays that might occur if customers attempted to opt in at the last minute. Once a customer opted in for a given event, this decision was final; no mechanism existed for customers to subsequently opt out of participation.

This opt-in mechanism was designed to ensure that any observed participation reflected deliberate customer engagement, with the aim of reducing remuneration to customers who had not actively participated.

2.7 Sample sizes

Table 3: Sample size within trials arms

Trial arm	# start	# end	% dropout
Control	12,590	11,778	6.40%
£0.25/kWh	12,605	11,574	8.20%
£0.25/kWh + Down events	14,037	12,892	8.20%
£0.20/kWh + community pot	12,592	11,568	8.10%
Free electricity (volume)	12,601	11,587	8.00%
Free electricity (consistency)	12,593	11,538	8.40%
Total	77,018	70,937	7.90%

Table 4: Notice period trial arms trials arms

Trial arm	# start	# end	% dropout
Control	12,590	11,778	6.40%
2 hours	21,480	19,770	8.00%
5 hours	21,468	19,728	8.10%
Day ahead	21,480	19,661	8.50%
Total	77,018	70,937	7.75%

2.8 Introduction of SMS event notifications and SMS reminders

A primary focus of this trial was to assess the impact of different notice periods on electricity consumption during events, encompassing both turn-down and turn-up scenarios. This primary research question guided the design and analytical strategy.

During the trial, OVO introduced supplementary communications to measure participant engagement and address concerns about event recall.

- **Notifications:** In addition to all customers receiving event notifications via emails, OVO Energy introduced SMS event notifications for a selected random sample of customers on 16 August. These notifications came at the *same time* as the email – for this reason, the notification send time was 2 hours before the event for the 2-hour-notice customers, 5 hours for the 5-hour-notice customers, and day-ahead for the day-ahead-notice customers.
- **Reminders as well:** OVO introduced SMS reminders for the same sample from 17 September. Reminder SMSs were sent a half-hour before event start regardless of which notice-period group customers had been allocated to.

Importantly, OVO sent these SMSs to only half of the customers in the sample who had a valid phone number (referred to as eligible customers). The recipient vs non-recipient split was determined via randomisation. This randomisation enables identification of the causal impact of the SMSs on consumption during events.

2.9 Analytical methodology

2.9.1 Main analysis: Effect of notice period

For the main analyses assessing the impact of incentives and notice periods, all customers were pooled. This included customers without a phone number, those eligible for SMS notifications and reminders who were not selected, and those selected to receive SMS notifications and reminders.

The analysis encompassed all events throughout the trial period.

The introduction of SMS reminders did not bias the estimation of the effect of notice or incentive on consumption or opt-in. This was ensured because reminders were randomised evenly across all treatment arms (i.e., treatment and notice periods), meaning their presence did not skew the comparison of treatments to each other.

2.9.2 Exploratory analysis: Effect of SMS notifications and reminders

A separate, exploratory analysis was conducted to specifically evaluate the additional effect of SMS reminders. This analysis was restricted to eligible customers and only included events after 16 August through 16 September (for analysis of the SMS notifications) or after 17 September (for analysis of the SMS reminders).

The analysis proceeded in two steps:
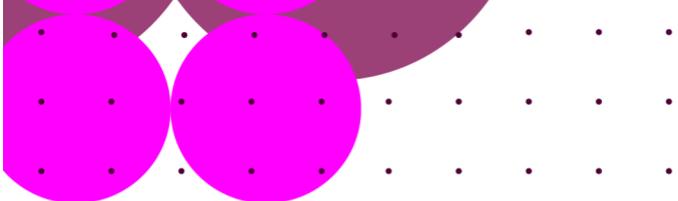
- Initially, the average treatment effect of reminders among the eligible population was estimated by pooling all notice periods and comparing those who received a reminder against those who did not.
- This was then refined by comparing reminders versus no reminders within each specific notice period (short, medium, long). This allowed for an investigation into whether reminders had a differential impact across varying notice lengths.

3. Results

3.1 Primary analysis

3.1.1 Effect of interventions on consumption during turn-up events

Analysis of the trial's turn-up events, where participants were encouraged to increase their electricity usage, revealed a clear and positive influence of all tested incentive structures on consumption during turn-up events. The Control group, which received no monetary incentive nor communications about events, had an average demand of 0.304 kW during these events. All five intervention arms had higher demand (all P-values <0.001).

The intervention where participants received £0.25/kWh for increased consumption proved to be the most effective overall, achieving the highest average demand of 0.366 kW. This represented an increase of 19.55% over the Control group. Across the incentive structures, the increases from the Control group ranged from 16.29% to 19.55%.

Note: *The Percentage difference refers to the coefficient (treatment effect) vs. the Control group. Significance levels: * P<0.05, ** P<0.01, *** P<0.001.*

Figure 1: Average kW demand across trial arms during turn-up events
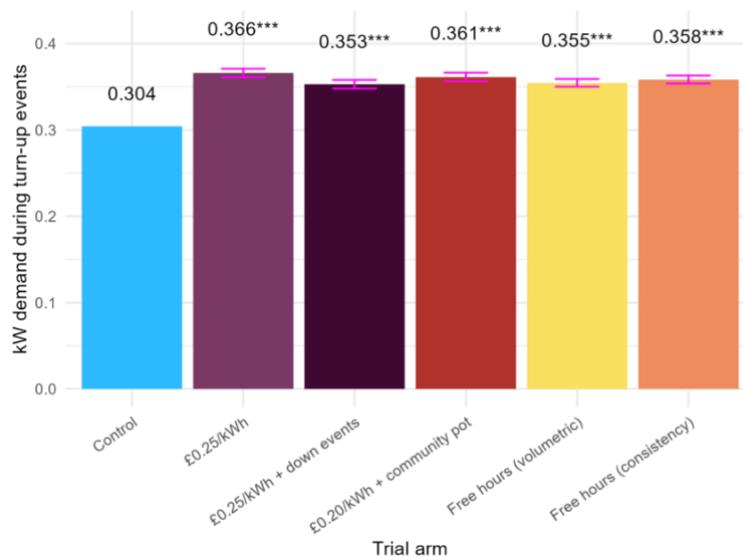
Figure 2: Regression coefficients for turn-up event kW demand

The comparison of the £0.25/kWh arm (only exposed to turn-up events) with the £0.25/kWh + Down Events arm (exposed to the same number turn-up events, plus one additional turn-down event, per week) provided a novel test of whether the extra turn-down event changed turn-up performance. This could be due to behavioural fatigue, habit formation, or even complementarities in participating in opposite-direction events.

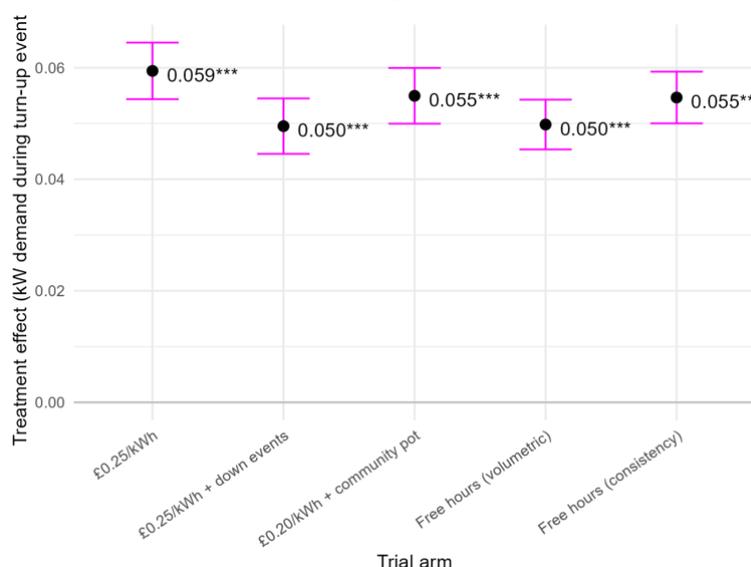The results suggested that fatigue outweighed any habit formation effects participants exposed to both types of events had slightly lower demand during the turn-up events than those exposed only to turn-up events. The £0.25/kWh + Down Events group recorded an average demand of 0.353 kW, representing a 2.67% reduction in performance compared to the £0.25/kWh up-only group average demand (0.366 kW). This difference was statistically significant (P=0.0016), suggesting that the extra event did introduce a minor but detectable dampening effect on turn-up performance.

A similar, statistically significant reduction was observed between the Free Hours (Volumetric) group and the £0.25/kWh group (-2.63%). No other statistically significant differences were observed across treatment groups.

Table 5: Pairwise comparison of average turn-up demand across treatment arms

Baseline Arm	Compared arm	Baseline avg.	Compared avg.	Coefficient	% difference
£0.25/kWh	£0.25/kWh + Down Events	0.366	0.353	-0.010**	-2.67%
£0.25/kWh	£0.20/kWh + Community Pot	0.366	0.361	-0.004	-1.23%
£0.25/kWh	Free Hours (Volumetric)	0.366	0.355	-0.010***	-2.63%
£0.25/kWh	Free Hours (Consistency)	0.366	0.358	-0.005	-1.24%

Baseline Arm	Compared arm	Baseline avg.	Compared avg.	Coefficient	% difference
£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.353	0.361	0.005.	1.54%
£0.25/kWh + Down Events	Free Hours (Volumetric)	0.353	0.355	0.000	0.08%
£0.25/kWh + Down Events	Free Hours (Consistency)	0.353	0.358	0.005.	1.44%
£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.361	0.355	-0.005.	-1.44%
£0.20/kWh + Community Pot	Free Hours (Consistency)	0.361	0.358	0.000	-0.07%
Free Hours (Volumetric)	Free Hours (Consistency)	0.355	0.358	0.005.	1.37%

3.1.2 Effect of interventions on consumption during turn-down events

The treatment group successfully reduced electricity demand during the 12 turn-down events, achieving a statistically significant 1.43% reduction over the Control group (P=0.0263). This result confirms that the incentive mechanism was effective at driving measurable change in participant behaviour.

Figure 3: Average kW demand across trial arms during turn-down events
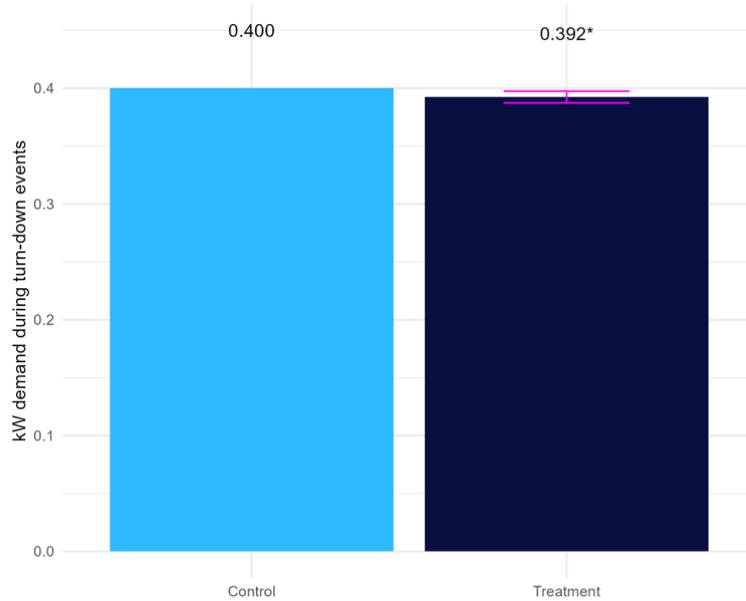
Figure 4: Regression coefficients for turn-down event kW demand
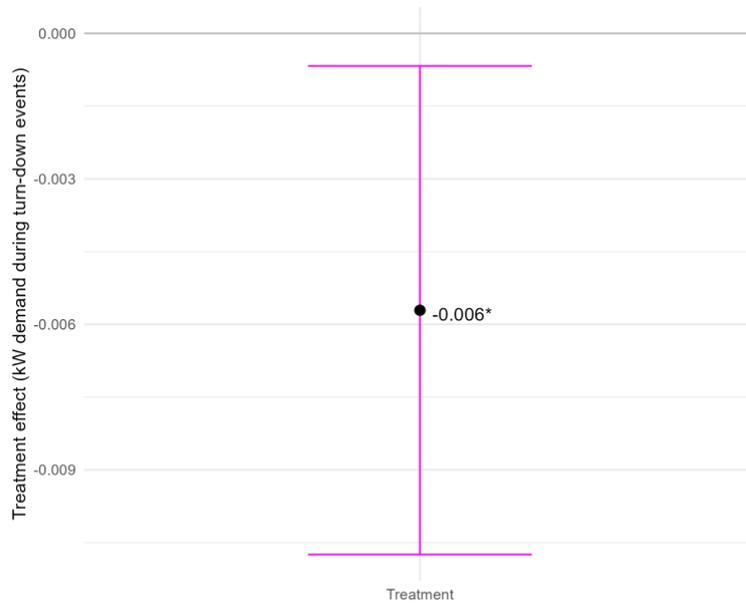
3.1.3 Effect of notice period on consumption during turn-up and turn-down events

The other primary analysis concerned the impact of shorter / longer notice periods. By "notice period", we mean the time between event notification and the event start. We found that there was a small dampening in turn-up for 2 hours' notice, but 5 hours and day-ahead notice had similar levels of turn-up. For turn-down, day-ahead notified customers reduced demand whereas both 2- and 5-hours' notice customers did not.

3.1.3.1 Effect of notice period on consumption during turn-up events

For turn-up events (increasing consumption), all notice period groups had a significant increase in consumption compared to the Control group (receiving no intervention), which, as mentioned above, averaged 0.304 kW demand.

- The greatest response came from the 5 hours' notice period, which had an average demand of 0.365 kW (a 19.49% increase over the Control group) and the DA notice period group, which had an average demand of 0.361 kW (an 18.55% increase). Note the 5-hours and day-ahead groups' demand were *not* statistically significantly different from each other.
- The shortest notice period group, 2 hours' notice, resulted in a significant response, reaching 0.350 kW (14.94% increase). This level of demand *was* statistically significantly different from both longer notice-period groups.

In summary, the 2-hour group had slightly lower turn-up than the other two groups.

Figure 5: Average kW demand in turn-up events by notice period
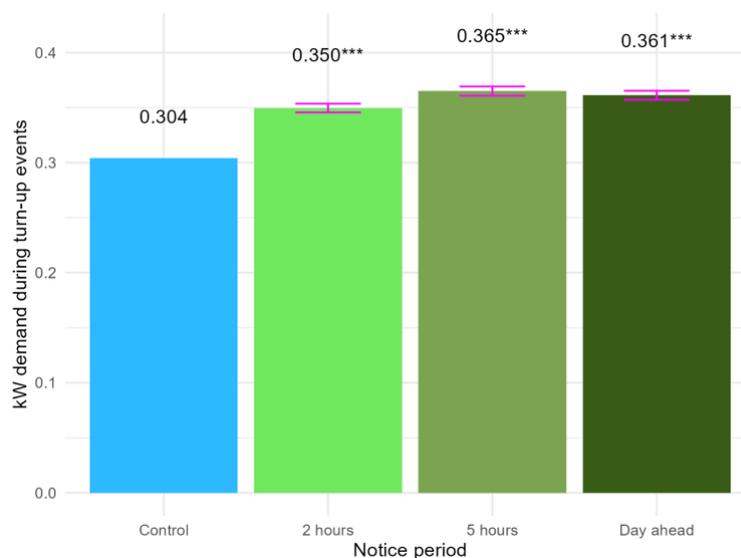
Figure 6: Regression coefficients for turn-up event kW demand by Notice Period
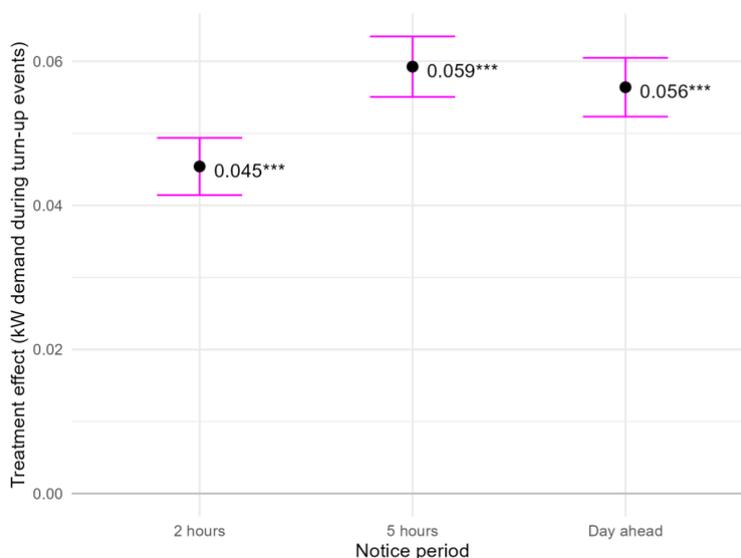
3.1.3.2 Effect of notice period on consumption during turn-down events

We used the single treatment group that faced the extra turn-down event once per week to investigate the impact of notice period on *turn-down*, as well.

The Control group for these events averaged demand of 0.400 kW. The 2 hours and 5 hours' notice groups had demand that was not statistically significantly different from the Control group, with average demand of 0.396 kW for both (P=0.7939 and P=0.2507, respectively). Note that the control-variable adjusted coefficients show *suggestive* evidence of slightly more turn-down in the 5-hours' notice group than the 2-hours' notice group, but we cannot rule out no difference in demand to the Control group.

In contrast, the DA notice period demonstrated a statistically significant reduction, showing a 2.95% decrease in average demand, to 0.385 kW (P=0.0012).

These findings suggested that for turn-down during this trial, customers need a longer preparation window to achieve a significant reduction, and that 2 and even 5 hours were not sufficient. Note that this result is in tension with CrowdFlex's winter utilisation payments trial, where customers had meaningful turn-down in <4-hour notice events, and it was not meaningfully different from their turn-down during longer events.

Figure 7: Average kW demand in turn-down events by notice period
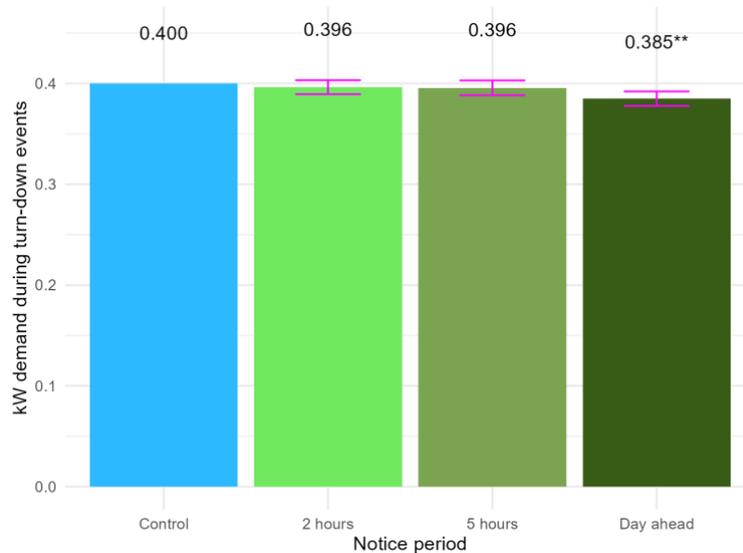
Figure 8: Regression coefficients for turn-down event kW demand by Notice Period
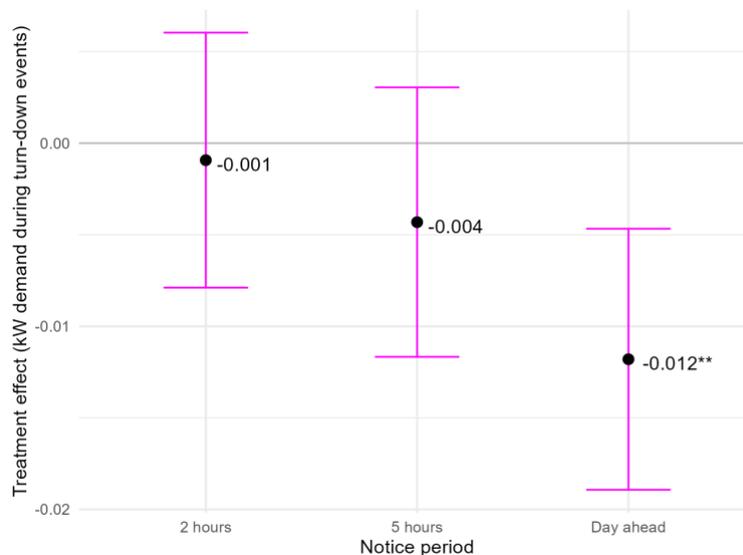
The overall contrast between turn-up and turn-down events suggests longer level of preparation and planning needed for turn-down than for turn-up, suggesting a different behavioural response.

3.2 Secondary analysis

3.2.1 Effect of incentives on opt-in to turn-up events

This analysis compared the event opt-in rates for turn-up events across the five incentive groups. Overall, the incentive groups' opt-in rates were similar to each other. However, we

have high statistical precision given the large sample sizes and number of events in the trial; thus, many of the small differences in opt-in rate *were* statistically significantly different.

- The Free Hours (Consistency) treatment consistently achieved higher opt-in rates than all other treatment groups, indicating that this structure led to increased participation.
- The £0.25/kWh + Down Events arm showed higher opt-in rates than the £0.25/kWh baseline, £0.20/kWh + Community Pot, and Free Hours (Volumetric) schemes, suggesting that combining turn-up and turn-down events encouraged engagement.
- The £0.20/kWh + Community Pot and Free Hours (Volumetric) groups also differed significantly, with slightly higher opt-in rates under the Free Hours (Volumetric) model, implying that individual rather than communal rewards were more motivating.
- Only one comparison showed no meaningful difference: the £0.25/kWh and £0.20/kWh + Community Pot arms had statistically indistinguishable opt-in rates, indicating that a change in monetary value and framing did not alter participation.

Figure 9: Turn-up opt-in rates by incentive group
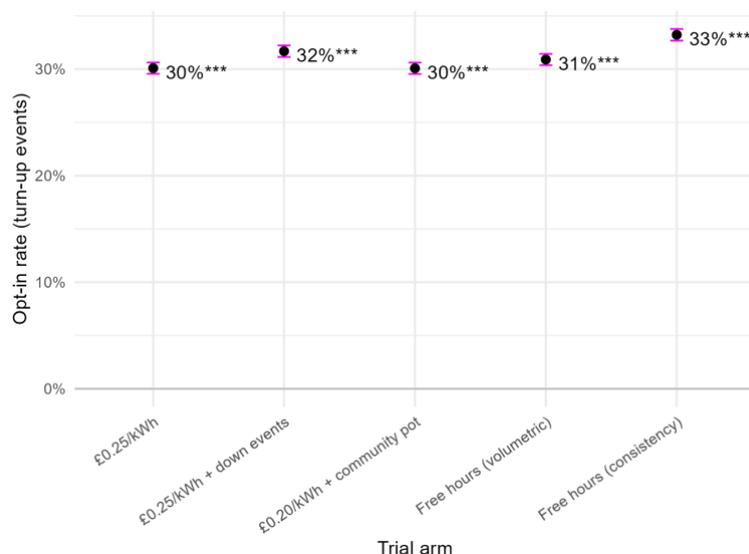
Table 6: Regression results for treatment vs. treatment comparison of turn-up event opt-in rates

Baseline Arm	Compared Arm	Baseline Avg.	Compared Avg.	Coefficient
£0.25/kWh + Down Events	£0.25/kWh	31.62%	30.24%	-0.016***
£0.20/kWh + Community Pot	£0.25/kWh	30.22%	30.24%	0.000
Free Hours (Volumetric)	£0.25/kWh	31.03%	30.24%	-0.008*
Free Hours (Consistency)	£0.25/kWh	33.27%	30.24%	-0.032***
£0.20/kWh + Community Pot	£0.25/kWh + Down Events	30.22%	31.62%	0.016***
Free Hours (Volumetric)	£0.25/kWh + Down Events	31.03%	31.62%	0.008*

Baseline Arm	Compared Arm	Baseline Avg.	Compared Avg.	Coefficient
Free Hours (Consistency)	£0.25/kWh + Down Events	33.27%	31.62%	-0.015***
Free Hours (Volumetric)	£0.20/kWh + Community Pot	31.03%	30.22%	-0.008*
Free Hours (Consistency)	£0.20/kWh + Community Pot	33.27%	30.22%	-0.032***
Free Hours (Consistency)	Free Hours (Volumetric)	33.27%	31.03%	-0.023***

3.2.2 Effect of incentives on opt-in to turn-down events

The overall opt-in rate for turn-down events was 35%. This figure is slightly higher than the average participation observed for turn-up events, where participation rates generally cluster around 32%. While this suggests a marginally greater willingness to participate in turn-down events, this result should be interpreted with caution due to the distinct scheduling patterns of each event type. Turn-down events were predominantly concentrated during the evening peak (17:00–20:00), accounting for 10 of the 12 events conducted. Conversely, the vast majority of turn-up events occurred in the early afternoon (12:00–15:00), representing 45 of the 48 total events. Consequently, the lower opt-in rate for turn-up events could partially reflect lower customer availability during working hours rather than a strict preference for the event type.

Figure 10: Turn-down opt-in rates by incentive group
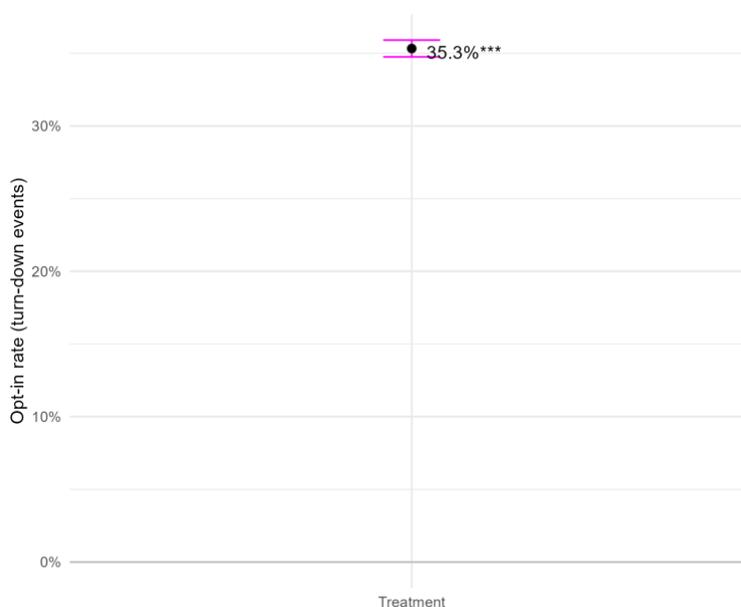
3.2.3 Local Average Treatment Effect (LATE) by incentive level among opt-in customers

While the previous section (3.1.1) examined overall treatment effects on electricity demand during turn-up events, averaging across all participants, including those who did not opt in,

this analysis isolates the impact *among those who actively participated*. This "Local Average Treatment Effect" (LATE) therefore represents the influence of each incentive structure on actual demand increases, "undiluted" by non-participation.

Results show that treatment effects on demand were substantially larger once restricted to participants who opted in, confirming that incomplete participation diluted the effects reported in Section 3.1.1. Across treatment arms, LATE estimates ranged from roughly 0.15 kW to 0.20 kW, compared with 0.05–0.06 kW in the full-sample analysis.[1] This indicates that, conditional on opting in, participants increased their electricity usage by around three to four times more than the average customer in the full sample.

Because opt-in rates differed modestly but significantly across treatments, comparisons of LATEs between groups are not causal. Groups with higher opt-in rates have LATEs that reflect a larger share of customers in that group, meaning the composition of the opt-ins may differ by group. That said, in this case the relative ranking of incentive structures was consistent with the primary analysis, which compared demand across treatment groups regardless of opt-in: The £0.25/kWh and £0.25/kWh + Down Events arms produced the strongest increases in demand among opt-in participants, while the Free hours and Community pot treatments generated smaller but still statistically significant responses (when compared with the control group).

---

[1] In other words, only 1/3 of participants opted in, on average, and so the effect of treatment among opt-ins was approximately three times higher than the effect of treatment among all participants.

Figure 11: LATEs on kW demand during turn-up events among customers who opted in, by incentive group
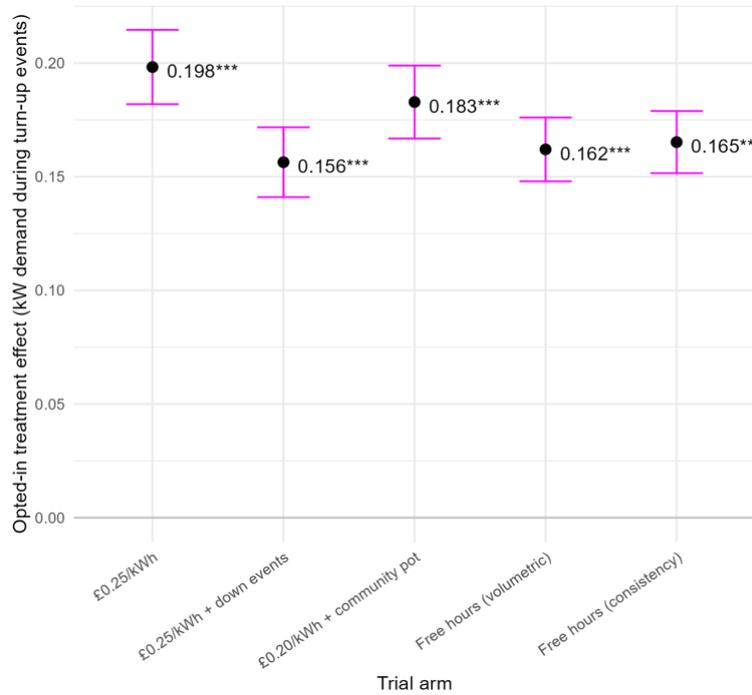
3.2.4 Effect of notice period on opt-in rate during turn-up and turn-down events

This analysis examined the effect of varying the event notice period on customer opt-in rates, comparing Day Ahead (DA), 5-hour, and 2-hour notifications for both turn-up and turn-down events.

The results demonstrated a consistent and significant relationship between the length of the notice period and customer participation across both event directions ($P < 0.001$ for all comparisons).

Figure 12: Opt-in rates by notice period for turn-up events
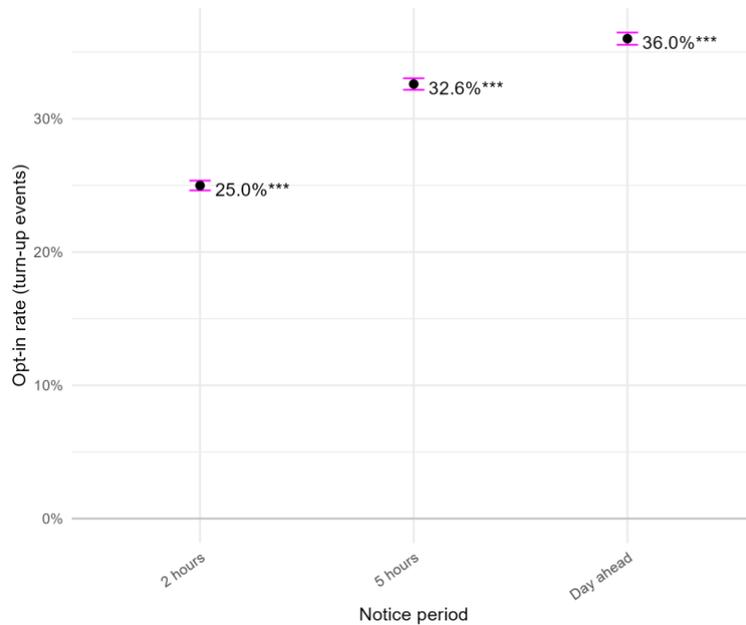
Figure 13: Opt-in rates by notice period for turn-down events
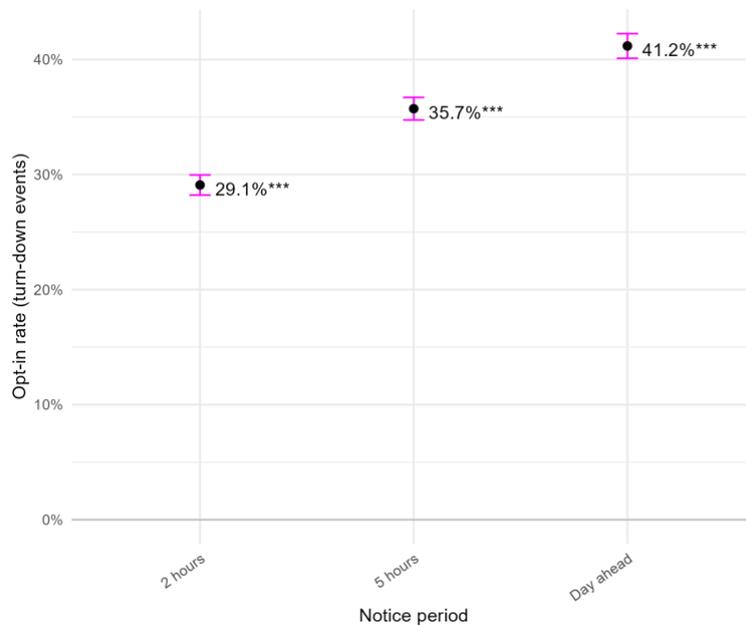
Increasing the notice provided to participants increased opt-in to trial events. The shortest 2-hour notice period had the lowest opt-in across both event directions. It is interesting to note that while the DA notice period proved the most effective lever for customer engagement (achieving the highest opt-in rates), its corresponding kW performance, as demonstrated in Section 3.1.3, did not result in different turn-up than the 5-hours' notice group.

Table 7: Regression results for effect of event notice period on opt-in rates

Event Direction	Baseline Arm (Avg. Opt-in)	Comparison Arm (Avg. Opt-in)	Coefficient	Interpretation
Turn-Up	5 hours **(32.68%)**	2 hours **(25.02%)**	−0.076***	2 hour notice reduced opt-in by 7.6% (vs 5h). Shorter notice significantly reduced opt-in.
	DA **(36.13%)**	5 hours **(32.68%)**	−0.034***	5 hour notice reduced opt-in by 3.4 % (vs DA). Shorter notice significantly reduced opt-in.
	2 hours **(25.02%)**	DA **(36.13%)**	0.110***	DA notice increased opt-in by 11.0% (vs 2h). Shorter notice significantly reduced opt-in.
Turn-Down	5 hours **(35.57%)**	2 hours **(29.02%)**	−0.067***	2 hour notice reduced opt-in by 6.7% (vs 5h). Shorter notice significantly reduced opt-in.
	DA **(41.18%)**	5 hours **(35.57%)**	−0.053***	5 hour notice reduced opt-in by 5.3% (vs DA). Shorter notice significantly reduced opt-in.
	2 hours **(29.02%)**	DA **(41.18%)**	0.120***	DA notice increased opt-in by 12.0% (vs 2h). Shorter notice significantly reduced opt-in.

3.2.5 LATE analysis: Relationship between notice period and kW demand among opt-in customers

This section explores how the length of notice affected customers' actual electricity demand during turn-up events, focusing only on those who opted in. The LATE captures the realised behavioural response among participating customers, isolating the effect of notice period from broader differences in participation.

The analysis indicates that participants who received shorter notice periods delivered stronger increases in kW demand. Estimated LATEs for the 2-hour and 5-hour groups were around 0.18–0.19 kW, compared with approximately 0.15 kW for the day-ahead group, with differences statistically significant ($p < 0.05$).

Variations in LATEs across notice periods are not causal. However, there is a suggestive pattern. One hypothesis is that the longer notice period seems to induce participation from a greater spread of customers, including some lower-response customers as well as some who tend to participate regardless of notice period. Another hypothesis is that customers in the day-ahead group are both more likely to opt-in but also more likely to forget about the event between opt-in and event. In surveys by the Centre for Sustainable Energy with customers who participated in this and previous CrowdFlex utilisation payment trials, customers noted that forgetting events *was* a barrier to participation – it was the *most* commonly reported challenge, with between 37% and 40% of respondents reporting in all three trials that they sometimes did not remember to take part. In addition, the Centre for Sustainable Energy found that those with short notice periods were more likely to find the notice period challenging but were less likely to say remembering to take part was an issue.

Figure 14: LATEs on kW demand during turn-up events among customers who opted in, by notice period
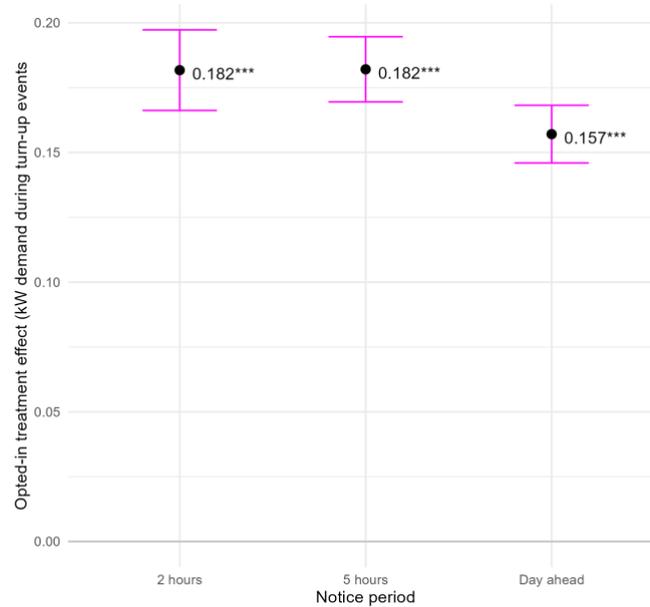
LATEs for demand (kW) during turn-down events showed a more similar pattern to the main analysis: measurable turn-down among opted-in customers in the day-ahead group, compared with levels of turn-down indistinguishable from zero in the shorter notice groups.

Note that the LATE where we pool notice periods – i.e., from the £0.25/kWh + Down events group – was –0.016 kW, and this effect was statistically significant.

Figure 15: LATEs on kW demand during turn-down events among customers who opted in, by notice period
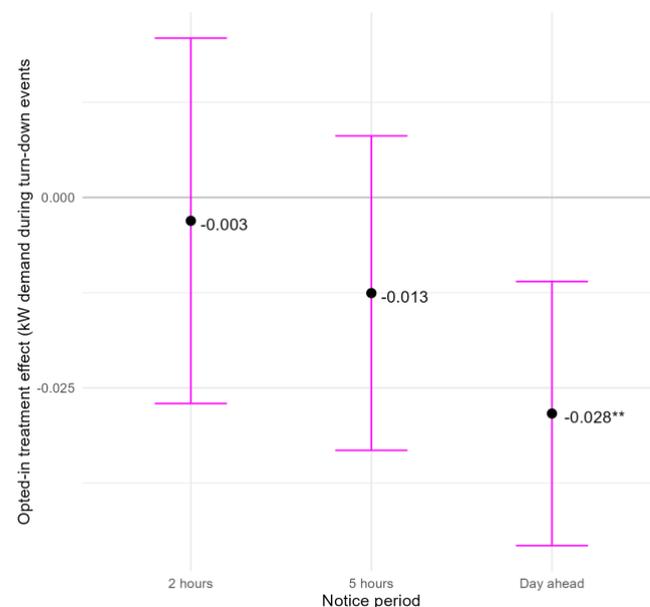
4.1 Event study: dynamic kW response around flexibility events

This analysis uses an event study methodology to examine the difference in kW demand between the Pooled Treatment group and the control groups during the hours surrounding a turn-up trial event. The analyses track the kW differential over 30-minute Settlement Periods (SPs), covering six hours before (−12 to −1 SPs), the (up to) two hours of the event itself (e1 to e4 SPs), and six hours after (1 to 12 SPs). Statistical significance is indicated by the confidence intervals not crossing zero.

4.1.1 Turn-up events

In the lead-up to the event, there appeared to be a slight but consistent dip in demand – i.e., the treatment effect for all periods up to the event were negative. Some of these results were statistically significant; in addition, it is important to note that we may be underpowered to detect diffuse demand reduction in the lead-up to the turn-up event.

During the event itself, consumption rose sharply among the treatment group relative to control. The estimated treatment effect was approximately +0.06 kW; note that this replicates our primary analysis, indicating that the intervention successfully increased electricity use during turn-up periods.

Following the event, consumption levels in the treatment group returned to near-baseline levels. However, there was some evidence of spillover demand increase into the half-hours after the event. We also see suggestive evidence of demand reduction in the four to six hours after an event (settlement periods 8-12), but these effects were not statistically significant.

Overall, these findings indicated that the intervention effectively shifted electricity consumption into the designated event period. We lack strong evidence of significant pre- or post-event displacement; however, we caution that it may be too diffuse to reliably detect, even given the relatively large sample size of the Treatment and Control groups.

For these event studies, we first show the average demand by Treatment versus Control during the settlement periods in advance, during, and after events. We then show the differences in these levels by Settlement Period, as represented by settlement-period-specific coefficients in regressions of demand on Treatment.

Figure 16: Average kW demand in Control and Treatment around turn-up events
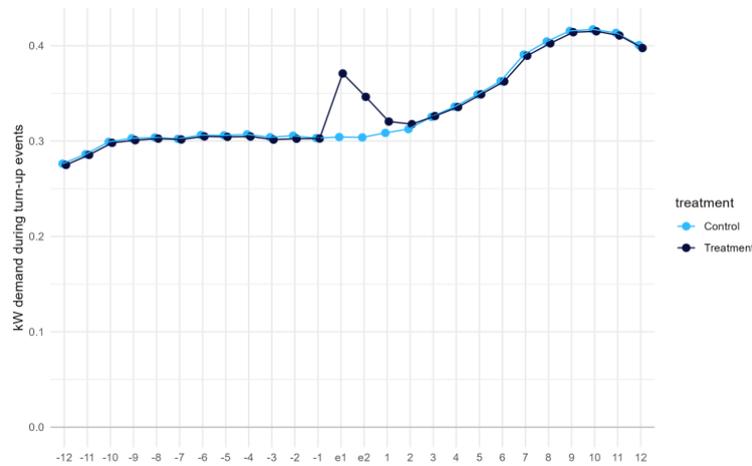
Figure 17: Estimated treatment effects (coefficients) around turn-up events
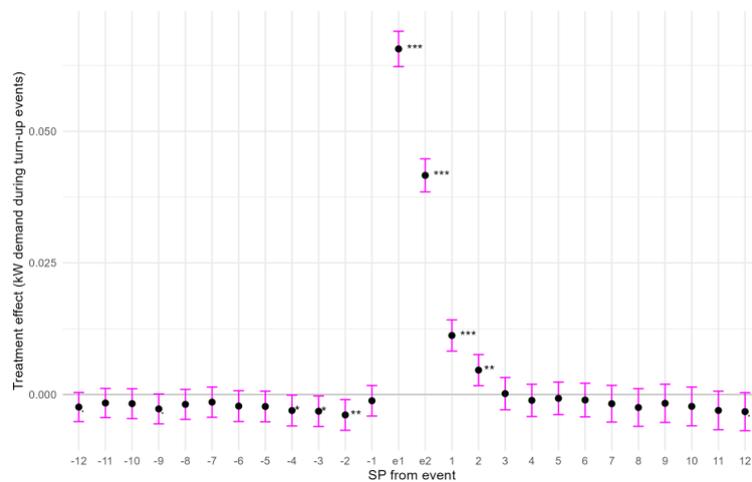
4.1.1 Turn-down events

Prior to event start, the treatment and control groups have similar demand, suggesting no or limited anticipatory effects. During the event window, average demand in the treatment group declined slightly relative to the control group, consistent with participants reducing consumption in response to the turn-down signal. Interestingly, the turn-down only noticeably occurred in the second settlement period. We hypothesise that this pattern may be due to the shorter-notice groups being more able to turn down in the second settlement period than the first, given the penalty from short notice we found in primary analysis. However, this is a speculative, exploratory hypothesis, and the pattern may indeed be statistical noise.

Following the event, consumption levels in both groups were similar. Interestingly, there was some evidence of spillover demand reduction in subsequent settlement periods. We are not certain how to interpret this pattern; as before, we caution that summer turn-down

effects were modest and based on a relatively small sample (only the £0.25/kWh + Down Events group), making the event study results noisy and potentially difficult to interpret.

Figure 18: Average kW demand in Control and Treatment around turn-down events
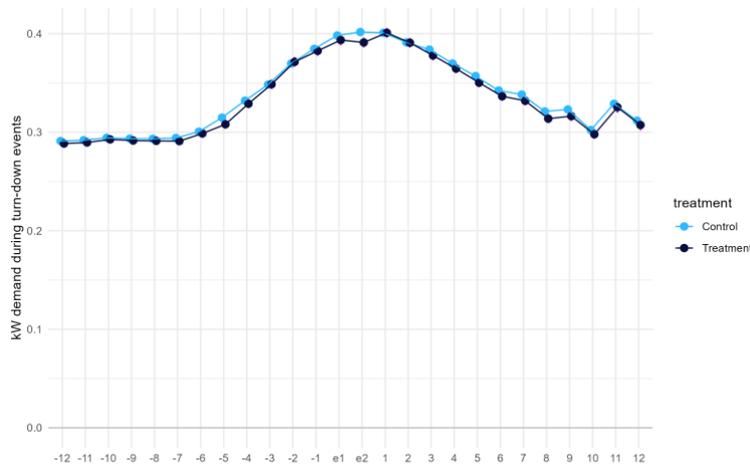
Figure 19: Estimated treatment effects (coefficients) around turn-down events
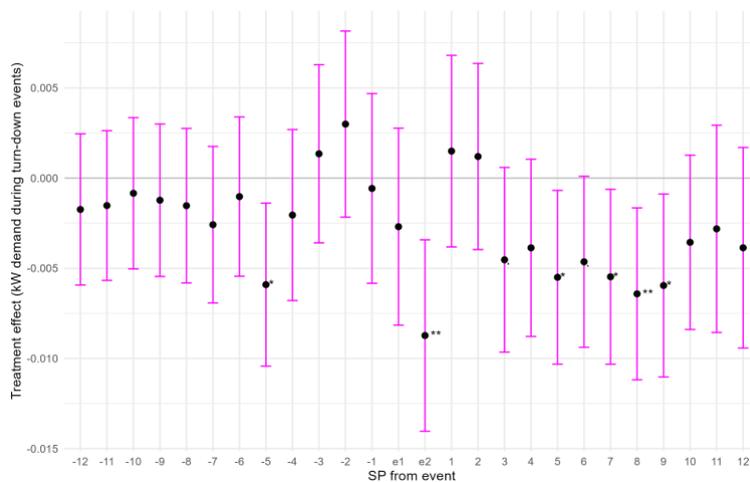
4.2 Subgroup analysis: Comparing treatment and control within participant segments

This analysis explored whether the treatment effect for turn-up event demand varied across participant subgroups.

Across almost all subgroups, turn-up events produced consistent, statistically significant increases in average kW demand among treated participants, with treatment effects typically ranging from 15–25%. The consistency of these effects suggests that response to treatment was broad-based rather than concentrated within specific groups. The largest relative increases were observed among households with LCTs (48%).

In contrast, turn-down events showed small and inconsistent reductions in kW demand, with some subgroups displaying statistically significant but small effects (generally below 3%) and others showing non-statistically significant treatment effects. We caution that statistical power here was low: we had small main effects of treatment on consumption to begin with, and here have subdivided our sample further to produce subgroup-specific effects. For this reason, we caution against over-interpretation of differences in turn-down effects by subgroup.

Note that in the tables below, Baseline represents the control group customers within that subgroup (for example, Control customers on weekdays vs. Treatment customers on weekdays). The % difference reflects the difference between the coefficient and subgroup-baseline average.

Two results are potentially of particular policy interest. First, customers on time-of-use (ToU) tariffs – predominantly "Economy 7", but also including other such tariffs such as "Economy 10" – exhibited higher turn-up than non-ToU customers – a pattern we also saw in CrowdFlex's winter utilisation payments trial. Second, customers on the Priority Services Register (PSR) showed higher turn-up than non-PSR customers.

Table 8: Estimated treatment effects on average consumption during turn-up events, by subgroup

Subgroup	Baseline avg.	Compared avg.	Coefficient	% difference	kW difference
Recruited	0.317	0.369	0.050***	15.89%	0.052
Auto-enrolled	0.293	0.352	0.056***	19.22%	0.059
Weekday	0.301	0.356	0.054***	17.93%	0.055
Weekend	0.320	0.373	0.052***	16.24%	0.053
Non ToU customer	0.304	0.356	0.052***	16.93%	0.052
ToU customer	0.301	0.375	0.069***	22.92%	0.074
High EAC: False	0.195	0.238	0.044***	22.36%	0.043
High EAC: True	0.425	0.491	0.065***	15.25%	0.066
EFA 4: 11:00 - 15:00	0.299	0.354	0.054***	18.20%	0.055
EFA 5: 15:00 - 19:00	0.292	0.344	0.052***	17.73%	0.052
EFA 6: 19:00 - 23:00	0.405	0.458	0.052***	12.85%	0.053
Power Move status: Never active	0.317	0.361	0.042***	13.22%	0.044
Power Move status: Previously active	0.289	0.337	0.049***	16.92%	0.048
Power Move status: Active	0.304	0.375	0.069***	22.70%	0.071
Non PSR customer	0.306	0.358	0.051***	16.59%	0.052
PSR customer	0.300	0.361	0.06***	20.19%	0.061

Subgroup	Baseline avg.	Compared avg.	Coefficient	% difference	kW difference
No children in household	0.280	0.358	0.071***	25.28%	0.078
Children in household	0.315	0.395	0.078***	24.62%	0.080
No seniors in household	0.285	0.36	0.069***	24.34%	0.075
Seniors in household	0.298	0.384	0.079***	26.61%	0.086
No LCT in household	0.297	0.369	0.066***	22.08%	0.072
LCT in household	0.245	0.367	0.12***	48.86%	0.122
EAC <2700 kWh	0.195	0.238	0.044***	22.36%	0.043
EAC >=2700 kWh	0.425	0.491	0.065***	15.25%	0.066

Table 9: Estimated treatment effects on average consumption during turn-down events, by subgroup

Subgroup	Baseline avg.	Compared avg.	Coefficient	% difference	kW difference
Recruited	0.420	0.411	-0.006	-1.35%	-0.009
Auto-enrolled	0.384	0.380	-0.006	-1.48%	-0.004
Weekday	0.403	0.395	-0.006*	-1.39%	-0.008
Weekend	0.367	0.359	-0.007	-1.78%	-0.008
Non ToU customer	0.401	0.393	-0.006*	-1.46%	-0.008
ToU customer	0.395	0.387	-0.009	-2.25%	-0.007
High EAC: False	0.256	0.249	-0.004*	-1.64%	-0.007
High EAC: True	0.559	0.554	-0.005	-0.94%	-0.006
EFA 4 (11:00–15:00)	0.292	0.290	0.000	-0.08%	-0.001
EFA 5 (19:00–23:00)	0.419	0.409	-0.008**	-1.99%	-0.011
EFA 6 (23:00–03:00)	0.429	0.424	-0.003	-0.78%	-0.005
Power Move status: Never active	0.423	0.409	-0.012*	-2.77%	-0.014
Power Move status: Previously active	0.390	0.388	-0.001	-0.32%	-0.002
Power Move status: Active	0.385	0.381	-0.004	-1.16%	-0.005
Non-PSR customer	0.408	0.401	-0.005	-1.26%	-0.007
PSR customer	0.381	0.371	-0.008	-1.97%	-0.010
No children in household	0.370	0.367	-0.009*	-2.44%	-0.002
Children in household	0.407	0.396	-0.007	-1.67%	-0.011

Subgroup	Baseline avg.	Compared avg.	Coefficient	% difference	kW difference
No seniors in household	0.373	0.369	-0.009	-2.30%	-0.005
Seniors in household	0.393	0.387	-0.010	-2.60%	-0.005
No LCT in household	0.382	0.376	-0.009**	-2.45%	-0.006
LCT in household	0.379	0.374	-0.007	-1.74%	-0.005
EAC <2700 kWh	0.256	0.249	-0.004*	-1.64%	-0.007
EAC >=2700 kWh	0.559	0.554	-0.005	-0.94%	-0.006

4.3 Relationship between customer participation and kWh shift

We conducted a series of exploratory, descriptive analyses comparing opt-in rates per event with shift per event. These analyses should not be treated as causal – certain events may elicit both more opt-in and more shift due to, for example, their timing, where it may not necessarily be the higher opt-in that caused the higher shift. However, overall, we see patterns that are reassuring as to what we might expect from our causal analysis, in that higher opt-in was correlated with greater group-level demand response magnitudes.

4.3.1 Overall relationship between opt-in rate and kWh shift across all events

This figure aggregated all notice periods to show the overall relationship between treatment opt-in rate and the kWh shift (Treatment – Control) for turn-down (left) and turn-up (right) events. Note that the shift here is between the whole Treatment group versus the whole Control group (not just among opt-ins). Each point represented an individual event, with a fitted linear regression line and 95% confidence interval.

When pooling across notice periods, turn-up events displayed a strong positive relationship between opt-in rate and aggregate kWh shift ($R^2 = 0.44$), indicating that higher opt-in was correlated with stronger collective load increases.

Conversely, turn-down events showed a relatively weaker – though still noticeable – negative trend ($R^2 = 0.15$). We speculate that turn-down effects were simply more modest at all levels of participation, and that extra participation had a more modest impact again on turn-down than turn-up, potentially reflecting saturation effects or behavioural limits to turning down demand.

Figure 20: Overall relationship between treatment opt-in rate and average kWh shift (Treatment – Control) across all events
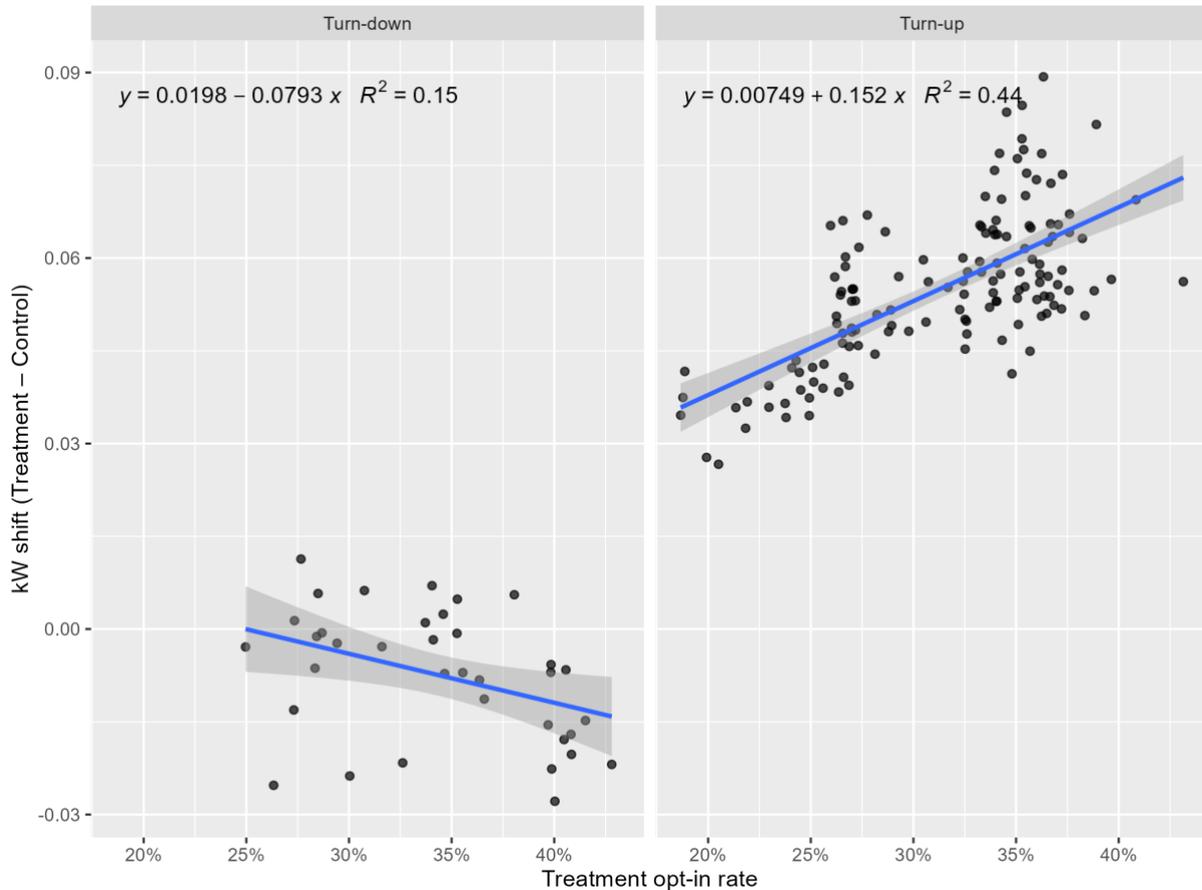
4.3.2 Relationship between opt-in rate and kWh shift by notice period

We repeated the analysis above, but where the analyses were split by notice period (2 hours, 5 hours, or day-ahead). As before, a linear regression line (blue) with a 95% confidence band (grey shading) illustrated the strength and direction of the relationship.

Across turn-up events, there was a strong positive correlation between opt-in rate and kWh shift ($R^2 ≈ 0.5$ across notice periods). This indicated that as more customers opted in, the aggregate increase in energy use during turn-up events grew proportionally, suggesting a scalable flexibility response. This relationship was weaker for the day-ahead group; we are not certain why this would be the case. One potential hypothesis is that there was more forgetting of the event among opt-ins in the day-ahead group, as discussed previously.

In contrast, turn-down events exhibited weak and inconsistent relationships (R^2 values between 0.02 and 0.11). There may have been too few events to draw reliable conclusions, even noting that these analyses were purely descriptive.

Figure 21: Relationship between treatment opt-in rate and average kWh shift (Treatment – Control), disaggregated by event notice period
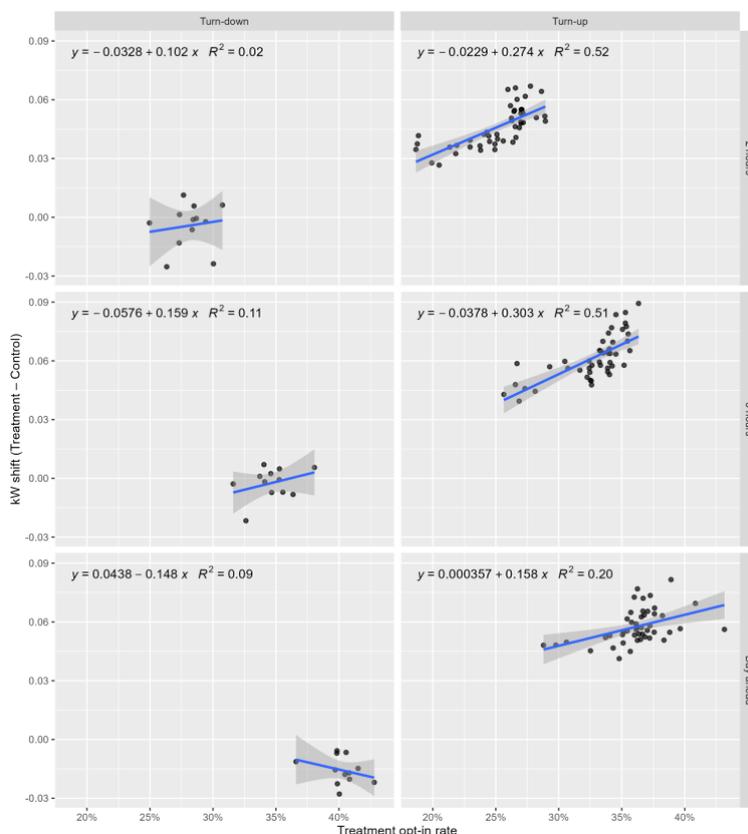
4.4 Examining baseline accuracy

This analysis compared three methods of estimating shifts in electricity consumption during turn-up events across treatment arms: two P376-based shifts, and a control-based shift. The objective was to assess whether the P376 methodology produced systematically higher or lower estimates of turn-up response than those implied by the observed control-adjusted treatment averages.

P376 is a Balancing and Settlement Code modification that defines a baseline derived from meter data against which metered event demand is compared to calculate demand shift. Essentially, DSRSPs compare actual consumption to recent historical consumption in the same settlement periods to estimate shift, especially when remunerating shift directly, as was the case in this trial. Given its wide use in demand-side response operations, assessing its accuracy in our trial provides a useful contribution to wider flexibility work across the energy industry.

4.4.1 Methodology

For each treatment group, the P376-based shift was estimated as the mean clipped P376 shift among opt-in customers. "Clipped" means that individual-level shifts in the opposite direction are not penalised; that is, they are treated as 0 kWh rather than included as

negative shifts. The term refers to this adjustment, which effectively clips negative values to zero. This is the current methodology for remunerating customers in current live markets, such as the demand flexibility services (DFS), so serves as a useful comparison. We then multiplied this mean by the opt-in rate for that treatment arm to produce an aggregate shift value representative of the full trial sample base, rather than only those participants who opted in to events.

We also examined the unclipped P376-based shift among all trial participants. This estimate was not used for remuneration. However, comparing it with the remuneration-based estimate provides an indication of the reduced accuracy introduced by clipping.

The control-based shift was calculated as the difference between the average consumption observed in the treatment group and the corresponding value in the control group. This approach provided a "ground truth" against which to measure the accuracy of the baseline-based measure.

As in previous CrowdFlex trials, we found that clipping reduced the accuracy of measured flexibility. In contrast, the unclipped baseline was a fairly good approximation of group-level shift magnitude.

4.4.2 Results – Turn-up

The P376-derived shifts consistently exceeded those calculated using the control-based method. The mean clipped P376-derived shift among opt-in customers was 40–54% higher than the control-based estimate, whereas the unclipped P376-derived shift across all participants was 5–8% higher. This pattern suggests that the P376 baseline may have overestimated the extent of the true behavioural shift during turn-up events when applied across all customers. The magnitude of overestimation was relatively consistent across treatment arms, implying that the bias likely reflected systematic issues in the baseline methodology rather than differences in participant response.

The accuracy was much greater for the unclipped comparisons P376 estimate (i.e. with unclipped opt-ins only and unclipped all participants) than the clipped (opt-ins only) estimate. The unlipped comparisons were much closer to the control-based estimate, suggesting that most of the P376 inaccuracy stemmed from the clipping process. Some residual differences remained, which we speculate may be due to weather effects: specifically, the baseline period being consistently warmer than the event period, as the trial spanned late summer into autumn.

Figure 22: P376 comparison – turn-up events
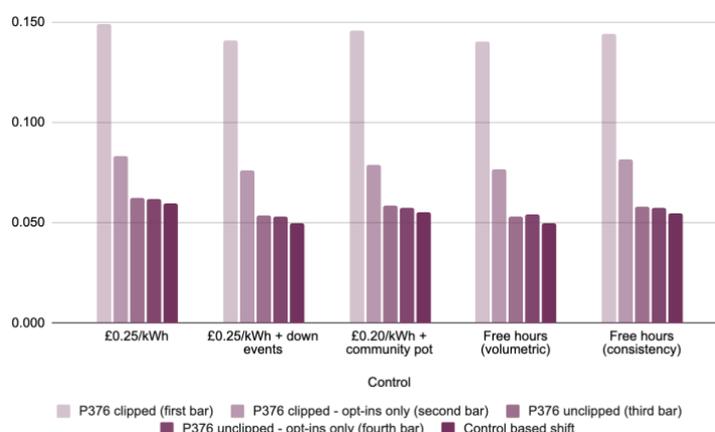
Table 10: P376 analysis across turn-up events

Trial arm	P376 clipped events only (all participants)	P376 clipped (opt-in only)	P376 unclipped (all participants)	P376 unclipped (opt-in only)	Control based shift
£0.25/kWh	0.149	0.083	0.062	0.062	0.059
£0.25/kWh + down events	0.141	0.076	0.054	0.053	0.050
£0.20/kWh + community pot	0.146	0.079	0.059	0.057	0.055
Free hours (volumetric)	0.140	0.077	0.053	0.054	0.050
Free hours (consistency)	0.144	0.082	0.058	0.058	0.055

4.4.2 Results – Turn-down

Turn-down results were considerably noisier.

The modest true turn-down effect appears to have been swamped by P376 baseline noise and the influence of clipping, together producing an overestimate of turn-down of 95.13% in the remuneration method (i.e., clipped, among opt-ins only and all participants).

In contrast, the unclipped P376 estimates incorrectly indicated a backfire (i.e., an apparent turn-up). We speculate that this occurred because, during the trial period, baseline consumption was lower than the true counterfactual – as in the P376 turn-up section above, likely due to the timing of the trial – leading to an underestimate of turn-down so large that it appeared as a backfire.

Figure 23: P376 comparisons – turn-down events
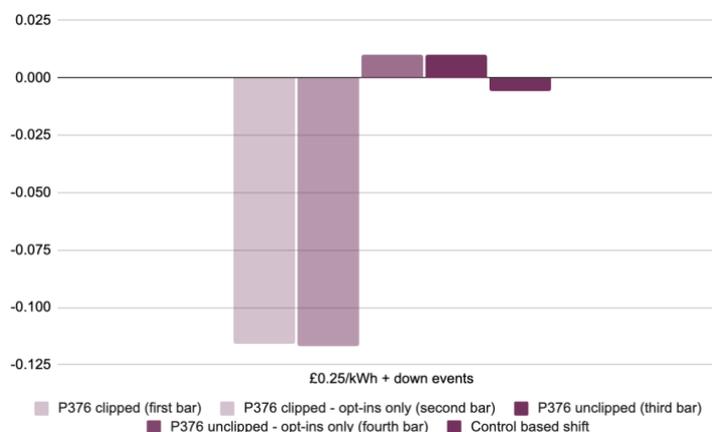
Table 11: P376 analysis across turn-down events

Trial arm	P376 clipped events only (all participants)	P376 clipped (opt-in only)	P376 unclipped (all participants)	P376 unclipped (opt-in only)	Control based shift
£0.25/kWh + down events	-0.116	-0.117	0.010	0.010	-0.006

4.5 Exploratory analysis: Effect of SMS on kW demand

As discussed in Section 2, OVO introduced SMS notifications on 16 August, later adding SMS reminders from 17 September. Among customers eligible to receive these messages (i.e., those for whom OVO held a valid phone number), we randomised half to receive the SMSs and half not to. This randomisation enabled identification of the causal impact of the SMSs on consumption during events.

To do so, our analysis was restricted to eligible customers (62,689 across Treatment and Control, approximately 84% of the sample) and to turn-up events occurring after the introduction of each SMS type. Accordingly, the *notification-only* analysis draws on events from 16 August through 16 September, while the *notification-and-reminder* analysis uses events from 17 September through the end of the trial. Within this restricted sample and period, we found results consistent with the main analysis: all treatments significantly increased kW demand relative to the control group, with average demand rising from around 0.30 kW to between 0.34 and 0.36 kW, equivalent to gains of 14–21% (p < 0.001).

We found that the introduction of SMS notifications further increased kW demand, producing a 2.6% greater turn-up (with respect to the average in the Treatment group not receiving SMSs, p < 0.001). Adding SMS reminders seemed to strengthen these effects, though we should be cautious about comparing treatment effects for different sets of events. Compared with the Control, treatments with reminders increased kW demand by 15–24% (p < 0.001), and within-treatment comparisons showed an additional ~4% increase (all p < 0.01).

Table 12: Average treatment effects of SMSs on turn-up kW demand

SMSs	Group	Avg.	Coefficient	SE	P-value	% difference
Notification only	Control	0.298				
	Treatment	0.348	0.049	0.002	<0.001***	16.51%
	Treatment + SMS	0.357	0.058	0.002	<0.001***	19.58%
Notification and reminder	Control	0.318				
	Treatment	0.375	0.055	0.003	<0.001***	17.17%
	Treatment + SMS	0.389	0.068	0.003	<0.001***	21.53%

For *notification only*, these effects were most pronounced for the 2-hour (p < 0.001) and day-ahead (p = 0.015) notice periods, while the 5-hour period showed no significant change – an odd pattern indicating some potential noisiness in these effects by treatment group. *Notification-and-reminder* effects were extremely consistent in their impact across notice periods – producing an extra 3.6-3.7% more turn-up for all three groups. Taken together, these results demonstrate that text messages (notifications and reminders) increased turn-up demand, perhaps particularly at shorter notice.

Table 13: Effect of SMS notifications on kW demand during turn-up events, split by notice period

Baseline	SMS type	Baseline avg.	Comparison avg. (+ SMS)	Coefficient	% difference
2 hours	Notification only	0.341	0.353	0.014***	3.96%
	Notification + reminder	0.367	0.379	0.014***	3.74%
5 hours	Notification only	0.354	0.361	0.004	1.26%
	Notification + reminder	0.381	0.399	0.014**	3.67%
Day Ahead	Notification only	0.349	0.357	0.009*	2.71%
	Notification + reminder	0.376	0.388	0.014**	3.61%

By contrast, during turn-down events, neither treatments nor SMS communications produced detectable reductions in demand among customers eligible to receive SMSs. Note that this does not mean that the treatment worked better for SMS-ineligible customers and failed to work for SMS-eligible customers; the events themselves are different, being only a subset of the full event season. In addition, we face a reduction in statistical precision caused by the sample here being reduced, due to fewer events and customers being involved in the analysis, meaning greater difficulty detecting small effects.

Table 14: Average treatment effects of SMSs on turn-down kW demand

SMSs	Trial arm	Avg.	Coefficient	SE	P-value	% difference
Notification only	Control	0.383				
	Treatment	0.376	-0.004	0.004	0.3601	-0.98%
	Treatment + SMS	0.382	0.003	0.004	0.4408	0.83%
Notification and reminder	Control	0.470				
	Treatment	0.460	-0.008	0.006	0.1918	-1.71%
	Treatment + SMS	0.465	-0.002	0.006	0.7625	-0.39%

Table 15: Effect of SMS notifications on kW demand during turn-down events, split by notice period

Baseline	SMS type	Baseline avg.	Comparison avg.	Coefficient	% difference
2 hours	Notification only	0.386	0.378	-0.003	-0.76%
	Notification + reminder	0.462	0.472	0.012	2.53%
5 hours	Notification only	0.373	0.395	0.018	4.92%
	Notification + reminder	0.467	0.464	-0.003	-0.66%
DA	Notification only	0.371	0.375	0.005	1.29%
	Notification + reminder	0.451	0.458	0.008	1.88%

4.6 Exploratory analysis: Effect of SMS on event opt-in

SMS messages significantly increased opt-in rates during turn-up events. Compared to the pooled treatment group, adding an SMS notification raised participation by 3.5–4.7% (p < 0.001), with both the initial notification and reminder contributing positively. For turn-down events, SMS messages also increased opt-ins slightly (+1.6–3.2%), though these effects were smaller in magnitude and only marginally significant. Overall, SMS communication improved opt-in rates, especially for turn-up events, and we hypothesise that it was likely the main mechanism for the improved turn-up they caused.

Table 16: Effect of SMS notifications on opt-in rates across treatment groups during turn-up and turn-down events (Pooled treatment groups)

SMS type	Event direction	Baseline avg.	Comparison avg. (+SMS)	Coefficient	% difference
Notification only	Turn-up	32.1%	33.4%	0.011***	3.50%

SMS type	Event direction	Baseline avg.	Comparison avg. (+SMS)	Coefficient	% difference
Notification + reminder		32.2%	34.0%	0.015***	4.69%
Notification only	Turn-down	34.9%	36.2%	0.011	3.16%
Notification + reminder		36.2%	37.1%	0.006	1.64%

4.7 Exploratory analysis: Cost per kWh of flexibility

The cost-effectiveness analysis, measured by £/kWh of demand increase (compared to the control group), demonstrated that the economic value of flexibility was maximised through non-monetary, in-kind rewards. All five treatment arms were initially designed to offer an approximately equal expected monetary reward expenditure of around £0.25/kWh of turn-up by the operator, yet the realised financial value delivered varied substantially.

The Free hours (volumetric) group proved to be the most cost-effective, delivering flexibility at £0.13/kWh, followed by the Free hours (consistency) group at £0.14/kWh. These incentive structures successfully generated significant changes in consumption (when compared with the control group), with the consistency group achieving 0.055 kW turn-up and the volumetric consistency group achieving 0.050 kW. These turn-up coefficients were very similar to the other treatment groups; yet the total remuneration paid out for these groups was significantly lower (£3,635.14 and £4,584.38, respectively).

Direct financial payment schemes were more expensive to procure flexibility. The treatment providing a direct payment of £0.25/kWh, which achieved the highest efficacy with a 0.059 kW coefficient, cost £0.32/kWh. The £0.20/kWh + community pot structure offered similar performance (0.055 kW coefficient) at a similar cost, £0.33/kWh.

The least cost-effective approach was the £0.25/kWh + down events group, costing £0.36/kWh. This cost reflects the total remuneration paid for the turn-up (£10,337) events only and therefore excludes the turn-down portion. The group also exhibited a dampened turn-up performance (0.050 kW coefficient) compared to the up-only arm (which, as discussed previously, suggests that exposure to both event types reduced customers' delivery of turn-up).

In conclusion, the £/kWh results highlight that the utilisation of free hours of electricity as an incentive structure delivered comparable flexibility shifts to the direct payment mechanisms but much more cost-efficiently.

Table 17: Cost per kWh of flexibility by treatment group

Treatment group	Remuneration	Turn-up coeff	Turn-up hours	Avg start, end	kWh shift	£/kWh
£0.25/kWh	£11,111.27	0.059	48	12,090	34,487	£0.32

Treatment group	Remuneration	Turn-up coeff	Turn-up hours	Avg start, end	kWh shift	£/kWh
£0.25/kWh + down events	£10,337.38	0.050	48	11,996	28,516	£0.36
£0.20/kWh + community pot	£10,369.79	0.055	48	12,080	31,870	£0.33
Free hours (volumetric)	£3,635.14	0.050	48	12,094	28,917	£0.13
Free hours (consistency)	£4,584.38	0.055	48	12,066	31,653	£0.14

5. Discussion

As in previous CrowdFlex trials, this trial's results showed that incentives can be used to influence household consumption during explicit flexibility events. Across all treatment groups, participants responded in the expected direction during events. The Summer trial – focused predominantly on turn-up events – further considered how the way incentives were framed and communicated affected both participation and demand response.

Direct monetary rewards of £0.25/kWh produced the largest average effect – approximately 20% turn-up compared to the control group. However, the community and free electricity schemes achieved similar shifts (only 1–3% lower than the turn-up in the direct reward group); and the free electricity groups did this at roughly a third of the cost. The close similarity in response generated by financial and in-kind incentives suggested that participants may have been motivated more by the chance to take part and the perceived value of the reward than by the exact monetary value or incentive structure. Community-based rewards did not noticeably raise engagement, implying that collective incentives may require stronger social framing or feedback to be effective; but neither did they generate a strong backfire.

For customers exposed to both turn-up and turn-down events, we observed a small but statistically significant decline in turn-up performance relative to those facing only turn-up events. This may reflect mild behavioural fatigue, given that participants were asked to respond to an extra event per week – one of a different "direction" than their typical four turn-up events. Demand response programmes may need to balance the frequency and diversity of events to maintain participant engagement and minimise cognitive load.

The length of notice influenced turn-up and turn-down response. For turn-down, day-ahead notified customers had statistically significant response, while the other groups did not; for turn-up, the day-ahead and 5-hour notice period groups had greater turn-up than the 2-hour notice group. Day-ahead notifications produced the highest opt-in rates (36% for turn-up and 41% for turn-down), compared with 25–30% under two-hour notice. The impact of day-ahead notice reinforces results from the 2022/23 Demand Flexibility Service (e.g., Jacob et al., 2024[2]). This pattern suggests that consumers may benefit from more time to plan reductions in electricity use, whereas increasing consumption during surplus periods can perhaps be achieved with shorter lead times.

---

[2] Jacob, M., Jenkinson, R., Lopez Garcia, D., Metcalfe, R. D., Schein, A. R., Simpson, C. R., & Yu, L. (2024). The impact of demand response on energy consumption and economic welfare [Working paper]. Centre for Net Zero.

Interestingly, when examining only customers who opted in, the pattern of kW response reversed: shorter-notice groups displayed stronger average turn-up per participant. One possibility is that shorter notice primarily screened out less engaged participants, leaving a subset of highly responsive households whose flexibility was relatively unaffected by timing, whereas day-ahead events attracted more participants overall but some of the "extra" responders contributed less per household. Another possibility is that day-ahead notice encouraged high opt-in rates, but some customers who opted in subsequently forgot about the event or were otherwise less committed to following through, simply because there was a longer gap between opting in and the event start.

The introduction of SMS notifications and reminders proved effective in boosting both participation and performance, particularly for turn-up events. Messages increased participation by approximately 3% and consumption during turn-up events by a further 3-4 percentage points (on top of the 16-17% increase in the absence of notifications). The impact was perhaps most pronounced for shorter-notice events, suggesting that timely reminders can make up for less notice. These results reinforce the importance of salient reminders to convert initial interest into active participation.

In examining the accuracy of baselines, we found that clipping reduced the accuracy of baseline-based estimates of flexibility. This effect was smaller when using only opted-in customers (excluding non-participating customers from the flexibility calculation). CrowdFlex was able to quantify the accuracy of different baselining approaches with high evidentiary quality, thanks to its RCT structure. These findings showed that some commonly used practices – such as clipping and the inclusion of customers not opted in – can introduce substantial inaccuracies. As flexibility services expand, operators should continue to explore ways to develop robust and common baselining methods to establish a fair playing field for market participants and consumers.

Treatment effects were broadly consistent across customer subgroups, suggesting that the interventions engaged a wide cross-section of participants. Nonetheless, the strongest relative increases were observed among households with low-carbon technologies, who delivered approximately 48% higher proportional response. This likely reflects their greater capacity for flexible load shifting through assets such as EV chargers, heat pumps, and smart appliances. It also highlights the growing potential for coordination between flexibility programmes and electrification policies to unlock automated, scalable response.

Overall, the trial demonstrated that monetary payments were effective, but not uniquely so. Programmes offering in-kind rewards could harness comparable behavioural responses while reducing cost to the system operator. Longer notice periods may support manual participation, though targeted reminders can partially offset this barrier. The combination of intuitive incentive design and timely prompts will likely be central to scaling domestic flexibility.

6. Appendix

6.1 Treatment group details

1. Turn-up and turn-down group

- **Definition:** This group incentivised participants to both increase their electricity demand (turn-up) at specific times and decrease their electricity demand (turn-down) at other specific times in response to signals.
- **Criteria:**
    - Turn-up: Participants in this group received a payment of £0.25 per kWh for every kilowatt-hour of increased demand during a designated turn-up event, as measured against a baseline.
    - Turn-down: Participants in this group received a payment of £0.50 per kWh for every kilowatt-hour of decreased demand during a designated turn-down event, as measured against the same baseline.
    - The P376 baseline methodology was used to calculate the change in demand for both demand turn-up (DTU) and demand turn-down (DTD) events across all notice periods (Short, Medium, Long).

2. Turn-up only group

- **Definition:** This group focused solely on incentivising participants to increase their electricity demand (turn-up) at specific times in response to signals.
- **Criteria:**
    - Participants in this group received a payment of £0.25 per kWh for every kilowatt-hour of increased demand during a designated turn-up event, as measured against a baseline.
    - The P376 baseline methodology was used to calculate the change in demand for both DTU and DTD events across all notice periods (Short, Medium, Long).

3. Community pot group

- **Definition:** This group introduced a community-based reward system where participants contributed to a shared pot based on their flexible electricity consumption. The top-performing communities then shared this pot.
- **Criteria:**
    - Participants received a headline reward of £0.20 per kWh for each kilowatt-hour of delivered flexibility.
    - An additional £0.05 per kWh of delivered flexibility was contributed to a community pot.
    - Top 10 performing communities were ranked based on the total kWh shifted compared to the P376 baseline, combined across all participants in that community, over an entire month.
    - Communities were defined by the first two letters of participants' postcodes. When a community size became too large, they were further subdivided to

ensure an acceptable average payout. Participants were made aware of their assigned Community at the start of the trial.
   o There was a minimum guaranteed pot size of £1000. The minimum pot was estimated to equate to a payout of approximately £1.50 per winning customer.
   o The "expected" pot size was around £1800, based on an average event contribution of 0.25 kWh per hour-long event. This estimation used observed shifts from the control group and P376 "correct direction" shifts to simulate an expected P376-with-opt-in shift.
- The goal was to provide an overall, average reward for turn-up similar in cost to the reward for Group 1 and 2 (£0.25/kWh).
- To be eligible for the bonus, customers needed to opt-in and participate in at least 50% of events each month (approximately 8 of 16 events expected per month). Payout was equal regardless of individual performance, conditional on eligibility.
   o "Participate" means: Participants must have opted in and turned up by 0.05 kWh/hour in an event. This threshold is based on analysis of winter consumption data in the control group.

4. Free electricity group (volumetric)

- Definition: This group rewarded participants with free electricity, with the reward structure based on the amount (i.e., volume) of energy flexibility delivered.
- Criteria: Participants in this group earned free electricity based on the total amount (in kWh) of demand response they provided during events. The indicative scale for earning free electricity was as follows:
   o 0.5 kWh of delivered flexibility = 0.5 free hours of electricity
   o 1 kWh of delivered flexibility = 1 free hour of electricity
   o 1.5 kWh of delivered flexibility = 1.5 free hours of electricity
   o 2 kWh of delivered flexibility = 2 free hours of electricity
- Note that the total free electricity earned in any given week was capped at a maximum of 2 hours.

5. Free electricity group (consistency)

- **Definition:** This group rewarded participants with free electricity, with the reward structure based on the consistency of energy flexibility delivered
- **Criteria:** Participants in this group earned free electricity based on the consistency with which they participate in demand response events. The indicative scale for earning free electricity was as follows:
   o 1 event = 0.5 free hours of electricity
   o 2 events = 1 free hour of electricity
   o 3 events = 1.5 free hours of electricity
   o 4 events = 2 free hours of electricity
- To achieve "consistency", participants must have opted in and turned up by 0.05 kWh/hour in an event. This threshold is based on analysis of winter consumption data in the control group.
- As with group 4, the total free electricity earned in any given week was capped at a maximum of 2 hours.

6.2 Balance check results

Table 18: EAC kW/h balance check - Treatment group pairwise comparisons

Trial arm (1)	Trial arm (2)	p
Control	£0.25/kWh + Down Events	0.665
Control	£0.25/kWh	0.038
Control	Community	0.098
Control	Free Hours (Volumetric)	0.735
Control	Free Hours (Consistency)	0.591
£0.25/kWh + Down Events	£0.25/kWh	0.103
£0.25/kWh + Down Events	Community	0.222
£0.25/kWh + Down Events	Free Hours (Volumetric)	0.918
£0.25/kWh + Down Events	Free Hours (Consistency)	0.919
£0.25/kWh	Community	0.69
£0.25/kWh	Free Hours (Volumetric)	0.077
£0.25/kWh	Free Hours (Consistency)	0.177
Community	Free Hours (Volumetric)	0.182
Community	Free Hours (Consistency)	0.26
Free Hours (Volumetric)	Free Hours (Consistency)	0.836

Table 19: EAC kW/h balance check: Notice period treatment group pairwise comparisons

Trial arm (1)	Trial arm (2)	p
2hr	5hr	0.283
2hr	DA	0.224
5hr	Control	0.635
5hr	DA	0.886
8hr	Control	0.167
DA	Control	0.134

Table 20: EAC balance check - Treatment group pairwise comparisons within notice period

Notice period	Trial arm (1)	Trial arm (2)	p
2hr	£0.25/kWh + Down Events	£0.25/kWh	0.396
2hr	£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.825

Notice period	Trial arm (1)	Trial arm (2)	p
2hr	£0.25/kWh + Down Events	Free Hours (Volumetric)	0.705
2hr	£0.25/kWh + Down Events	Free Hours (Consistency)	0.517
2hr	£0.25/kWh	£0.20/kWh + Community Pot	0.532
2hr	£0.25/kWh	Free Hours (Volumetric)	0.634
2hr	£0.25/kWh	Free Hours (Consistency)	0.137
2hr	£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.876
2hr	£0.20/kWh + Community Pot	Free Hours (Consistency)	0.387
2hr	Free Hours (Volumetric)	Free Hours (Consistency)	0.304
5hr	£0.25/kWh + Down Events	£0.25/kWh	0.346
5hr	£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.646
5hr	£0.25/kWh + Down Events	Free Hours (Volumetric)	0.366
5hr	£0.25/kWh + Down Events	Free Hours (Consistency)	0.956
5hr	£0.25/kWh	£0.20/kWh + Community Pot	0.628
5hr	£0.25/kWh	Free Hours (Volumetric)	0.063
5hr	£0.25/kWh	Free Hours (Consistency)	0.374
5hr	£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.17
5hr	£0.20/kWh + Community Pot	Free Hours (Consistency)	0.686
5hr	Free Hours (Volumetric)	Free Hours (Consistency)	0.335
DA	£0.25/kWh + Down Events	£0.25/kWh	0.303
DA	£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.162
DA	£0.25/kWh + Down Events	Free Hours (Volumetric)	0.732
DA	£0.25/kWh + Down Events	Free Hours (Consistency)	0.451
DA	£0.25/kWh	£0.20/kWh + Community Pot	0.679
DA	£0.25/kWh	Free Hours (Volumetric)	0.48
DA	£0.25/kWh	Free Hours (Consistency)	0.78
DA	£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.274
DA	£0.20/kWh + Community Pot	Free Hours (Consistency)	0.496
DA	Free Hours (Volumetric)	Free Hours (Consistency)	0.671

Table 21: EAC balance check - Treatment group pairwise comparisons within notice period

Treatment group	Notice period (1)	Notice period (2)	p
£0.25/kWh + Down Events	2hr	5hr	0.606
£0.25/kWh + Down Events	2hr	DA	0.973
£0.25/kWh + Down Events	5hr	DA	0.597
£0.25/kWh	2hr	5hr	0.514
£0.25/kWh	2hr	DA	0.843
£0.25/kWh	5hr	DA	0.647
£0.20/kWh + Community Pot	2hr	5hr	0.445
£0.20/kWh + Community Pot	2hr	DA	0.228
£0.20/kWh + Community Pot	5hr	DA	0.659
Free Hours (Volumetric)	2hr	5hr	0.437
Free Hours (Volumetric)	2hr	DA	0.96
Free Hours (Volumetric)	5hr	DA	0.477
Free Hours (Consistency)	2hr	5hr	0.228
Free Hours (Consistency)	2hr	DA	0.167
Free Hours (Consistency)	5hr	DA	0.864

Table 22: Peak (kWh consumption) balance check - Treatment group pairwise comparisons

Trial arm (1)	Trial arm (2)	p
Control	£0.25/kWh + Down Events	0.456
Control	£0.25/kWh	0.387
Control	£0.20/kWh + Community Pot	0.431
Control	Free Hours (Volumetric)	0.678
Control	Free Hours (Consistency)	0.691
£0.25/kWh + Down Events	£0.25/kWh	0.106
£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.13
£0.25/kWh + Down Events	Free Hours (Volumetric)	0.245
£0.25/kWh + Down Events	Free Hours (Consistency)	0.733
£0.25/kWh	£0.20/kWh + Community Pot	0.961
£0.25/kWh	Free Hours (Volumetric)	0.653

Trial arm (1)	Trial arm (2)	p
£0.25/kWh	Free Hours (Consistency)	0.209
£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.699
£0.20/kWh + Community Pot	Free Hours (Consistency)	0.243
Free Hours (Volumetric)	Free Hours (Consistency)	0.418

Table 23: Peak (kWh consumption) balance check - Notice period treatment group pairwise comparisons

Trial arm (1)	Trial arm (2)	p
2hr	5hr	0.651
2hr	DA	0.51
2hr	Control	0.911
5hr	DA	0.846
5hr	Control	0.778
DA	Control	0.647

Table 24: Peak kW/h balance check - Treatment group pairwise comparisons within notice period

Notice period	Trial arm (1)	Trial arm (2)	p
2hr	£0.25/kWh + Down Events	£0.25/kWh	0.574
2hr	£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.767
2hr	£0.25/kWh + Down Events	Free Hours (Volumetric)	0.956
2hr	£0.25/kWh + Down Events	Free Hours (Consistency)	0.592
2hr	£0.25/kWh	£0.20/kWh + Community Pot	0.389
2hr	£0.25/kWh	Free Hours (Volumetric)	0.534
2hr	£0.25/kWh	Free Hours (Consistency)	0.283
2hr	£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.808
2hr	£0.20/kWh + Community Pot	Free Hours (Consistency)	0.797
2hr	Free Hours (Volumetric)	Free Hours (Consistency)	0.625
5hr	£0.25/kWh + Down Events	£0.25/kWh	0.177
5hr	£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.151
5hr	£0.25/kWh + Down Events	Free Hours (Volumetric)	0.421
5hr	£0.25/kWh + Down Events	Free Hours (Consistency)	0.311
5hr	£0.25/kWh	£0.20/kWh + Community Pot	0.842

Notice period	Trial arm (1)	Trial arm (2)	p
5hr	£0.25/kWh	Free Hours (Volumetric)	0.577
5hr	£0.25/kWh	Free Hours (Consistency)	0.735
5hr	£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.477
5hr	£0.20/kWh + Community Pot	Free Hours (Consistency)	0.611
5hr	Free Hours (Volumetric)	Free Hours (Consistency)	0.828
DA	£0.25/kWh + Down Events	£0.25/kWh	0.378
DA	£0.25/kWh + Down Events	£0.20/kWh + Community Pot	0.162
DA	£0.25/kWh + Down Events	Free Hours (Volumetric)	0.215
DA	£0.25/kWh + Down Events	Free Hours (Consistency)	0.889
DA	£0.25/kWh	£0.20/kWh + Community Pot	0.584
DA	£0.25/kWh	Free Hours (Volumetric)	0.695
DA	£0.25/kWh	Free Hours (Consistency)	0.451
DA	£0.20/kWh + Community Pot	Free Hours (Volumetric)	0.883
DA	£0.20/kWh + Community Pot	Free Hours (Consistency)	0.199
DA	Free Hours (Volumetric)	Free Hours (Consistency)	0.262

Table 25: Peak kW/h balance check - Notice period pairwise comparisons within treatment group

Treatment group	Notice period (1)	Notice period (2)	p
£0.25/kWh + Down Events	2hr	5hr	0.416
£0.25/kWh + Down Events	2hr	DA	0.609
£0.25/kWh + Down Events	5hr	DA	0.769
£0.25/kWh	2hr	5hr	0.988
£0.25/kWh	2hr	DA	0.841
£0.25/kWh	5hr	DA	0.854
£0.20/kWh + Community Pot	2hr	5hr	0.331
£0.20/kWh + Community Pot	2hr	DA	0.228
£0.20/kWh + Community Pot	5hr	DA	0.897
Free Hours (Volumetric)	2hr	5hr	0.965
Free Hours (Volumetric)	2hr	DA	0.42
Free Hours (Volumetric)	5hr	DA	0.448
Free Hours (Consistency)	2hr	5hr	0.466

Treatment group	Notice period (1)	Notice period (2)	p
Free Hours (Consistency)	2hr	DA	0.861
Free Hours (Consistency)	5hr	DA	0.56

6.3 Statistical power

6.3.1 Turn-down events

In the previous winter trial, turn-down incentives ranged from £0.20/kWh to £1.25/kWh, with a medium incentive of £0.50/kWh. The corresponding treatment effect at the medium incentive was approximately −0.020 kWh per event. The standard error for pooled comparisons across treatment arms (each with ~9,100 participants) was 0.002 and pairwise treatment comparison was 0.003, based on 39 observed events.

In the Summer 2025 trial, each arm had 13,500 participants, with 12 events scheduled during the summer time, which usually means lower variance of electricity consumption at the household level. Adjusting for the increase in sample size, the new standard errors were approximately 0.0016 for pooled regression and 0.002 for pairwise treatment comparison. If we adjust the standard errors similarly for the decrease in events, we obtain a new standard error of 0.003 and 0.004. However, since statistical power is more sensitive to total participants than event counts, we assume that the overall increase in participants will contribute more than the decrease in number of events. Nonetheless, with an expected effect size of −0.020 kWh and at most SE of 0.004 , power remains very high for detecting the primary treatment effect, even in individual pairwise comparisons.

6.3.2 Turn-up events

In the previous design, turn-up incentives ranged from £0.10 to £0.25, with two further groups who had per kWh incentives accompanied by a consistency bonus. A £0.20 incentive alone produced a treatment effect of +0.034 kWh, while £0.20 with a consistency bonus produced +0.049 kWh, resulting in a pairwise difference between treatments of 0.015 kWh. Standard errors for pairwise comparisons in that trial were approximately 0.003.

In the trial, the incentive levels were set at a direct £0.25/kWh or the approximate equivalent thereof, but with differing incentive structures; these treatments were expected to yield similar or slightly larger treatment effects. With a sample size of 13,500 per arm and no significant change in the number of events (a total of 48 events), the SE remained approximately 0.002 for comparison with the control and 0.003 for pair-wise treatment comparison. This ensured high power for detecting both individual treatment effects and pairwise differences as small as 0.015 kWh.

6.4 Supporting tables

6.4.1 Primary analysis

Effect of interventions' on consumption during turn-up events

Table 26: Effect of incentives on electricity consumption during turn-up events

Trial arm	Avg.	Coefficient	SE	P value	% difference
Control	0.304				
£0.25/kWh	0.366	0.059	0.003	<0.001***	19.55%
£0.25/kWh + Down Events	0.353	0.050	0.003	<0.001***	16.29%
£0.20/kWh + Community Pot	0.361	0.055	0.003	<0.001***	18.08%
Free Hours (Volumetric)	0.355	0.050	0.002	<0.001***	16.39%
Free Hours (Consistency)	0.358	0.055	0.002	<0.001***	17.98%

Effect of dual event exposure on turn-up performance

Table 27: Comparison of turn-up' performance with and without exposure to turn-down events

Trial arm	Avg.	Coefficient	SE	P-value	% difference
£0.25/kWh + Down Events	0.353				
£0.25/kWh	0.366	0.010	0.003	0.0016**	2.76%

Effect of interventions' on consumption during turn-down events

Table 28: Effect of incentives on electricity consumption during turn-down events

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Control	0.400				
£0.25/kWh + Down Events	0.392	-0.006	0.003	0.0263*	-1.43%

Effect of notice period on consumption during turn-up events

Table 29: Effect of notice period on electricity consumption during turn-up events

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Control	0.304				
2 hours	0.350	0.045	0.002	<0.001***	14.94%
5 hours	0.365	0.059	0.002	<0.001***	19.49%
DA	0.361	0.056	0.002	<0.001***	18.55%

Effect of notice period on consumption during turn-down events

Table 30: Effect of notice period on electricity consumption during turn-down events

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Control	0.400				
2 hours	0.396	-0.001	0.004	0.7939	-0.23%
5 hours	0.396	-0.004	0.004	0.2507	-1.08%
DA	0.385	-0.012	0.004	0.0012**	-2.95%