Public

CrowdFlex report: Utilisation trial

Winter 2025

May 2025

Version: 1.0:

Written by: Centre for Net Zero

Reviewers: CrowdFlex consortium partners

Final Review Date: 12/06/2025

Approval Date: 30/06/2025

NESO
National Energy System Operator

Executive summary

Demand flexibility, whereby consumers adjust their electricity consumption based on external signals related to price, carbon intensity, or grid constraints, can play an important part in balancing the electricity grid. This report presents the results of a randomised controlled trial testing the extent to which different financial incentives shift electricity consumption during specific time windows. It includes events which incentivise consumers to turn down, as well as turn up, electricity use in response to grid conditions. The focus of the trial is important in the context of the UK government's 2030 Clean Power target, which includes an ambition to increase levels of demand flexibility in the system four- to five-fold, in just five years.

The trial involved over 55,000 participants and took place across 76 events from October 2024 to March 2025. It was delivered by OVO Energy, with trial design and analysis led by Centre for Net Zero as part of CrowdFlex. CrowdFlex is a National Energy System Operator (NESO) led Innovation project delivered by NESO, OVO, Ohme, Centre for Net Zero, ERM, AWS, National Grid Electricity Distribution, and Scottish and Southern Electricity Networks, with the support of Smart Grid Consultancy, CGI, Smith Institute, and the Centre for Sustainable Energy. It is funded by energy network users and consumers through the Strategic Innovation Fund, a programme from the UK's independent energy regulator Ofgem managed in partnership with Innovate UK.

Customers were randomly assigned to one of six groups: a control group, three groups with varying incentive levels, and two additional groups where incentives per kWh were combined with a £3/month consistency bonus. The consistency bonus was awarded to participants achieving positive response in at least 50% of the events. This randomised controlled trial design ensures that, on average, observed and unobserved characteristics are balanced across groups. As a result, differences in outcomes can be causally attributed to the interventions rather than to pre-existing differences between customers.

Executive Summary Table 1: Trial arms

Trial Arms	Incentive levels	
	Turn-up	Turn-down
Control Group (no message)	n/a	n/a
Low	£0.10	£0.20

Trial Arms	Incentive levels	
	Turn-up	Turn-down
Medium	£0.15	£0.50
High	£0.25	£1.25
Low + Consistency Bonus	£0.10 + CB	£0.20 + CB
Medium + Consistency Bonus	£0.15 + CB	£0.50 + CB

The trial's key findings were:

1. **Financial incentives changed electricity use in the desired direction.** All treatment effects were statistically significant, showing that financial incentives effectively change electricity use during events. Overall effect sizes for turn-up events were more than double those for turn-down events.

2. **Higher incentive levels and the addition of the consistency bonus both increased demand response.** The High Incentive group showed a similar response to the Consistency Bonus groups. Both groups increased turn-down compared to the Low Incentive group and increased turn-up compared to the Low and Medium Incentive groups. This points to higher utilisation prices being as effective as a separate financial bonus. Lower incentives delivered more flexibility per pound spent, while higher incentives delivered greater volumes, pointing to a trade-off between maximising response and minimising cost.

3. **We found little evidence of demand displacement.** Electricity use outside of event windows appeared to remain consistent across all groups, which would suggest observed changes represent demand destruction or demand creation to align with balancing needs during flexibility windows. However, it is also possible that changes in demand outside events were too diffuse to detect in our sample or occurred outside of the observation window.

4. **Exploratory analysis shows varying effects based on customer and event characteristics.** While not causal, this early analysis identified several promising areas for further investigation to help inform future service design.

The table below summarises the results underpinning these findings. It is important to note that the figures presented in this table reflect the average change in consumption across each group. Within each cohort, however, individual responses likely varied considerably: some customers actively reduced usage, others may have shown little change, and a subset may have increased usage. As a result, the observed average effects may understate the magnitude of response among the most engaged participants due to these offsetting behaviours within the overall portfolio.

Executive Summary Table 2: Key findings

Finding	Results
1. Financial incentives changed electricity use	• Overall, turn-down events led to a 4.1% reduction • Overall, turn-up events led to a 11.1% increase
2. Higher incentives and the addition of a consistency bonus increased response	**Turn-down**: • Low Incentive: -3.2% (95% confidence interval: -4.3% to -2.0%) • Medium Incentive: –3.8% (-4.9% to -2.6%) • High Incentive: –4.3% (-5.4% to -3.1%) • Low + Consistency Bonus: –5.1% (-6.2%, to -3.9%) • Medium + Consistency Bonus: –4.3% (-5.5% to -3.2%) Note: Differences between the high incentive group and Consistency Bonus groups were not statistically significant, and none of the medium-level incentive groups (with or without CB) significantly outperformed one another - indicating diminishing returns or equivalence at moderate incentive levels. **Turn-up**: • Low Incentive: +8.2% (6.85% to 9.44%) • Medium Incentive: +10.1% (-1.66% to 24.13%) • High Incentive: +12.8% (-3.85% to 29.51%) • Low + Consistency Bonus: +11.6% (-2.42% to 25.69%) • Med + Consistency Bonus: +12.7% (-1.40% to 26.71%) Note: Most comparisons showed significant differences, except when comparing High Incentives to either Low + CB or Med + CB.
3. Little evidence of displacement	**We found no evidence of displacement in either turn-down or turn-up events**: all treatment groups had consumption levels outside of events comparable to the control group.

Exploratory findings investigated how treatment effects varied based on customer and event characteristics. The results include:

- **Critical peak pricing**: elevated incentive levels (£1.75/kwh) led to a further 2.1% reduction (on top of the 3.1% reduction from turn-down events during a similar period of the day).
- **Notice period**: short notice saw slightly less turn-up (and had no statistically significant effect on turn-down).
- **Event duration**: short events saw no change in turn-down, but slightly more turn-up per half-hour.
- **Type of day**: weekdays and weekends saw no difference in effects.
- **Ovo Power Move**: greater response from those enrolled in this other demand response programme.
- **Auto-enrolment**: no difference in effects between new recruits and those auto-enrolled from previous trial.
- **Estimated annual consumption**: higher consumers had greater change in kW, but not as a proportion of consumption.
- **Month-to-month effects**: while effect sizes varied, they were broadly consistent across the trial, suggesting no fatiguing effects.
- **Tariff type**: customers on time-of-use tariffs (Economy 7) had greater response than those on standard tariffs.
- **Flexibility value**: lower incentives had lower costs per unit of flexibility on a £/kWh basis, while higher-cost incentives delivered greater overall volumes.
- **Price elasticity**: based on estimates from the price levels tested, turn-down events had relatively low price elasticities, while turn-up events had higher elasticities with a larger proportional demand response to incentive levels.

In conclusion, the trial demonstrates that financial incentives can effectively change electricity use in response to grid conditions, with particularly strong response to turn-up events. Higher incentives, whether as direct additional £/kWh or as a consistency bonus, led to stronger responses and higher volumes of flexibility when compared with the Low Incentive treatment group, but will need to be balanced against budgetary constraints of future programmes. Flexibility events appear to deliver demand destruction and creation in response to grid balancing needs, although it is possible demand changes outside events are too diffuse to detect. Survey evidence indicates that customers report relatively low use of automated tools to respond, suggesting primarily manual response. Overall, these findings are important for programme and policy design for demand response, as system operators and suppliers consider how to scale the procurement of demand flexibility to help balance the future low-carbon electricity system.

1. Background

With increasing reliance on variable renewable supply such as wind and solar, the transition to a low-carbon electricity system requires greater flexibility in demand to balance the electricity grid. Demand flexibility, where consumers adjust their electricity consumption based on external signals related to price, carbon intensity, or grid constraints, can help reduce peak demand, balance supply and demand, mitigate system constraints, and support the integration of renewables. The government has identified demand flexibility as a critical part of its Clean Power Action Plan, following advice from NESO, with an ambition to increase levels four- to five-fold in just five years.[1]

Explicit demand flexibility (i.e., flexibility dispatched by grid operators) is an important part of this, providing demand-side mechanisms for consumers to change electricity use in response to requests aligned with system needs. The use of flexibility services has increased in recent years, but we need strong evidence to understand the effectiveness of flexibility events to inform their future design.

The Utilisation Trial was undertaken by OVO Energy (OVO) and ran from October 2024 to March 2025. It was designed to assess how financial incentives impact consumer participation and response in flexibility events. To do so, the trial examines whether varying levels of payments per kilowatt-hour (kWh) and the introduction of a consistency bonus encourage households to change electricity consumption during designated event windows.

To rigorously evaluate these effects, the trial was structured as a randomised controlled trial (RCT), randomly assigning participants to different incentive structures. This method allows for a robust comparison between treated and untreated households, ensuring that observed differences in behaviour can be attributed to the incentives rather than external factors. By examining the causal impact of different incentives, the study strengthens the reliability of insights for future demand-side response programs.

[1] Department for Energy Security and Net Zero, Clean Power Action Plan, December 2024.

2. Research questions and hypotheses

This study investigates how different financial incentives influence household participation in demand flexibility events (Table 1). Beyond direct participation in an event, the analysis explores whether treatments impact consumption patterns outside event windows, lead to longer-term behavioural changes, or vary in effectiveness based on household characteristics. Additionally, we assess how event design factors, such as timing, frequency, and notice periods, shape response rates.

The following research questions and hypotheses provide a structured framework for evaluating these effects.

Table 1: Research questions and hypotheses

Category	Research Question	Hypotheses
Primary	To what extent do different monetary incentive structures influence domestic energy consumers' participation in turn-up and turn-down events?	Higher incentives lead to greater turn-down and turn-up.
		Consistency bonuses lead to greater turn-down and turn-up.
		Higher incentives and consistency bonuses interact, with the combination leading to greater turn-down and turn-up than either factor alone.
Secondary	To what extent does being treated (as opposed to the control group) cause domestic customers to turn-down or up their electricity consumption outside event windows (same day, lead-up, post-event)?	Turn-up events result in both new demand creation and demand displacement. That is, in turn-up events, the net increase in demand during event hours exceeds the net reduction in demand outside event hours.
		Turn-down events result in both pure demand destruction and demand displacement. That is, in turn-down events, the net decrease in demand during event hours exceeds the net increase in demand outside event

Category	Research Question	Hypotheses
		hours.
		Behavioural spillover effects occur, where customers adjust consumption patterns at other times.
Exploratory	How do customers respond to "critical peak events" where payments (for all groups) are higher? With this we can answer whether there are any carryover effects across different levels of incentives (e.g., anchoring).	Higher payments will increase participation across all groups.

2.1 Additional exploratory analysis

We will examine how behaviours vary across the below event parameters. These factors, which differ between events but are not randomly assigned, include:

- **Effect of notification period** – investigating whether the amount of advance notice affects engagement with events.
- **Effect of event duration** – evaluating whether longer or shorter events lead to greater responsiveness.
- **Effect of day type** – comparing responsiveness on weekdays versus weekends to identify behavioural differences.
- **Effect of being on OVO's Power Move** – assessing whether customers enrolled in the OVO Power Move scheme exhibit different engagement patterns compared to those who are not.
- **Effect of auto-enrolment (summer trial participants vs. new winter trial recruits)** – exploring whether customers automatically enrolled from the summer trial behave differently from those who were freshly recruited for the winter trial.
- **Effect of estimated annual consumption (EAC)** – investigating whether households with higher or lower annual electricity consumption respond differently to incentives.
- **Month-to-month treatment effects** – evaluating whether performance changes over time.

- **Effect of tariff type (with add-on vs. without add-on)** – comparing customers with OVO add-ons to their tariff to those on standard tariffs to assess whether tariff structures influence responsiveness.

These additional analyses will provide deeper insights into the contextual factors that shape consumer engagement with demand flexibility initiatives.

3. Methodology

3.1 Utilisation payments trial design

The trial follows a randomised control trial (RCT) design, with approximately 55,000 of OVO's domestic energy customers participating. Customers were randomly assigned to one of six trial arms, including a control group. A small but important deviation from pure randomisation is that customers who were in the control group during the summer utilisation payments trial were randomly assigned to one of the five treatment groups, ensuring they were not placed in the control group a second time.

Each treatment group experienced different levels of monetary incentives, with two groups also exposed to consistency bonuses. The trial ran from October 2024 to March 2025 and consisted of multiple flexibility events, where participants were encouraged to adjust their electricity consumption in response to price signals.

3.2 Participants

Participant recruitment was conducted through auto-enrolment of non-EV add-on customers who had previously engaged in the summer utilisation trial. EV add-on customers, instead, were allocated to the winter Availability Payments Trial, which ran parallel to this winter Utilisation Payments Trial. Additional participants were recruited through direct invitations, ensuring a diverse and representative sample. The final recruited sample consists of 55,097 customers, including 1,442 prepay meter customer (PAYGO) participants.

3.3 Interventions

The incentive structure comprised three levels of utilisation payments, ranging from low to high, with two groups also receiving an additional consistency bonus (Table 2). Customers in the control group were not exposed to these financial incentives and did not receive messaging about events at all. They served as a baseline for comparison. A subset of participants included PAYGO, who were only invited to participate in turn-down events to mitigate financial risk to customers and risk of potential supply disruptions.

Table 2: Trial arms

Trial Arms	Incentive levels	
	Turn-up	Turn-down
Control Group (no message)	n/a	n/a

Trial Arms	Incentive levels	
	Turn-up	Turn-down
Low	£0.10	£0.20
Medium	£0.15	£0.50
High	£0.25	£1.25
Low + Consistency Bonus	£0.10 + CB	£0.20 + CB
Medium + Consistency Bonus	£0.15 + CB	£0.50 + CB

Each Consistency Bonus group had the potential to earn a consistency bonus of £3 per month.

Consistency bonus criteria: To qualify for the consistency bonus, customers must meet the following criteria:

- Turn-up events: Customers must increase their energy usage above their P376 baseline during the event.
- Turn-down events: Customers must reduce their energy usage below their P376 baseline during the event.
- Customers must successfully increase (if turn-up) / reduce (if turn-down) their consumption in 50% of the events each month to qualify for the consistency bonus.

All treatment groups were also exposed to critical peak events, where customers earned £1.75/kWh. This higher incentive level was intended to test consumer elasticity under peak pricing conditions and provide insights into the upper limit of price responsiveness at times of simulated critical stress (where OVO gave customers only 2-hour notice).

Note: Customer baselines were estimated in accordance with [Balancing and Settlement Code P376](), which provides the regulatory framework for how Demand Side Response Service Providers (DSRSPs) should determine baseline consumption. Performance was considered positive when a customer's consumption during an event was lower than their P376 baseline during turn-down events, or higher during turn-up events. In brief, OVO calculated the baseline by averaging consumption across the same half-hour period over the ten most recent working days; or four non-working days (depending on whether the event happened on a working or non-working day); skipping days when there had been a CrowdFlex or DFS event; and excluding the highest and lowest values (leaving eight working days or two non-working days). OVO applied a "clipping" rule whereby shifts in the incorrect direction were set to zero, thereby avoiding penalisation of customers for adverse deviations. Note that the demand shift identified in this report are from comparisons to the control group, and so these results are not affected by "clipping" or other decisions regarding how to use a customer's own baseline.

3.4 Outcome measures

The primary outcome of interest is the impact of financial incentives on participants' electricity consumption (kWh per half-hour) during event windows.[2]

The secondary outcome examines spillover effects, assessing whether participants alter their electricity consumption (kWh) outside of event windows.

Note: Electricity consumption was analysed at the half-hourly level (in kWh). However, for ease of interpretation, results are presented at the hourly level; that is, values are scaled by a factor of two (2 × kWh/half-hour), equivalent to average power in kilowatts (kW).

3.5 Event parameters and variability

The Utilisation Payments Trial events were structured across different Electricity Forward Agreement (EFA) blocks to assess demand flexibility at different times of the day. The trial was designed to capture variability in domestic manual flexibility based on underlying grid conditions, consumer responsiveness, and potential system constraints. By distributing events across multiple EFA blocks, the study provides insights into how the effectiveness of financial incentives varies with time of day (Table 3).

3.5.1 EFA block prioritisation and event distribution

- **Turn-up events**: Prioritised in EFA 4 (11:00-15:00) and EFA 6 (19:00-23:00), where a higher number of turn-up targets were set. This structure allows for testing demand flexibility during mid-day and evening periods when system balancing may be needed.
- **Turn-down events**: Concentrated in EFA 3 (07:00-11:00), EFA 4 (11:00-15:00), and EFA 5 (15:00-19:00), with a particularly high focus in EFA 5 (22 turn-down targets). These periods align with system peak times, where reducing consumption can

[2] This means that longer events influence the results "more" than shorter events because they contribute more observations.

alleviate grid constraints. Note that all critical peak events were undertaken during EFA 5.

- **Anti-symmetric events:** Used selectively in EFA 3 (07:00-11:00) and EFA 4 (11:00-15:00), where Scotland was targeted for turn-up while the rest of Great Britain was targeted for turn-down. This structure allows for assessing regional variations in response to demand flexibility signals.

This event structure ensures that demand flexibility is tested across a range of system conditions, allowing for a detailed evaluation of how households respond to financial incentives at different times of the day.

Table 3: The distribution of utilisation trial events by time of day and direction.

At	EFA	Up target	Down target	Anti target
23:00	EFA 1	4	1	0
03:00	EFA 2	3	0	0
07:00	EFA 3	2	7	1
11:00	EFA 4	9	5	4
15:00	EFA 5	2	22	0
19:00	EFA 6	7	4	2

3.6 Sample size

The trial initially aimed to recruit 55,000 participants from OVO's domestic customers. Following two phases of randomisation, the final sample included 55,109 participants. The study's sample size ensures sufficient statistical power to detect meaningful differences in consumer behaviour across experimental groups. The large sample helps improve external validity, making the findings more generalisable.

3.7 Randomisation

Randomisation was performed in two stages. The first round in October 2024 assigned 49,000 participants, while an additional 6,000 were assigned in November 2024, once the trial had begun. Customers were allocated across the six trial arms to ensure balanced trial arm sizes. (OVO checked balance of treatment on key observables; see our appendix section Balance Checks for details.)

The randomisation process ensures that treatment and control groups are comparable, reducing potential confounding factors. Random assignment of participants allows for a robust evaluation of treatment effects by mitigating biases that might arise from self-selection into flexibility schemes. However, we acknowledge that the sample is itself composed of customers who have actively signed up to participate, and they may differ from the general population.

3.8 Analysis methods

To assess the impact of the interventions on electricity consumption, regression models were employed. The analyses are divided into primary, secondary, and exploratory components, each with different comparison structures.

3.8.1 Primary analysis specification

The primary analysis focused on estimating the overall treatment effects of the interventions by comparing:

- **Treatment vs. control (during events):** Measuring the impact of different interventions on electricity consumption during event periods.
- **Treatment vs. treatment (during events):** Comparing the relative effects of different treatment groups to determine whether varying incentive levels or introducing consistency bonuses led to statistically significant differences in electricity consumption.

We used linear regression models to estimate the effects of the interventions on electricity consumption, controlling for a range of customer characteristics and time-fixed effects, as pre-specified in our trial design summary. Specifically, we included controls for customer region (GSP Group), projected annual consumption, current and past participation in Power Move, auto-enrolment status, tariff type (Economy 7, fixed tariff, or pay-as-you-go), tenure with OVO (in years), and whether the customer receives (SMS) notifications. We also included fixed effects for the date and time of each observation to adjust for systematic differences by time of day and day of week. All predictors were included as main effects without interactions. Standard errors were clustered at the customer level to account for repeated measurements over time. We summarise regression results using the following statistical measures:

- **Coefficient:** Measuring the magnitude of the intervention effect relative to the control group or other treatment groups.
- **Standard error:** Indicating the precision of the estimated effect.
- **P-value:** Assessing statistical significance to determine whether observed differences are likely due to chance. Note we use the following notation to mark p-values: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$

3.8.2 Secondary analysis

The secondary analysis focused on Treatment vs. Control comparisons of electricity consumption *outside* of event periods, defined as consumption non-event half-hours on the day of events.

This analysis identified whether the interventions had spillover effects or broader behavioural impacts beyond the designated event windows. For example, did participants alter their typical electricity consumption patterns even when no specific event was occurring?

Linear regression models were employed, with results summarised using coefficients, standard errors, and p-values.

3.8.3 Exploratory analysis

Exploratory analyses were conducted to gain additional insights into the factors influencing electricity consumption beyond the structured comparisons of the primary and secondary analyses. Unlike the primary and secondary analyses, the exploratory analyses are not based on randomisation and should be interpreted as indicative rather than causal. The findings are intended to uncover patterns and relationships that may inform future research rather than provide definitive conclusions about causal effects.

The exploratory analysis examined various factors and are outlined in section 2 above.

Linear regression models were used to estimate the impact of these factors on average kW hourly consumption during events. The results are summarised using differences in average kW hourly consumption, regression coefficients, standard errors, and p-values.

3.8.4 Sensitivity analysis

Additional robustness checks were conducted to assess the stability of results across alternative model specifications and subgroups. Specifically, the appendix presents results from both the primary and secondary analyses without covariate adjustments, as well as from the primary analysis with critical pricing events excluded.

3.9 Balance checks

Balance checks were conducted for the trial to ensure baseline comparability across the trial arms. These checks aimed to validate the randomisation process and confirm that pre-existing differences in electricity consumption or group composition would not confound the trial's results.

The following baseline variables were tested:

- Estimated Annual Consumption (EAC): Mean, standard deviation, minimum, and maximum.
- Peak energy use: Average electricity consumption during peak periods.
- Off-peak energy use: Average electricity consumption during off-peak periods.

Pairwise t-tests were used to compare these variables between groups (a statistical test that evaluates whether the means of two groups are significantly different), with a p-value threshold of 0.05 used to determine statistical significance.

4. Results

4.1 Balance check results

No statistically significant differences were observed between trial groups in either rounds of randomisation. The distribution of values remains balanced across groups, ensuring that randomisation was successful and that there is no indication of systematic bias in allocation. Please see the appendix for more details on the results.

4.2 Participant flow and dropout analysis

4.2.1 Recruitment and allocation

A total of 55,097 participants were initially enrolled in the OVO utilisation trial across six trial arms. Of these, 48,772 participants remained in the trial representing an overall dropout rate of 11.48% (Table 4). Analysis of dropout by trial arm indicates that the randomisation and allocation process was robust, with minimal differences in attrition across the groups. This indicates that the treatments had minimal effect on dropout.

Table 4: Participant dropout rates

Group	# start	# end	% decrease
1 - Control	9,214	8,296	9.96%
2 - Low (no CB)	9,180	8,066	12.14%
3 - Med (no CB)	9,167	8,098	11.66%
4 - High (no CB)	9,186	8,146	11.32%
5 - Low + CB	9,177	8,046	12.32%
6 - Med + CB	9,173	8,120	11.48%
Total	55,097	48,772	11.48%

4.2.2 Dropout reasons

The primary reasons for participant dropout were:

- Account not live (8.5%) – Participants whose accounts became inactive due to change of supply or property occupancy at any point during the trial.
- Smart meter issues (1.7%) – Participants whose smart meters experienced technical issues, preventing DSRSP access to half-hourly consumption data.
- Customer trial opt-out (1.3%) – Participants who withdrew after initial engagement with the trial.

4.2.3 Impact on trial execution

The dropout rate was broadly consistent across trial arms, suggesting relatively small attrition bias that could impact the validity of the trial outcomes. The dropout percentages remained within expected ranges, ensuring that statistical power was retained for analysis.

4.3 Primary analysis results

The primary analysis estimated the overall treatment effects of the interventions during event periods. By comparing treatment groups against the control group (treatment vs. control) and evaluating differences between treatment groups themselves (treatment vs. treatment), this analysis sought to identify the causal impact of interventions on electricity consumption. Randomisation was employed to ensure unbiased estimates of treatment effects.

4.3.1 Pooled treatment effect on electricity consumption during events

This section presents the pooled effect of treatment on electricity consumption during turn-down and turn-up events. Across both event types, the treatment led to statistically significant changes in consumption relative to the control group.

During turn-down events, the pooled treatment group (i.e., all treatment groups treated as one large group) consumed less electricity than the control group. Average electricity consumption was 0.5128 kW in the treatment group compared to 0.5329 kW in the control group (Table 5; Figure 1). The regression-adjusted treatment effect was −0.02198 kW, a 4.12% reduction (SE = 0.00246, $p < 0.001$), indicating a measurable reduction in consumption during these events.

During turn-up events, the treatment group consumed more electricity than the control group. The treatment group averaged 0.4652 kW compared to 0.4184 kW in the control group (Table 6; Figure 2). The regression-adjusted effect was +0.04634 kW, an 11.08%

increase (SE = 0.002, $p < 0.001$), confirming a statistically significant increase in usage (Table 6; Figure 2).

These results show that the treatment successfully changed electricity usage in the expected direction: decreasing during turn-down events and increasing during turn-up events. This adds to the limited evidence base for the causal impact of flexibility events on consumer response. The findings for turn-up events are particularly novel, with very little research to date on their efficacy in absorbing excess supply.

For turn-down events, the results confirm those seen in previous studies (see Fowlie et al., 2021[3], Schofield et al., 2014[4], and Jacob et al., 2024)[5]. CrowdFlex appears to be consistent with this literature suggesting that consumers respond to higher financial incentives.

[3] Fowlie, M., Wolfram, C., Baylis, P., Spurlock, C. A., Todd-Blick, A., & Cappers, P. (2021). Default effects and follow-on behaviour: Evidence from an electricity pricing program. *The Review of Economic Studies*, *88*(6), 2886-2934.

[4] Schofield, J., Carmichael, R., Tindemans, S., Woolf, M., Bilton, M., & Strbac, G. (2014). Residential consumer responsiveness to time-varying pricing: Report A3 for the "Low Carbon London" LCNF project.

[5] 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.

Figure 1: Bar chart of pooled treatment effects on hourly electricity demand (kW) during "down" and "up" events

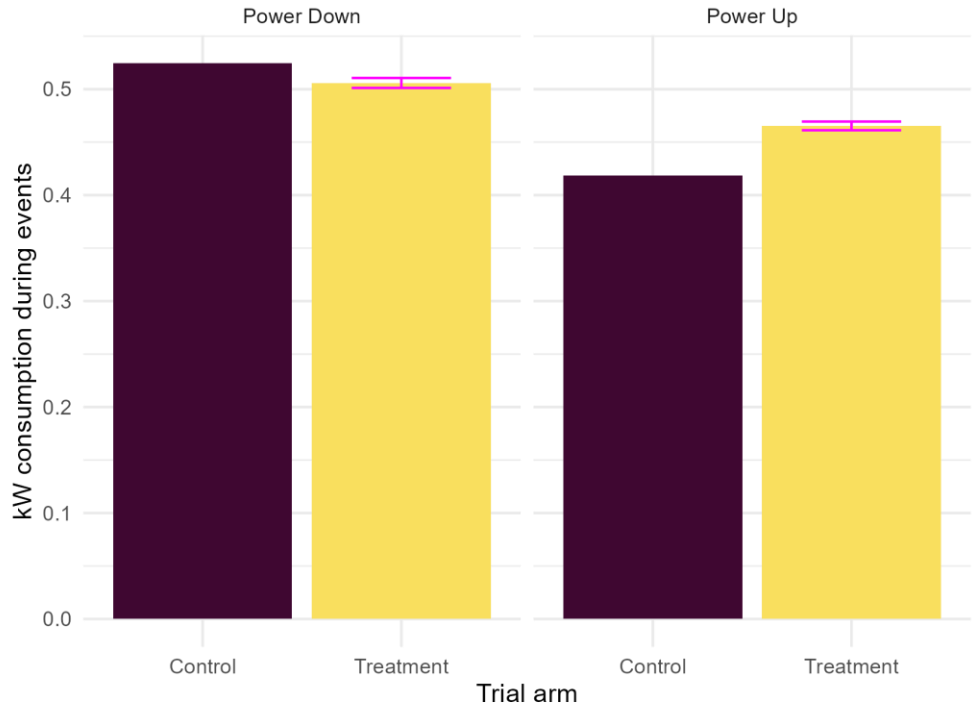


Figure 2: Coefficient plot of pooled treatment effects on hourly electricity demand (kW) during "down" and "up" events

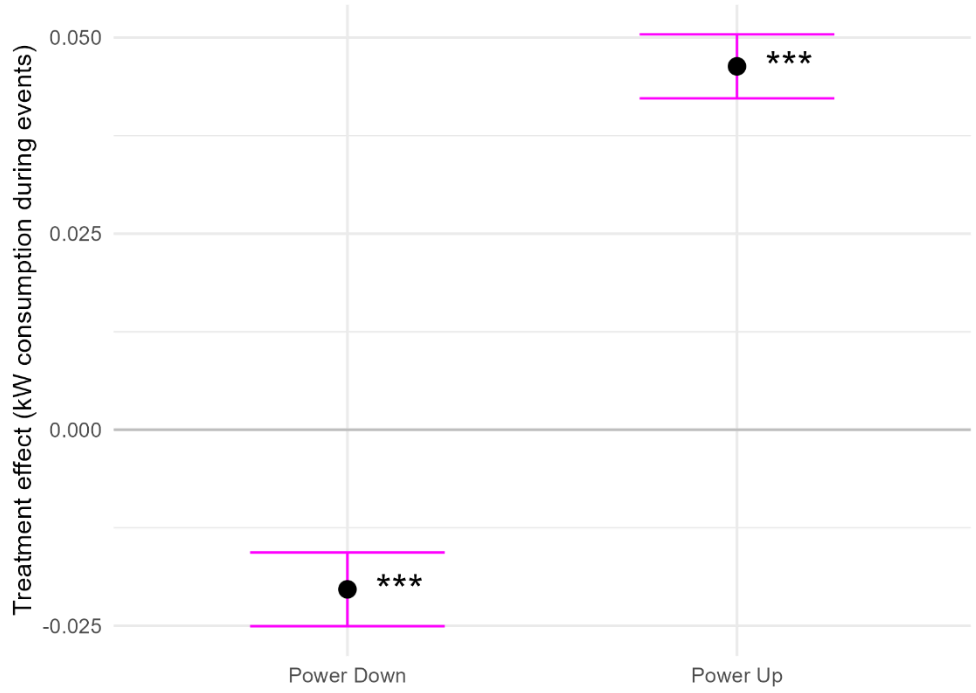


Table 5: Pooled treatment effect on hourly demand (kW) during turn-down events

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value	% difference (regression coefficient vs. control)
Control	0.533				
Treatment	0.513	-0.022	0.002	<0.001***	-4.12%

Table 6: Pooled treatment effect on hourly demand (kW) during turn-up events

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value	% difference (regression coefficient vs. control)
Control	0.418				
Treatment	0.465	0.046	0.002	<0.001***	11.08%

4.3.2 Turn-down events

4.3.2.1 Difference between treatment groups and control

All treatment groups showed lower electricity demand during turn-down events compared to the control group (0.5329 kW) (Table 7; Figure 3, Figure 4).

We also compared treatment groups to each other, as discussed in the next section (where we show pairwise comparisons and statistical significance testing in the next section).

The Low + Consistency Bonus group (0.5096 kW) had the largest reduction relative to the control group (−0.0269 kW, -5.05%), followed by Medium + Consistency Bonus (−0.0231 kW, -4.34%), and High (−0.02286 kW -4.29%). Note that we show pairwise comparisons and statistical significance testing in the next section.

The Low and Medium groups also showed significant reductions in consumption compared to control, with regression coefficients of −0.01694 and −0.02012, respectively. However, these reductions were smaller than those observed in the High and Consistency Bonus groups.

Overall, the results show that both increased incentive levels and the addition of a consistency bonus led to lower consumption during turn-down events. The three groups with the greatest reductions – High, Low + Consistency Bonus, and Medium + Consistency Bonus – had similar average consumption levels and regression-adjusted effects.

Figure 3: Bar chart of treatment effects on hourly demand (kW) during turn-down events

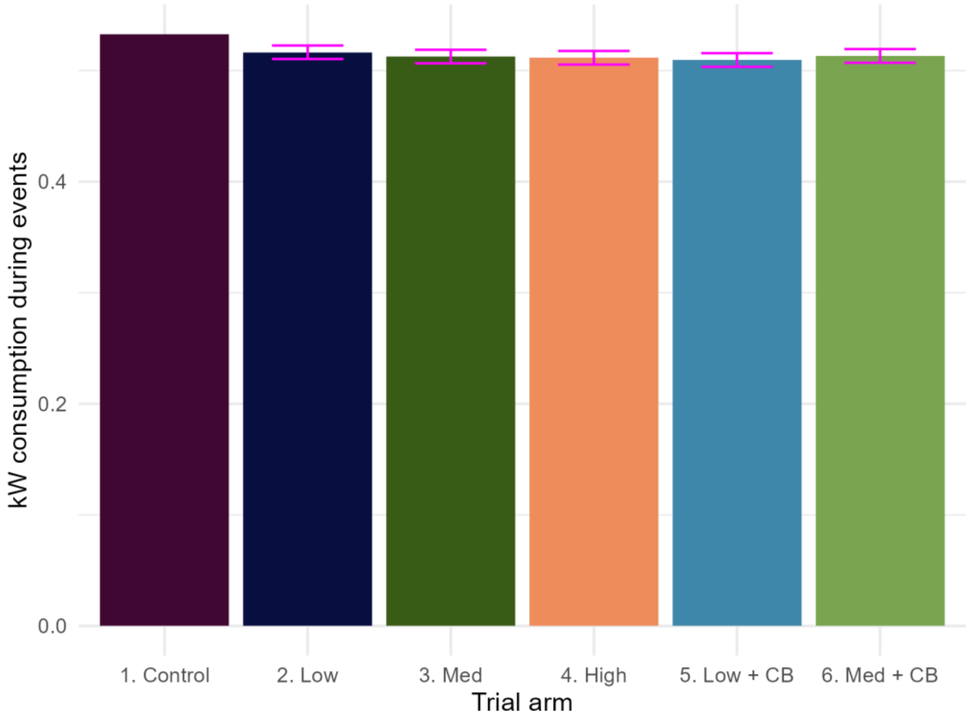
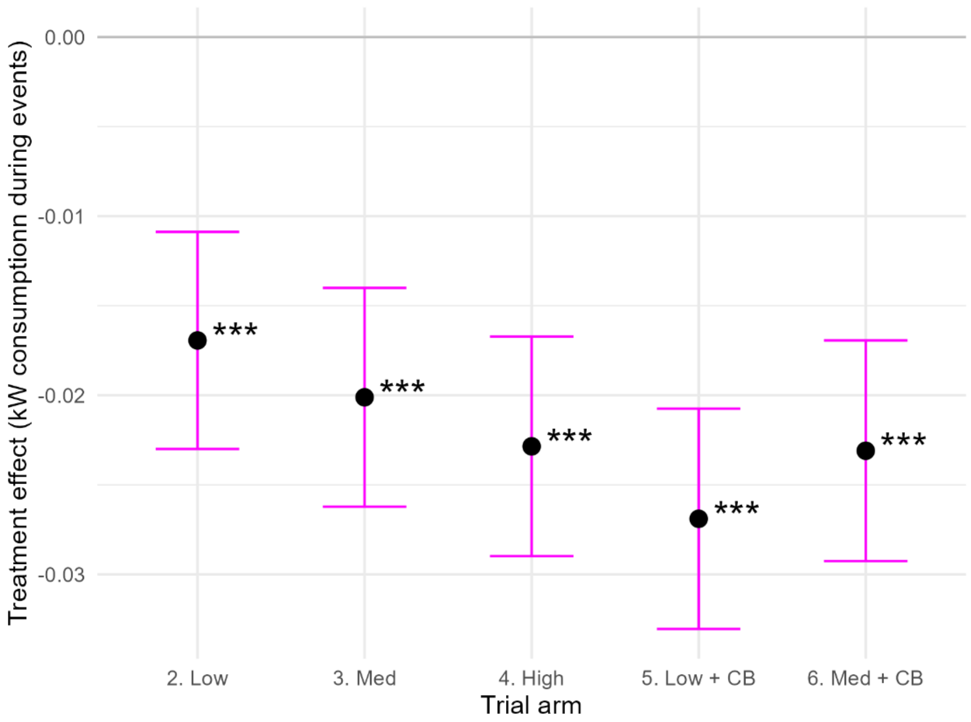


Figure 4: Coefficient plot of estimated treatments effects on hourly demand (kW) during turn-down events



*Note: 95% confidence intervals are indicated in pink. Significance levels: *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.10.*

Table 7: Treatment effects on hourly demand (kW) during turn-down events

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value	% difference (regression coefficient vs. control)
Control	0.533				
Low	0.517	-0.017	0.003	<0.001***	-3.18%
Medium	0.513	-0.020	0.003	<0.001***	-3.78%
High	0.512	-0.023	0.003	<0.001***	-4.29%
Low + Consistency Bonus	0.510	-0.027	0.003	<0.001***	-5.05%
Medium + Consistency Bonus	0.513	-0.023	0.003	<0.001***	-4.34%

4.3.2.2 Treatment versus treatment analysis

Pairwise comparisons across treatment arms suggest subtle but meaningful differences in electricity consumption behaviour during turn-down events. The High Incentive group and both Consistency Bonus groups had greater turn-down than the low-incentive group. The difference in response between the High Incentive group and both Consistency Bonus groups was not statistically significant.[6]

Table 8: Pairwise treatment comparisons for turn-down events

Group 1	Group 2	Avg. kW per hour (G1)	Avg. kW per Hour (G2)	% difference (vs. G2)	Regression Coefficient (vs. G2)	Standard Error	P-value
Low + CB	Low	0.510	0.517	-1.95%	**-0.010**	0.002	0.001**
Low + CB, Med + CB	Low, Med	0.512	0.514	-1.26%	**-0.006**	0.001	0.003**
Med + CB	Low	0.513	0.517	-1.21%	**-0.006**	0.002	0.041*
High	Low	0.512	0.517	-1.17%	**-0.006**	0.002	0.046*
Med	Low	0.513	0.517	-0.63%	**-0.003**	0.002	0.290
Med + CB	Med	0.513	0.513	-0.58%	**-0.003**	0.002	0.330
High	Med	0.512	0.513	-0.54%	**-0.003**	0.002	0.370
High	Med + CB	0.512	0.513	0.04%	**0.000**	0.002	0.950
Med + CB	Low + CB	0.513	0.510	0.74%	**0.004**	0.002	0.220
High	Low + CB	0.512	0.510	0.78%	**0.004**	0.002	0.194
Med	Low + CB	0.512	0.510	1.29%	**0.007**	0.002	0.030*

[6] The impact of the consistency bonus when offered without an accompanying incentive remains unknown, as we did not test a standalone consistency bonus trial arm due to concerns about potential customer confusion or dissatisfaction.

4.3.3 Turn-up events

4.3.3.1 Difference between treatment groups and control

All treatment groups recorded significantly higher electricity consumption during turn-up events compared to the control group (0.4184 kW) (Table 9; Figure 5; Figure 6). The observed increases ranged from 0.4512 kW in the Low group to 0.4734 kW in the High group. These differences were statistically significant across all arms ($p < 0.001$) (Table 9).

We also compared treatment groups to each other, as discussed in the next section (where we show pairwise comparisons and statistical significance testing in the next section). The High group had a coefficient of 0.05368 kW (12.83% increase on control group), while the Medium + Consistency Bonus and Low + Consistency Bonus groups had coefficients of 0.05294 (12.65% increase on control group) and 0.04868 (11.63% increase on control group), respectively. These were larger than the increases observed in the Low and Medium groups (0.03408 and 0.04218 kW), indicating that either increasing the base incentive or adding a consistency bonus was associated with greater uptake during turn-up events.

The ascending order of consumption and regression coefficients suggests that higher incentives are associated with larger increases in event-period consumption. All treatment arms were effective at increasing usage relative to the control.

Figure 5: Bar chart of treatment effects on hourly demand (kW) during turn-up events

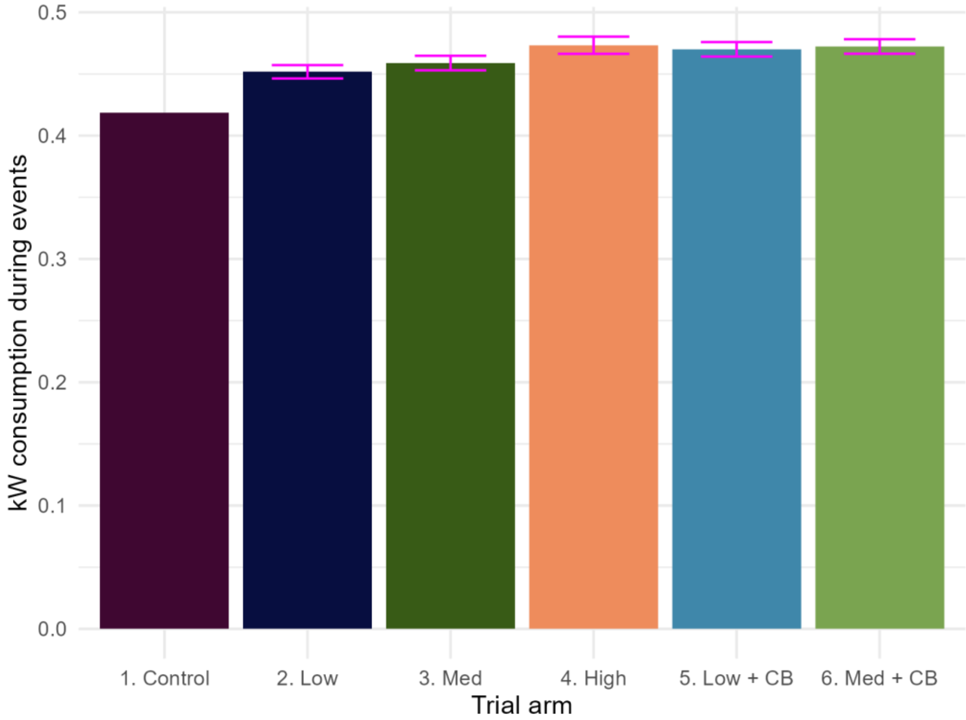
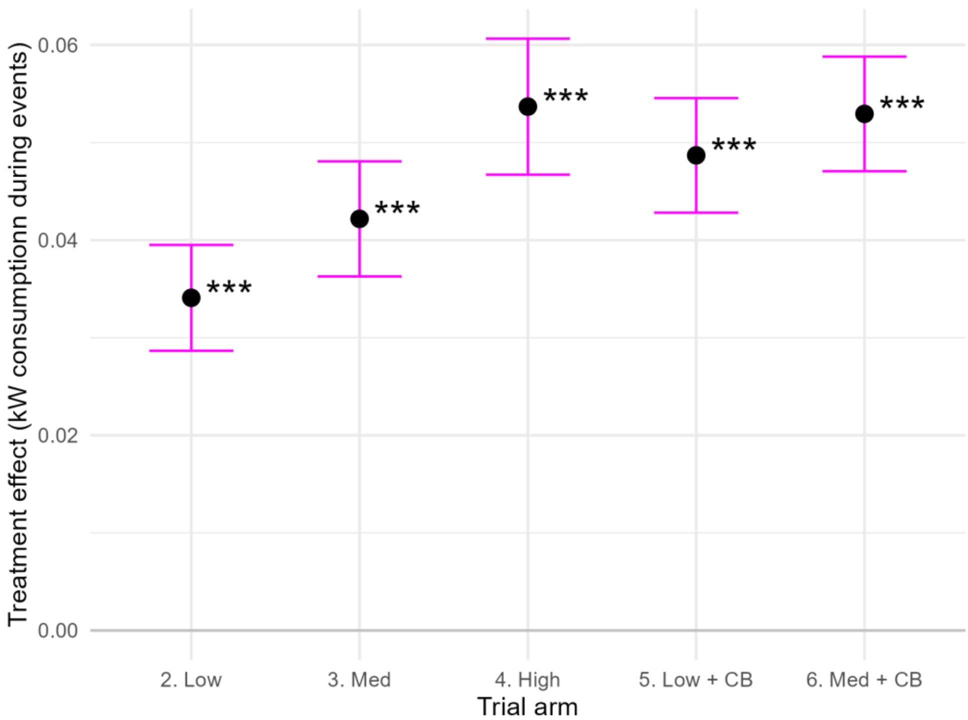


Figure 6: Coefficient plot of treatment effects on hourly demand (kW) during turn-up events



*Note: 95% confidence intervals are indicated in pink. Significance levels: *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.10.*

Table 9: Treatment effects on hourly demand (kW) during turn-up events

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value	% difference (regression coefficient vs. control)
Control	0.418				
Low	0.451	0.034	0.003	<0.001***	8.15%
Medium	0.459	0.042	0.003	<0.001***	10.08%
High	0.473	0.054	0.003	<0.001***	12.83%
Low + Consistency Bonus	0.470	0.049	0.003	<0.001***	11.63%
Medium + Consistency Bonus	0.472	0.053	0.003	<0.001***	12.65%

4.3.3.2 Treatment versus treatment analysis

The pairwise comparisons during turn-up events suggest that the High Incentive and Consistency Bonus treatments had the greatest turn-up. The difference in response between these groups was not statistically significant. At lower incentives levels, there is a clear trend of diminishing response, with Medium and Low incentive groups showing the lowest and second-lowest turn-up respectively.

Table 10: Pairwise treatment comparisons for turn-up events

Group 1	Group 2	Avg. kW per HH (G1)	Avg. kW per HH (G2)	% difference (vs. G2)	Regression Coefficient (vs. G2)	Standard Error	P-value
High	Low	0.473	0.451	4.17%	**0.020**	0.002	<0.001***
Med + CB	Low	0.472	0.451	3.99%	**0.019**	0.002	<0.001***
Low + CB	Low	0.470	0.451	3.11%	**0.015**	0.002	<0.001***
Low + CB, Med + CB	Low, Med	0.476	0.459	2.65%	**0.013**	0.001	<0.001***
High	Med	0.473	0.459	2.40%	**0.011**	0.002	0.004**
Med + CB	Med	0.472	0.459	2.26%	**0.011**	0.002	0.002**

Group 1	Group 2	Avg. kW per HH (G1)	Avg. kW per HH (G2)	% difference (vs. G2)	Regression Coefficient (vs. G2)	Standard Error	P-value
Med	Low	0.459	0.451	1.77%	0.008	0.002	0.013*
High	Low + CB	0.473	0.470	1.07%	0.005	0.002	0.196
Med + CB	Low + CB	0.472	0.470	0.91%	0.004	0.002	0.210
High	Med + CB	0.473	0.472	0.15%	0.001	0.002	0.852
Med	Low + CB	0.459	0.470	-1.39%	-0.006	0.002	0.065+

4.4 Secondary analysis results

The secondary analysis examined whether interventions had broader effects on electricity consumption outside of designated event periods. This analysis focused exclusively on treatment vs. control comparisons to determine if changes in behaviour extended beyond the targeted event windows, potentially indicating spillover effects or longer-term behavioural adjustments.

4.4.1 Summary of secondary analysis

- **Turn-down events:**
  - All treatment groups had baseline consumption levels comparable to the control group outside of turn-down events.
  - An event study of EFA 5 turn-down events (pooled treatments) found no evidence of demand displacement. However, there was some indication of a post-event spillover effect, particularly following 0.5-hour events, suggesting demand reduction extended slightly beyond the event window.
  - These results suggest demand destruction rather than displacement, but we caution that displacement may be too diffuse to detect, especially given that energy consumption from half-hour to half-hour and between participants is noisy (i.e., has high variance), and that events are typically 1 hour meaning that non-event half-hours on which calculations are performed may be up to 23 hours. For example, the Medium + Consistency Bonus group showed a material increase (0.01 kW) in non-event hours of the day – potentially *larger* than the treatment effect it was displacing),

which we expect to be a result of noisy data. Overall, we lack clear evidence of either demand displacement or demand destruction.

- **Turn-up events:**
    - All treatment groups had similar electricity consumption to the control group outside of turn-up events (0.43 kW), indicating demand creation rather than displacement.
    - If anything, our event study analysis again suggests spillover demand creation into nearby non-event half-hours.
    - However, again, we urge the same caution that displacement may be too diffuse to detect.

4.4.2 Turn-down events

4.4.2.1 Difference between treatment groups and control

The control group had an average electricity consumption of 0.42 kW outside of turn-down events. All treatment groups exhibited slightly higher consumption than the control (Table 11; Figure 7; Figure 8).

The Medium + Consistency Bonus group exhibited a statistically significant increase in average hourly electricity consumption outside of turn-down events compared to the control (0.43 kW vs. 0.42 kW, $p = 0.049^*$). While seemingly small on an hourly basis, the cumulative impact of such a difference over the substantial number of hours outside of events could be considerable, potentially even exceeding the magnitude of the observed in-event treatment effects. Furthermore, although the remaining treatment groups did not show statistically significant differences from the control in their average hourly consumption (ranging from 0.42 to 0.43 kW, $p > 0.05$), the observed trends warrant caution.

Therefore, while our findings currently indicate that turn-down events led to demand destruction rather than measurable displacement, this conclusion should be interpreted with caution. The potential for even small, diffuse increases in consumption outside of events to accumulate into meaningful totals over many hours suggests that our current analysis may be underpowered to definitively rule out displacement. This is particularly relevant given the statistically significant result for the Medium + Consistency Bonus group and the inherent challenges in detecting subtle, long-term shifts in consumption.

Figure 7: Bar chart of treatment effects on hourly demand (kW) outside of turn-down events

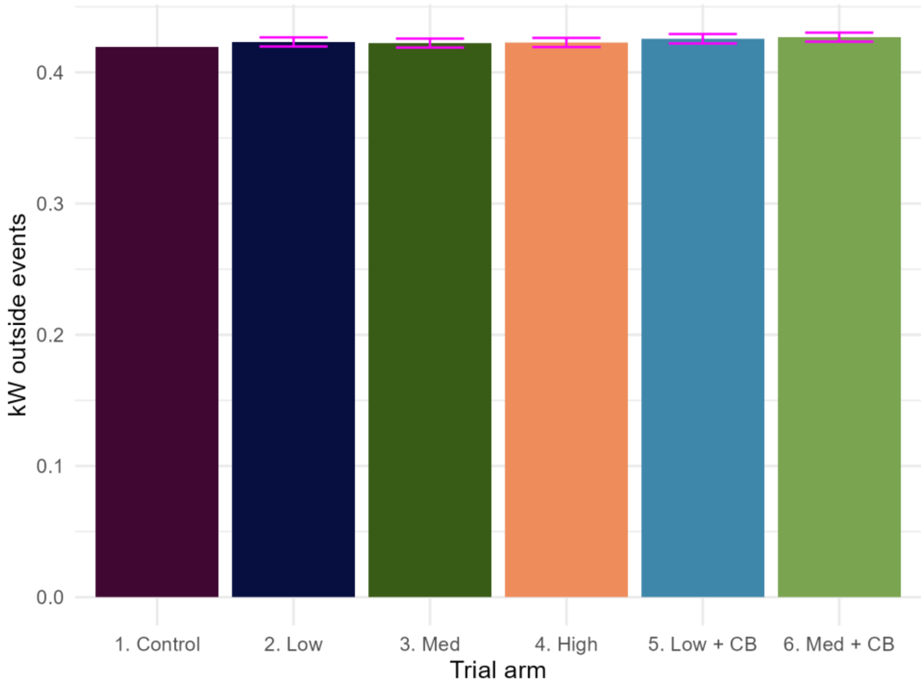
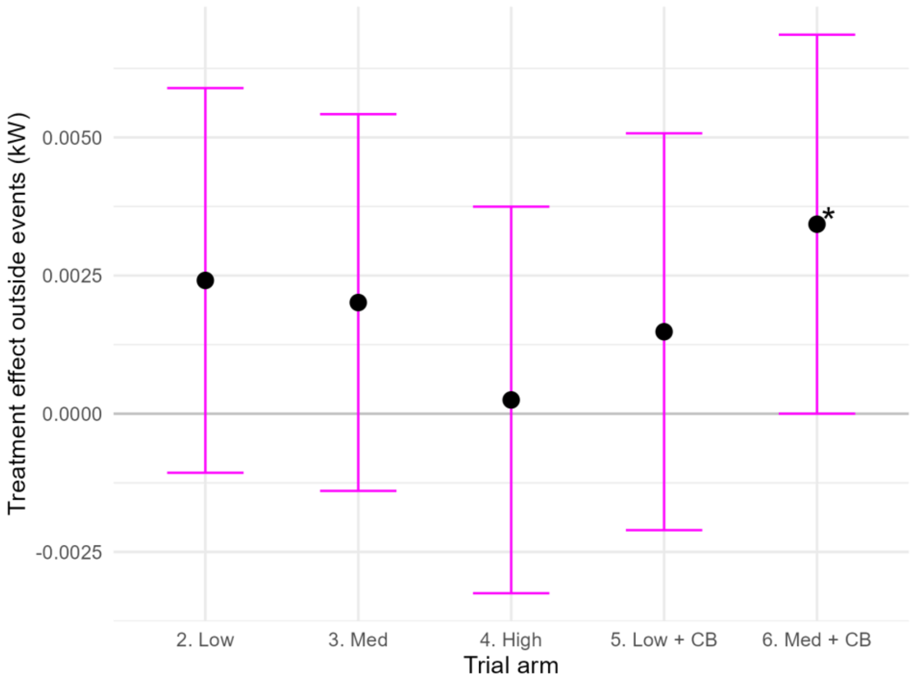


Figure 8: Coefficient plot of treatment effects on hourly demand (kW) outside turn-down events



*Note: 95% confidence intervals are indicated in pink. Significance levels: *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.10.*

Table 11: Treatment effects on hourly demand (kW) outside turn-down events

Trial Arm	Avg. kW outside turn-down events	Regression coefficient (vs. control)	Standard error	P-value	% difference (Regression vs. Control)
Control	0.42				
Low	0.42	0.00	0.00	0.17	0.58%
Medium	0.42	0.00	0.00	0.25	0.48%
High	0.42	0.00	0.00	0.89	0.06%
Low + Consistency Bonus	0.43	0.00	0.00	0.42	0.35%
Medium + Consistency Bonus	0.43	0.00	0.00	0.049*	0.82%

4.4.2.2 Event study (turn-down)

To take a finer-grained look at potential displacement and/or time-based spillover, we also present an event study of turn-down events occurring during EFA 5 (between 15:00 - 19:00; see Table 3), pooling all treatment groups (Figure 9; Figure 10). The analysis is shown in two forms: first, across all event durations combined; and second, separated by duration – specifically events lasting 1, 2, or 4 half-hour settlement periods (i.e., 0.5, 1, or 2 hours). Notably, even when pooling treatments, we find no clear evidence of demand displacement. If anything, there appears to be a slight spillover effect, with modest additional demand reduction in the period following the event, especially for the half-hour events. This pattern of destruction and spillover, rather than displacement, echoes the findings of Fowlie et al. (2021)[7] find in their seminal evaluation of critical peak pricing events, as well as the analysis by Jacob et al. (2024)[8] of Octopus Energy's implementation of the 2022-23 Demand Flexibility Service. However, as with the turn-down results, we recommend caution in assuming the absence of evidence of

[7] Fowlie, M., Wolfram, C., Baylis, P., Spurlock, C. A., Todd-Blick, A., & Cappers, P. (2021). Default effects and follow-on behaviour: Evidence from an electricity pricing program. *The Review of Economic Studies*, *88*(6), 2886-2934.

[8] 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.

displacement proves lack of displacement – the displacement may be too diffuse to detect with our sample sizes.

Figure 9: Treatment effects by half-hour before, during, and after EFA 5 turn-down events

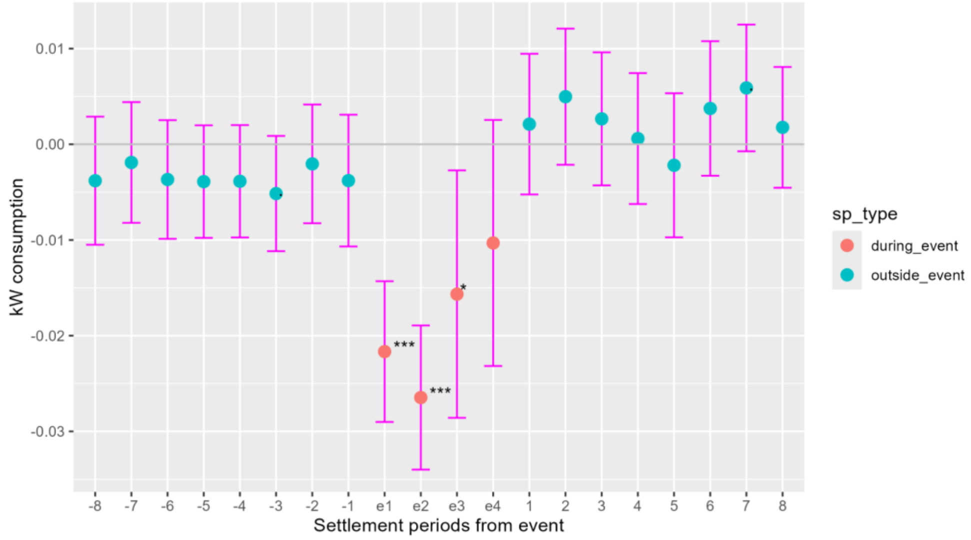
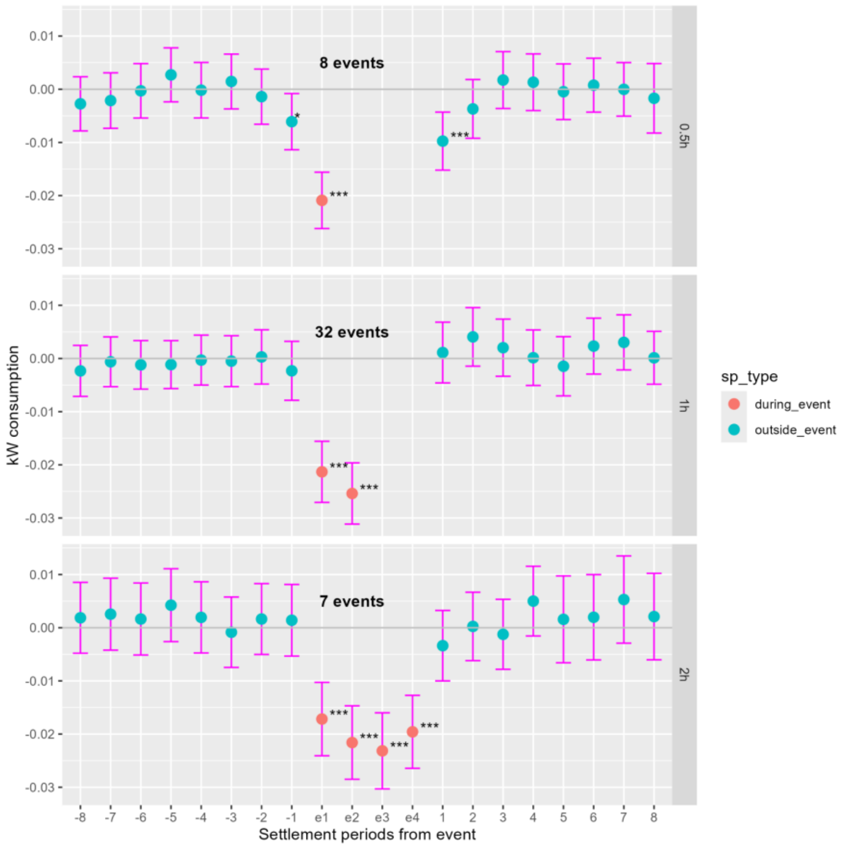


Figure 10: Treatment effects by half-hour before, during, and after EFA 5 turn-down events, by event duration



4.4.3 Turn-up events

4.4.3.1 Difference between treatment groups and control

The control group had an average electricity consumption of 0.43 kW outside of turn-up events (Table 12; Figure 11; Figure 12). All treatment groups exhibited nearly identical electricity consumption outside of turn-up events, averaging 0.43 kW – the same as the control group. The regression results show no statistically significant differences (all p-values > 0.05), and the effect sizes are effectively zero across the treatment. Given this, there is no meaningful variation between treatment groups or relative to the control, and no particular group stands out. The treatments can therefore be considered equivalent in their baseline consumption prior to and after events, though as noted previously the reader should be cautious that displacement effects may be too diffuse to easily detect.

Figure 11: Bar chart of treatment effects on hourly demand (kW) outside of turn-up events

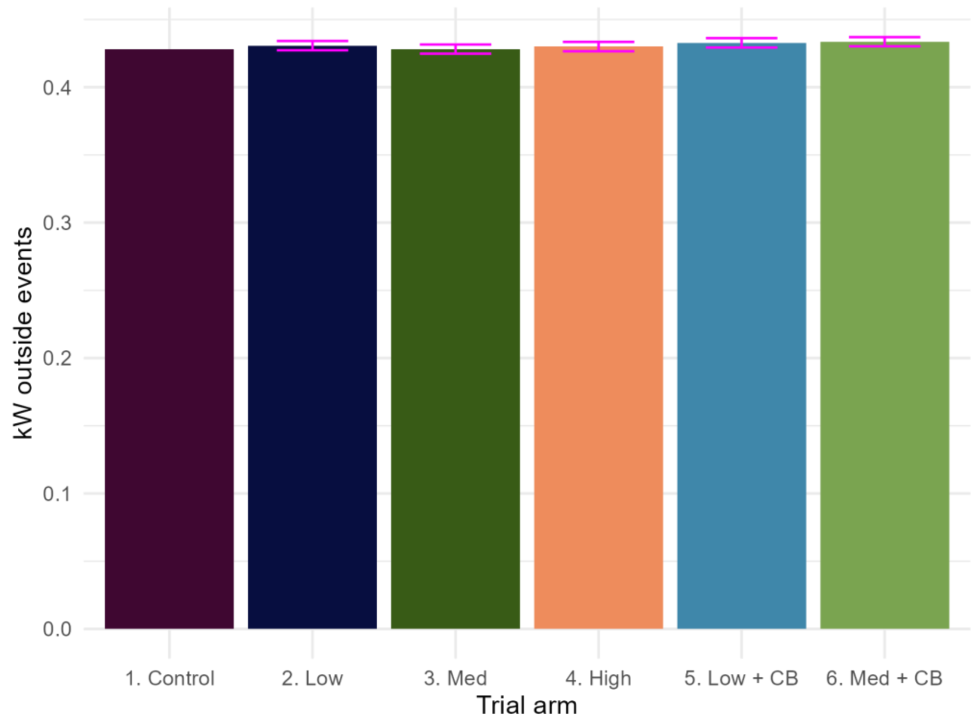


Figure 12: Coefficient plot of treatment effects on hourly demand (kW) outside of turn-up events

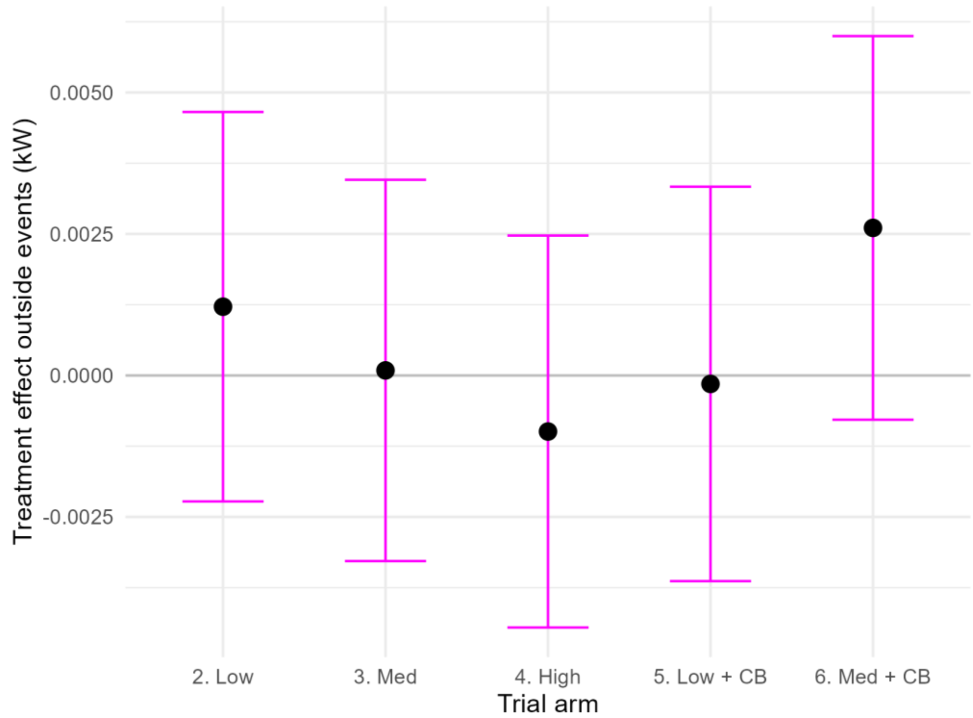


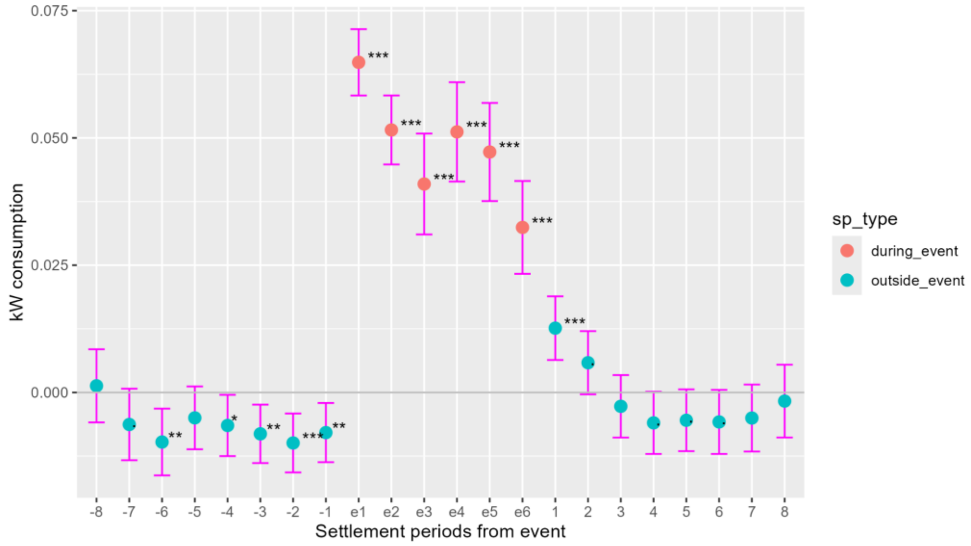
Table 12: Treatment effects on hourly demand (kW) outside of turn-up events

Trial Arm	Avg. kW outside turn-up events	Regression coefficient (vs. control)	Standard error	P-value	% difference (Coefficient vs. Control)
Control	0.43				
Low	0.43	0.00	0.00	0.49	0.28%
Medium	0.43	0.00	0.00	0.96	0.02%
High	0.43	0.00	0.00	0.57	-0.23%
Low + Consistency Bonus	0.43	0.00	0.00	0.93	-0.04%
Medium + Consistency Bonus	0.43	0.00	0.00	0.13	0.61%

4.4.3.2 Event study (turn-up)

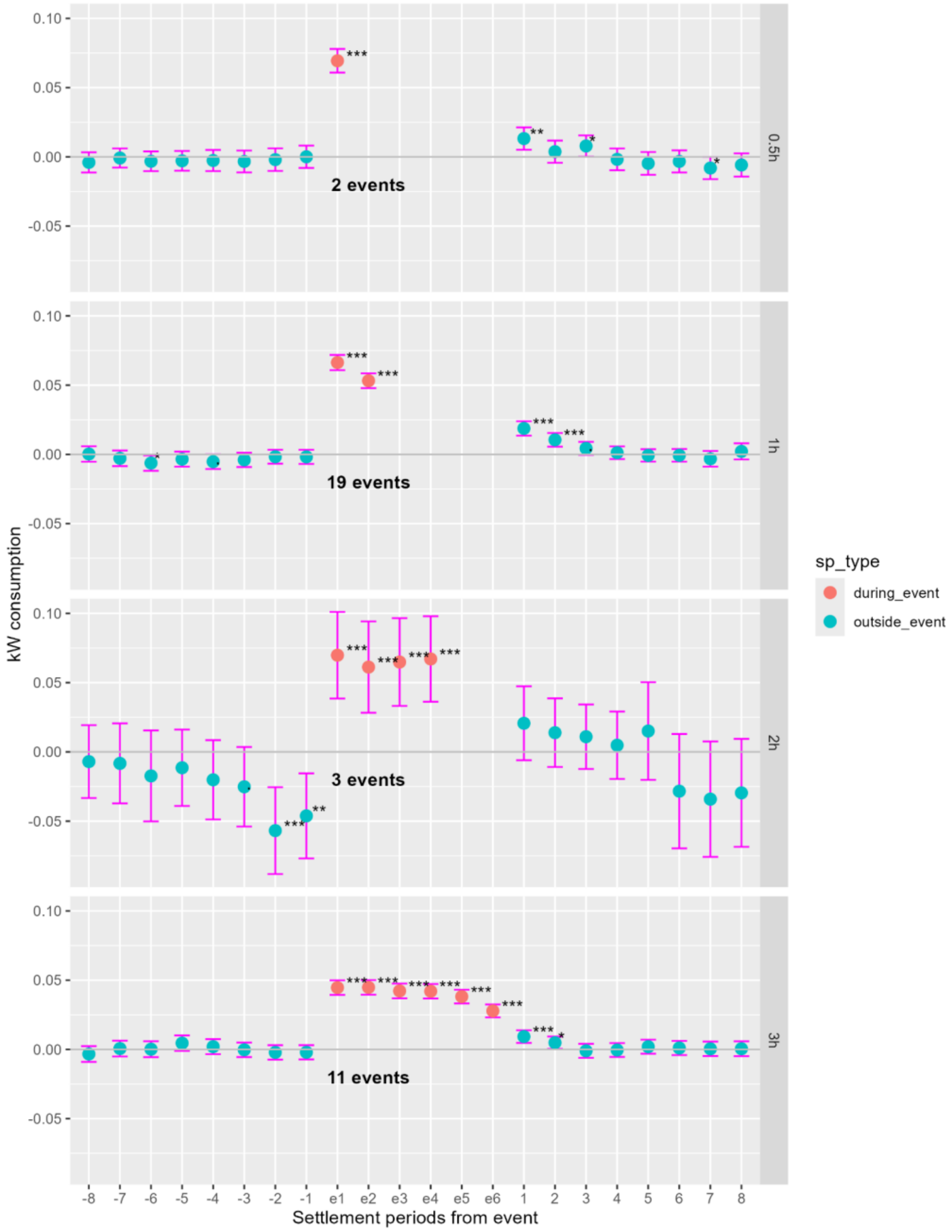
As with turn-down, we also present an event study of turn-up events during EFA 4 (between 11:00 - 15:00), again pooling all treatment groups (Figure 13; Figure 14). As before, we show results both across all event durations combined and separately by duration. Again, we observe limited evidence of displacement. There appears to be some demand reduction in the anticipation period before the event, though this seems to be driven by the three two-hour events only (see panel 3 of Figure 14). In addition, there appears to be some spillover demand *creation* into the subsequent periods. While this picture overall shows demand creation, we caution that there may be displacement that is too diffuse to detect or happens outside the window around the event that we have investigated.

Figure 13: Treatment effects each half-hour before, during, and after turn-up events during EFA 4 periods



*Note: 95% confidence intervals are indicated in pink. Significance levels: *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.10.*

Figure 14: Treatment effects each half-hour before, during, and after turn-up events during EFA 4, by event duration[9]



*Note: 95% confidence intervals are indicated in pink. Significance levels: *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.10.*

[9] Note that confidence intervals are wider for the three two-hour turn-up events due to a smaller sample size; these events were limited to Scottish customers participating in two-hour anti-symmetric trials. A few one-hour events follow a similar pattern, but the vast majority of one-hour turn-up events were conducted nationwide, resulting in larger sample sizes and narrower confidence intervals.

4.5 Exploratory analysis results

Unlike the primary and secondary analyses, exploratory analyses are not based on randomisation and should be interpreted as indicative rather than causal. Instead, they aim to uncover potential patterns and relationships that may inform future research.

Many of the analyses combine the five treatment arms into a single pooled treatment group to facilitate clearer comparisons with the control group. Unless otherwise specified, references to "treatment" refer to this pooled group.

We provide here a guide to interpret these regression outputs. To do so, we use Table 13 (treatment effects by notice period in turn-down events) below – reproduced here for convenience – as an example.

- The first row shows the raw average consumption in the control group under a particular category of event or customer characteristic, chosen essentially arbitrarily to be the reference category. "Short" notice events are the reference category. So, we see that the control group consumed 0.558 kW during short notice events.
- The second row's regression coefficient shows the difference in consumption between the control and treatment group during the reference category – here, treatment customers consumed 0.025 kW less during short-notice events than the control group. This difference was statistically significant (as indicated by the p-value being lower than 0.05).
- The third row's regression coefficient shows the difference in consumption between the reference and featured category among the control group – here, control customers consumed 0.06 kW more during long-notice events than short-notice ones. This difference was statistically significant, too.
- The final row's regression coefficient is called the *interaction*. It is like a "double difference" – it indicates whether treatment effects (treatment versus control) *differed* during long-notice events (versus short-notice events). Here, we see that treatment customers' turn-down was almost identical in long-notice events versus short-notice events: the 0.004 kW increase in consumption (i.e., reduction in turn-down) is small and not statistically significant. (In contrast, in Table 14: treatment effects by notice period in turn-*up* events, we see a statistically significant interaction between treatment and long-notice, with long-notice events having slightly greater turn-up than short-notice ones.)

Table 13 (duplicated from below): Treatment effects by notice period in turn-down events

Notification length	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Short (control group)	0.558				
Short + treated	0.537	-0.025	0.003	<0.001***	-4.52%
Long (vs. short)	0.516	0.060	0.002	<0.001***	10.75%
Long * treatment interaction	0.497	0.004	0.003	0.160	0.67%

4.5.1 Interactions with event and customer characteristics

A series of exploratory analyses investigated additional factors influencing electricity consumption that were not part of the core treatment design.

4.5.1.1 Effect of notification period

This exploratory analysis examines how the length of event notification, defined as short (≤ 4 hours) or long (> 4 hours), influenced electricity consumption during turn-down and turn-up events, and whether this effect varied by treatment status. Across the event schedule, the shortest notice given was 2 hours, and the longest was 21 hours.

In the summer 2024 utilisation payments trial, we found suggestive evidence that shorter-notice period events had similar response to longer-notice period events, but there were very few shorter-notice period events involved in the trial. This winter 2024-25 utilisation payments trial had many more events of each type of notice. However, the results are still only suggestive:

- During turn-down events, shorter-notice periods had greater turn-down than longer-notice periods, though the difference was not statistically significant ($p = 0.16$).
- During turn-up events, longer-notice periods had greater turn-up in electricity usage compared to shorter-notice periods ($p = 0.028$).

Note that, for the turn-down result, notice period was not randomised, and short-notice events may systematically differ from long-notice events in ways that affect consumption — for example, being more likely to occur during critical peaks. Further analysis is required to isolate the causal impact of notice period, ideally through randomised variation in future trials.

Turn-down events

In turn-down events, treated customers reduced their electricity consumption under both notification lengths. The interaction effect[10] for long-notice events was positive, indicating *less* turn-down in longer-notice period events, but it was not statistically significant, meaning we cannot distinguish a difference in treatment effect between long and short notice events (p = 0.16) (Table 13).

Table 13: Treatment effects by notice period in turn-down events

Notification length	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Short (control group)	0.558				
Short + treated	0.537	-0.025	0.003	<0.001***	-4.52%
Long (vs. short)	0.516	0.060	0.002	<0.001***	10.75%
Long * treatment interaction	0.497	0.004	0.003	0.160	0.67%

Turn-up events

In turn-up events, treated customers increased their electricity consumption under both notification lengths. The increase was 0.0419 kW for short-notice events ($p < 0.001$) and a *further* 0.0064 kW for long-notice events (p = 0.028) (Table 14).

[10] To interpret the interaction, it should be added to the main treatment effect to estimate the total effect under long-notice periods.

Table 14: Treatment effects by notice period in turn-up events

Notification length	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Short (control group)	0.450				
Short + treated	0.493	0.042	0.003	<0.001***	9.30%
Long (vs. short)	0.399	0.029	0.003	<0.001***	6.47%
Long * treatment interaction	0.449	0.006	0.003	0.028*	1.42%

4.5.1.2 Effect of event duration

This analysis investigates how the duration of flexibility events influenced electricity consumption, and whether the treatment effect varied by duration. Events were grouped into short (≤ 60 minutes) and long (> 60 minutes) durations.

Short-duration events led to no change in response during turn-down events but produced a more pronounced response per hour in consumption during turn-up events. Although the effect was small, long events produced a detectable reduction in response; in other words, the effect per hour was slightly lower than in shorter-duration turn-up events.

Turn-down events

During short-duration turn-down events, customers in the treatment group consumed 0.0233 kW less than the control group, a statistically significant effect ($p < 0.001$) (Table 15). No statistically significant difference was observed for long-duration turn-down events (p = 0.65) when compared with the short duration events, suggesting little difference in response by event duration.

Table 15: Treatment effects by event duration in turn-down events

Duration length	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Short (control	0.532				

Duration length	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
group)					
Short + treated	0.511	-0.023	0.003	<0.001***	-4.39%
Long (vs. short)	0.535	0.010	0.002	<0.001***	1.96%
Long * treatment interaction	0.517	0.001	0.003	0.650	0.23%

Turn-up events

Treated customers increased consumption during both short and long duration events. The estimated increase was 0.0587 kW for short events ($p < 0.001$), with this effect dampened by 0.0189 kW for long events ($p < 0.001$) (see Table 16). In other words, although longer-duration events provided more turn-up in kWh terms, the increase in consumption *per hour* (the key outcome in our analyses) was larger during shorter events.

Table 16: Treatment effects by event duration in turn-up events

Duration length	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Short (control group)	0.439				
Short + treated	0.499	0.059	0.003	<0.001***	13.36%
Long (vs. short)	0.408	-0.012	0.002	<0.001***	-2.81%
Long * treatment interaction	0.449	-0.019	0.002	<0.001***	-4.31%

4.5.1.3 Effect of day type

This analysis investigates whether the effectiveness of the treatment varied depending on whether events occurred on a weekday or a weekend. Treatment effects were nearly identical between weekdays versus weekend days (p=0.65), indicating no significant difference to response to the day type of the event.

Turn-down events

During weekdays, treated customers reduced electricity consumption by 0.0233 kW relative to the control group ($p < 0.001$) (Table 17). No statistically significant difference was observed for weekend events ($p = 0.65$), with a near-zero estimated coefficient (difference in treatment effect by weekday versus weekend day) of 0.0012 kW.

Table 17: Treatment effects by day type on turn-down events

Day type	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Weekday (control group)	0.544				
Weekend (vs weekday)	0.505	-0.023	0.003	<0.001***	-4.29%
Weekday + treated	0.523	0.010	0.002	<0.001***	1.92%
Weekend * treatment interaction	0.486	0.001	0.003	0.650	0.22%

Turn-up events

Electricity consumption was higher on weekends overall, but turn-up was again almost identical; the difference was not statistically significant (*p* = 0.78) (Table 18).

Table 18: Treatment effects by day type on turn-up events

Day type	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Weekday (control group)	0.401				
Weekend (vs weekday)	0.453	0.046	0.002	<0.001***	11.40%
Weekday + treated	0.448	0.004	0.002	0.110	0.90%
Weekend * treatment interaction	0.500	0.001	0.003	0.780	0.17%

4.5.1.4 Effect of being on OVO's Power Move

This section investigates whether customers enrolled in the OVO Power Move programme, who are incentivised to shift electricity away from peak hours (17:00–19:00) through prize draws, exhibited different responses during turn-down and turn-up events compared to non-members.

Customers enrolled in Power Move exhibited strong and statistically significant responses to both turn-down and turn-up signals. While differences in absolute consumption levels were modest, the treatment effects suggest that Power Move participants may be more responsive to demand flexibility signals than non-members. These results may be due to increased familiarity with demand flexibility, and/or customers who are particularly flexible in CrowdFlex may be more likely to select into a program like Power Move.

Turn-down events

Power Move members had a lower baseline consumption during events compared to non-members (0.512 kW vs. 0.538 kW) (Table 19). Despite this lower baseline, among those in the treatment group, non-members reduced consumption by 0.0185 kW ($p < 0.001$), while Power Move members reduced by a *further* 0.0198 kW ($p < 0.001$), over doubling the reduction in kW terms.

Table 19: Treatment effects by whether a customer is part of *Power Move* on turn-down events

Status	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not on Power Move (control group)	0.538				
On Power Move (vs not on Power Move)	0.512	-0.031	0.005	<0.001***	-5.83%
Not on Power Move + treated	0.522	-0.019	0.003	<0.001***	-3.45%
On Power Move * treatment interaction	0.460	-0.020	0.005	<0.001***	-3.68%

Turn-up events

Treated Power Move members increased their electricity consumption by 0.0568 kW ($p < 0.001$), almost tripling the 0.0357 kW turn-up observed among non-members ($p < 0.001$) (Table 20).

Table 20: Treatment effects by whether a customer is part of *Power Move* on turn-up events

Status	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not on Power Move (control group)	0.422				
On Power Move (vs not on Power Move)	0.402	0.013	0.004	<0.001***	3.16%
Not on Power Move + treated	0.462	0.036	0.002	<0.001***	8.44%
On Power Move * treatment interaction	0.483	0.057	0.005	<0.001***	13.46%

4.5.1.5 Effect of auto-enrolment (summer trial participants vs. new winter trial recruits)

This analysis compares responses between customers who were automatically enrolled from the summer trial (i.e. existing participants) and those who were newly recruited for the winter trial. The aim was to assess whether prior exposure influenced responsiveness to treatment.

Across both turn-down and turn-up events, newly recruited customers (not auto-enrolled) displayed almost identical response to treatment; differences were not statistically significant in either turn-down or turn-up events[11].

Turn-down events

Treated customers who were auto-enrolled showed no statistically significant difference in electricity use from not auto-enrolled customers (p = 0.85) (Table 21).

Table 21: Treatment effects by auto-enrolment on turn-down events

Status	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not auto enrolled (control group)	0.545				
Auto enrolled (vs not auto enrolled)	0.522	-0.003	0.004	0.520	-0.53%
Not auto enrolled + treated	0.523	-0.022	0.004	<0.001***	-4.12%
Auto enrolled * treatment interaction	0.506	0.001	0.005	0.850	0.16%

Turn-up events

[11] This null result offers reassuring evidence against a specific threat to internal validity arising from a randomisation decision discussed in Section 3: namely, the choice to exclude customers from the control group if they had previously been in the control group of the summer 2024 CrowdFlex utilisation payments trial. This approach meant that auto-enrolled customers were underrepresented in the control group, raising the possibility that differences in behaviour between auto-enrolled and non-auto-enrolled customers could bias results. The absence of treatment effect heterogeneity by auto-enrolment status suggests this is not a concern.

Similarly, among treated customers, new recruits increased their electricity consumption by the same amount as the auto-enrolled group (*p* = 0.83) (Table 22).

Table 22: Treatment effects by auto-enrolment on turn-up events

Status	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not auto enrolled (control group)	0.430				
Auto enrolled (control group)	0.409	-0.004	0.003	0.280	-0.82%
Not auto enrolled + treated	0.474	0.046	0.003	<0.001***	10.67%
Auto enrolled * treatment interaction	0.460	0.001	0.004	0.830	0.20%

4.5.1.6 Effect of estimated annual consumption (EAC)

This analysis examines whether customers with higher or lower annual electricity use, measured via their EAC, responded differently to the trial interventions. Customers were categorised into two groups: low EAC (< 2700 kWh) and high EAC (≥ 2700 kWh).

Customers with higher annual electricity consumption had greater change in terms of kW, though there is evidence that low-EAC customers' consumption change was greater relative to their own counterfactual consumption.

Turn-down events

Treated customers in the low EAC group reduced their consumption by 0.0186 kW, a statistically significant result ($p < 0.001$) (Table 23). The high EAC group reduced their consumption by a further 0.0060 kW, though note that this extra reduction is not statistically significant reduction (p = 0.22). As a percentage of their own consumption, though, the high-EAC increase is 3%.[12]

Table 23: Treatment effects by high/low estimated annual electricity consumption on turn-down events

EAC	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Low (below 2700 kW) (control group)	0.308				
High (≥ 2700 kW) (vs Low)	0.777	0.006	0.000	<0.001***	2.03%
Low + treated	0.289	-0.019	0.002	<0.001***	-6.04%
High * treatment interaction	0.759	-0.006	0.005	0.220	-1.96%

Turn-up events

For turn-up events, both EAC groups showed statistically significant increases in electricity use when treated, but the high-EAC group again showed greater increase in kW terms. The low EAC group increased usage by 0.0355 kW; the high EAC group by a *further* 0.0237 kW (*p* < 0.001 for both) (Table 24). As a percentage of their own consumption, again the low-EAC group's increase was greater, however.[13]

[12] Here we compare

- *High (the sum of the treatment effect in the low group + high * treatment interaction effect, divided by the avg. consumption in the high control group: (-0.019 + -0.006) / 0.777 = -3.22%*

  to:

- *Low (the treatment effect in the low group, divided by the avg. consumption in the low control group): -0.019 / 0.308 = 6.17%.*

[13] Here we compare

Table 24: Treatment effects by high/low estimated annual electricity consumption on turn-up events

EAC	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Low (below 2700 kW) (control group)	0.223				
High (≥ 2700 kW) (vs Low)	0.630	-0.036	0.006	<0.001***	-16.30%
Low + treated	0.256	0.036	0.002	<0.001***	15.93%
High * treatment interaction	0.696	0.024	0.004	<0.001***	10.61%

4.5.1.7 Month-to-month treatment effects

This section examines the change in electricity consumption during events over time, comparing treated and control customers each month. The analysis is presented for all EFA blocks as well as specifically for the subset of turn-down events that occurred during EFA 5 (15:00-19:00) and were "regular" (rather than "critical") and for the subset of turn-up events that occurred during EFA 4 (11:00-15:00); this goal of this sub-setting was to improve comparability between months, though results should still be treated with caution as they are non-causal and the type and spread of events still varied from month to month.

The results indicate that customers engaged with both turn-down and turn-up events across the trial period, with treatment effects observed in every month. While the size of

- *High: (0.035536 + 0.023674) / 0.62979 = -9.40%*

to:

- *Low: 0.035536 / 0.223124 = 15.9%.*

the response varied, the direction of effect was consistent: consumption decreased during turn-downs and increased during turn-ups, suggesting no fatiguing effects during the trial.

Turn-down events

All EFA blocks

Treated customers consistently used less electricity than the control group in every month from October to March. All regression coefficients were negative and statistically significant ($p < 0.001$), providing strong evidence that the treatment reduced consumption throughout the trial. The largest estimated effects occurred in:

- October (-6.5%)
- February (-5.37%)
- November (-4.67%)

Smaller reductions were seen in December (-2.9%) and January (-3.2%), indicating some seasonal variation in responsiveness (Table 25).

Table 25: Month to month treatment effects in turn-down events

Month	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
October control	0.365				
October treatment	0.342	-0.024	0.003	<0.001***	-6.47%
November control	0.465				
November treatment	0.444	-0.022	0.003	<0.001***	-4.67%
December control	0.597				
December treatment	0.581	-0.017	0.004	<0.001***	-2.86%
January	0.583				

Month	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
control					
January treatment	0.567	-0.019	0.004	<0.001***	-3.18%
February control	0.553				
February treatment	0.527	-0.030	0.003	<0.001***	-5.37%
March control	0.531				
March treatment	0.513	-0.022	0.004	<0.001***	-4.07%

EFA 5 only

Restricting the analysis to EFA Block 5 events, and avoiding "critical" down events, yields a broadly consistent pattern. Note two months are missing due to the absence of EFA 5 events (November and February).

- In all months with scheduled EFA 5 events (October, December, January, and March), treatment groups exhibited lower consumption than the control group (Table 26).
- All reported reductions are statistically significant, with p-values < 0.01, indicating high confidence in the observed effects.
- The largest relative reduction is observed in October (-8.22%).
- Reductions in other months are smaller but still consistent, ranging from -2.50% to -3.86%.

Table 26: Month to month treatment effects - EFA 5 only, and excluding critical pricing events

Month EFA 5 only	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. Control)
October control	0.321				

Month EFA 5 only	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. Control)
October treatment	0.294	-0.026	0.005	<0.001**	-8.22%
November control	**No EFA 5 events were scheduled.**				
November treatment					
December control	0.649				
December treatment	0.635	-0.017	0.005	<0.001**	-2.63%
January control	0.672				
January treatment	0.658	-0.017	0.005	0.002**	-2.50%
February control	**No EFA 5 events were scheduled.**				
February treatment					
March control	0.548				
March treatment	0.531	-0.021	0.005	<0.001***	-3.86%

Turn-up events

All EFA blocks

Treatment effects were positive and statistically significant across all months ($p < 0.001$), indicating that treated customers consistently increased their electricity usage during turn-up events relative to controls. Regression coefficients ranged from 0.0364 in March to 0.0600 in January (Table 27), with the largest relative increases observed in:

- October: +13.2%
- February: +12.5%
- December: +12.5%
- November: +11.8%

- January: +11.9%

These findings suggest a stable and sustained pattern of responsiveness to turn-up incentives throughout the winter trial period.

Table 27: Month to month treatment effects

Month	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
October control	0.320				
October treatment	0.361	0.042	0.003	<0.001***	13.15%
November control	0.389				
November treatment	0.437	0.046	0.005	<0.001***	11.81%
December control	0.389				
December treatment	0.437	0.049	0.003	<0.001***	12.47%
January control	0.504				
January treatment	0.566	0.060	0.004	<0.001***	11.92%
February control	0.401				
February treatment	0.450	0.050	0.004	<0.001***	12.53%
March control	0.451				
March treatment	0.489	0.036	0.003	<0.001***	8.08%

EFA 4 only

Analysis restricted to EFA Block 4 supports the main findings. Treatment effects remained positive and statistically significant across all months, with some variation in magnitude. Notable observations include:

- January again showed the strongest effect, with a 17.7% increase compared to control (Table 28).
- Other high months include November (+13.6%) and December (+13.2%) March and October also showed positive but smaller increases (8.4% and 1.3%, respectively).

Table 28: Month to month treatment effects - EFA 4 only

Month EFA 4 only	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
October control	0.381				
October treatment	0.440	0.064	0.005	<0.001***	16.73%
November control	0.506				
November treatment	0.576	0.069	0.007	<0.001***	13.63%
December control	0.440				
December treatment	0.496	0.058	0.004	<0.001***	13.15%
January control	0.397				
January treatment	0.460	0.070	0.006	<0.001***	17.70%
February control	0.473				
February treatment	0.525	0.054	0.007	<0.001***	11.40%
March control	0.320				

Month EFA 4 only	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
March treatment	0.347	0.027	0.004	<0.001***	8.36%

4.5.1.8 Effect of tariff type (Economy 7 versus not Economy 7)

This section examines whether customers' baseline electricity tariff type influenced their responsiveness to turn-down and turn-up events. Customers were categorised based on whether they were on Economy 7, the most common time-of-use (ToU) tariff, versus a non-ToU tariff.

Tariff type influenced the extent to which customers changed their electricity consumption during events. Similar to the Power Move subgroup analysis results, Economy 7 customers responded more than non-Economy-7 customers, though the magnitude of the enhanced response was less than it was for Power Move customers and was not statistically significant.

Turn-down events

Treated customers not on Economy 7 reduced their consumption during turn-down events by 0.0205 kW on average ($p < 0.001$), relative to their control counterparts. Treated customers on Economy 7 reduced by a further 0.0114 kW, but this difference was not statistically significant ($p = 0.33$) (Table 29).

Table 29: Effect of tariff type on turn-down events

Tariff type	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not on Economy 7 (control group)	0.520				
On Economy 7 (vs Not on Economy 7)	0.622	-0.101	0.011	<0.001***	-19.39%

Tariff type	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not on Economy 7 + treated	0.499	-0.021	0.002	<0.001***	-3.95%
On Economy 7 * treated interaction	0.606	-0.011	0.012	0.330	-2.20%

Turn-up events

Treatment effects during turn-up events were statistically significant across both tariff groups. Among customers not on Economy 7, electricity consumption increased by 0.0436 kW ($p < 0.001$), while Economy 7 customers increased their consumption by a further 0.0224 kW ($p = 0.016$) (Table 30).

Table 30: Effect of tariff type on turn-up events

Tariff type	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
Not on Economy 7 (control group)	0.388				
On Economy 7 (vs Not on Economy 7)	0.636	0.071	0.008	<0.001***	18.39%
Not on Economy 7 + treated	0.429	0.044	0.002	<0.001***	11.23%
On Economy 7 * treated	0.721	0.022	0.009	0.016*	5.78%

Tariff type	Avg. kW demand during event	Regression coefficient	Standard error	P-value	% difference (Regression vs. 1st row average)
interaction					

4.5.1.9 Effect of critical peak events

All treatment groups were also exposed to eight critical peak events, where customers could earn £1.75 per kWh reduced. This elevated incentive level was designed to test price elasticity under extreme conditions and to provide insight into the upper limit of consumer responsiveness.

These events were delivered between January and March, exclusively during EFA Block 5 (15:00–19:00), to test events during the period of the afternoon/evening peak load.

Treated customers reduced consumption by -0.018728 kW (3.09%) during non-critical turn-down events during EFA 5; they reduced consumption by a *further -0.012698 (-2.09%) during these "critical" events.* All of these effects were statistically significant ($p<0.001$) (Table 31).

Table 31: Effect of critical event pricing on hourly demand (kW) during EFA 5 events

Critical Peak	Avg. kW demand	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
Control - not critical (but yes EFA 5)	0.607				
Control - critical	0.577	-0.046	0.003	<0.001***	-7.65%
Treated - not critical (but yes EFA 5)	0.591	-0.019	0.004	<0.001***	-3.09%
Critical * treatment	0.550	-0.013	0.003	<0.001***	-2.09%

Critical Peak	Avg. kW demand	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
interaction					

We also examined how consumption during critical peak events varied with treatment – despite the fact that customers were all paid the same £/kWh during these events.

Across all treatment groups, electricity consumption was significantly lower than in the control group (0.5772 kW) (Table 32). Regression results show consistent and statistically significant reductions across all arms ($p < 0.001$).

We also see weak (not statistically significant, but perhaps suggestive) evidence that higher incentives may continue to elicit greater demand reduction even during the critical peak events where all groups received the same unit payment, a sort of "spillover" effect.

- The largest reductions were observed in the Low + Consistency Bonus (-0.0346 kW) and medium + Consistency Bonus (-0.0339 kW) groups.
- These were closely followed by the High (-0.0320 kW) and Medium (-0.0299 kW) groups.
- The Low arm also showed a meaningful reduction (-0.0251 kW), albeit slightly smaller.

Table 32: Treatment effects on hourly demand (kW) during critical events

Trial arm	Avg. kW demand during critical peak events	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
Control	0.577				
Low	0.555	-0.025	0.005	<0.001***	-4.35%
Medium	0.550	-0.030	0.005	<0.001***	-5.17%
High	0.551	-0.032	0.005	<0.001***	-5.55%
Low + Consistency	0.549	-0.035	0.005	<0.001***	-5.99%

Trial arm	Avg. kW demand during critical peak events	Regression coefficient	Standard error	P-value	% difference (Coefficient vs. Control)
Bonus					
Medium + Consistency Bonus	0.548	-0.034	0.005	<0.001***	-5.88%

4.5.2 Interactions with EFA

This analysis examines how treatment effects varied across the six standard EFA blocks, each representing a four-hour settlement window.

Turn-down events

Treatment effects during turn-down events varied across EFA blocks. The strongest reductions were observed in the three blocks that were the primary focus for turn-down targeting:

- **EFA 3 (07:00–11:00):** Treatment reduced consumption by −0.0205 kW (SE = 0.0039, $p < 0.001$).
- **EFA 4 (11:00–15:00)**: The effect was larger at −0.0289 kW (SE = 0.0032, $p < 0.001$).
- **EFA 5 (15:00–19:00):** Events in this block showed a treatment effect of −0.0231 kW (SE = 0.0035, $p < 0.001$). This block also included all critical peak events, which may have contributed to the strong and statistically significant change in consumption.

A smaller but still significant reduction was observed in EFA 6 (19:00–23:00) (−0.0126 kW, $p < 0.001$) (Table 33), though this block received fewer turn-down events. No effects were detected in overnight blocks (EFA 1 and 2), where no events were conducted.

These findings align with the targeting strategy, showing that customers responded to signals during the grid's peak loads and that peak-time reductions were achieved across all targeted daytime periods. While exploratory, this provides early evidence that demand-side flexibility can be delivered at a range of times, not just at times of peak demand.

Table 33: Treatment effects by EFA block on turn-down events

EFA Block	Trial Arm	Average kW	Regression Coefficient (vs. Control)	Standard Error	P-value	% difference (Coefficient vs. Control)
EFA 3: 07:00 – 11:00	Control	0.458				
	Pooled treatment groups	0.439	-0.021	0.004	<0.001***	-4.48%
EFA 4: 11:00 – 15:00	Control	0.464				
	Pooled treatment groups	0.436	-0.029	0.003	<0.001***	-6.23%
EFA 5: 15:00 – 19:00	Control	0.597				
	Pooled treatment groups	0.577	-0.023	0.004	<0.001***	-3.87%
EFA 6: 19:00 – 23:00	Control	0.514				
	Pooled treatment groups	0.504	-0.013	0.004	<0.001***	-2.45%

Turn-up events

Turn-up events were concentrated in EFA 4 (11:00–15:00) and EFA 6 (19:00–23:00), and these periods yielded some of the strongest treatment effects:

- EFA 4: Treated participants increased usage by 0.0537 kW (SE = 0.0030, $p < 0.001$).
- EFA 6: The effect was 0.0575 kW (SE = 0.0037, $p < 0.001$) (Table 34).

Notably, increases in electricity consumption were also observed across all other EFA blocks, despite these not being prioritised for turn-up. For example:

- EFA 3 (07:00–11:00): +0.0747 kW (SE = 0.0057)
- EFA 5 (15:00–19:00): +0.0444 kW (SE = 0.0048)

This suggests that customer response to turn-up events was not confined to prioritised periods, and that demand could be increased even outside of mid-day and evening windows. This is particularly relevant in a wind-dependent system like Great Britain's, where surplus generation can occur at all times of day due to its intermittency .

Table 34: Pooled treatment effects by EFA block on turn-up events

EFA Block	Trial Arm	Average kW	Regression Coefficient (vs. Control)	Standard Error	P-value	% difference (Coefficient vs. Control)
EFA 1: 23:00 – 03:00	Control	0.304				
	Pooled treatment groups	0.334	0.027	0.005	<0.001***	8.80%
EFA 2: 03:00 – 07:00	Control	0.293				
	Pooled treatment groups	0.326	0.032	0.004	<0.001***	10.81%
EFA 3: 07:00 – 11:00	Control	0.447				
	Pooled treatment groups	0.521	0.075	0.006	<0.001***	16.73%
EFA 4: 11:00 – 15:00	Control	0.413				
	Pooled treatment groups	0.465	0.054	0.003	<0.001***	13.00%
EFA 5: 15:00 – 19:00	Control	0.541				
	Pooled treatment groups	0.586	0.044	0.005	<0.001***	8.17%
EFA 6: 19:00 – 23:00	Control	0.514				
	Pooled treatment groups	0.572	0.058	0.004	<0.001***	11.18%

4.5.3 Interactions with customer characteristics obtained from onboarding survey analysis

Prior to enrolling in the trial, customers were invited to complete an onboarding survey to provide additional detail about household characteristics and energy-related behaviours. A total of 28,410 participants responded to the survey.

These survey results will be used later in the project to explore differences in demand response by demand archetype, and also to help in creating new demand response archetypes.

By way of exploration in this report, we have investigated differences in demand and demand response according to some of the key characteristics returned in the survey results: household composition, presence of children or older adults, and ownership of on-site generation. While not the primary focus of the trial, these findings provide early indications of how different customer types may engage with demand-side interventions.

4.5.3.1 Households with children (17 & below)

Households without children demonstrated similar responsiveness to treatment during turn-down events but lower increase during turn-up events.

Turn-down events

Households with children consumed more electricity per hour (0.654 kW) than those without (0.493 kW) during turn-down events, even before treatment. However, the reduced turn-down from households with children was not statistically significant (p=0.25) (Table 35).

Table 35: Treatment effects by presence of children in household on turn-down events

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
Without children (control	0.493				

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
group)					
With children (vs without children)	0.654	0.033	0.008	<0.001***	6.63%
Without children + treated	0.459	-0.029	0.004	<0.001***	-5.79%
With children * treatment interaction	0.637	0.009	0.008	0.250	1.90%

Turn-up events

A different pattern emerged during turn-up events (Table 36):

- Among households without children, treatment was associated with a significant increase in consumption (+0.0652 kW, $p < 0.001$).
- Among households with children, the increase was -0.028706 kW smaller ($p < 0.001$), suggesting either reduced responsiveness or potential difficulty in adapting demand upward during these windows.

Table 36: Treatment effects by presence of children in household on turn-up events

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
Without children (control group)	0.393				
With children (vs without children)	0.505	-0.010	0.006	0.060+	-2.63%
Without children + treated	0.456	0.065	0.003	<0.001***	16.57%

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
With children * treatment interaction	0.548	-0.029	0.007	<0.001***	-7.30%

4.5.3.2 Households with adults 65 & above

This analysis examined whether households with adults aged 65 or above responded differently to flexibility events compared to those without older adults.

Households with and without adults aged 65 or above responded similarly to turn-down events, with no statistically significant difference in treatment effects observed. However, during turn-up events, households with older adults demonstrated significantly greater increases in electricity consumption, suggesting that age composition may play a role in enhancing responsiveness to turn-up interventions.

Turn-down events

There was no statistically significant difference in treatment effects between households with and without adults aged 65 or above (p = 0.39) (Table 37). While treated households with adults aged 65 or above showed a slightly greater reduction in consumption (-0.006 kW) compared to those without, this difference was not significant, suggesting that the presence of older adults did not materially alter responsiveness to the intervention.

Table 37: Treatment effects by presence of 65+ adults in household on turn-down events

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
Without adults ≥65	0.535				
With adults ≥65 (vs without)	0.501	0.004	0.006	0.570	0.68%
Without adults ≥65 + treated	0.499	-0.025	0.005	<0.001***	-4.67%
With adults ≥65 * treatment interaction	0.476	-0.006	0.007	0.390	-1.12%

Turn-up events

Households with adults aged 65 or over were able to increase their electricity consumption during turn-up events, relative to households without older adults. Specifically, households without adults aged 65 or above increased their consumption by 0.0465 kW when treated, while households with older adults increased consumption by an additional 0.0272 kW, leading to a total treatment effect of approximately 0.0737 kW (Table 38). This difference is statistically significant ($p < 0.001$) and indicates that the presence of older adults may be associated with greater responsiveness to turn-up interventions.

Table 38: Treatment effects by presence of 65+ adults in household on turn-up events

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
Without adults ≥65	0.431				
With adults ≥65 (vs without)	0.387	-0.001	0.005	0.790	-0.28%
Without adults ≥65 + treated	0.470	0.047	0.004	<0.001***	10.81%
With adults ≥65 * treatment interaction	0.469	0.027	0.006	<0.001***	6.32%

4.5.3.3 Presence of solar, battery, micro-hydro, or other generation/storage

Of the 30,975 respondents, 7.8% reported having some form of on-site generation/storage (e.g. solar panels, battery storage, or micro-hydro). This subgroup was compared to the wider population to understand how embedded generation may influence baseline consumption and treatment responsiveness.

Turn-down events

Households with and without on-site generation/storage both exhibited reductions in electricity consumption during turn-down events when treated. The group with on-site generation/storage reduced consumption by an additional −0.0031 kW relative to the group without, but this difference was not statistically significant (p = 0.84), indicating that the presence of on-site generation/storage did not materially affect turn-down responsiveness (Table 39).

Table 39: Treatment effects by presence of on-site generation on turn-down events

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
Without presence	0.528				
With presence (vs without)	0.575	0.054	0.014	<0.001***	10.22%
Without presence + treated	0.492	-0.026	0.003	<0.001***	-4.91%
With presence * treatment interaction	0.594	-0.003	0.015	0.840	-0.60%

Turn-up events

During turn-up events, households with on-site generation increased their consumption significantly more than those without. The treated group with on-site generation increased usage by an additional 0.08 kW, a difference that was statistically significant (p < 0.001) (Table 40). This finding suggests that on-site generation may enable greater responsiveness to turn-up signals.

Table 40: Treatment effects by presence of on-site generation on turn-up events

Characteristic	Avg. kW per hour	Coefficient	Standard Error	P-value	% difference (Regression vs. 1st row average)
Without presence	0.415				
With presence (vs without)	0.465	0.051	0.010	<0.001***	12.30%
Without presence + treated	0.461	0.052	0.003	<0.001***	12.60%
With presence * treatment interaction	0.652	0.082	0.015	<0.001***	19.84%

4.6 Translating treatment effects into flexibility value and price elasticity

4.6.1 Flexibility value

All preceding analyses estimated the impact of treatments on electricity consumption (half-hourly kWh, converted to kW for ease of interpretation). When the outcome is expressed in kWh, it is natural to consider the associated value and cost of the flexibility delivered. Although a detailed assessment of these aspects will be undertaken by ERM as part of separate deliverables within the CrowdFlex programme, we provide high-level estimates here as a preliminary indication.

Total flexibility (kWh) by treatment group: We believe that the correct approach is to take the coefficient[14] on each treatment group from Table 7 and Table 9, multiplied by

[14] A 95% confidence interval around total flexibility would come from a similar calculation; for simplicity, however, we have not done so here.

the number of customers in that group[15], multiplied by the number of hours of events[16]. This product is the total flexibility (in kWh) delivered by a group over the course of the trial, where that estimate is based on the difference between that group and the control group – thus, by not relying on the P376 or any other baseline, it is not sensitive to any potential baselining biases (we discuss these potential biases in more depth immediately below).

We show the £/kWh for the non-consistency-bonus groups separately for turn-down and turn-up in Table 41.

However, to obtain £/kWh for *all groups*, we combine the flexibility from turn-up and turn-down events into a single overall measure, as remuneration in the consistency bonuses group is not tied to a particular event type (Table 41).

Total cost: The cost of this flexibility is the expenditure by NESO, based on *settled* flexibility volumes. By this, we mean flexibility based on customer baselines; this results in higher £/kWh figures than the official remuneration rates, for reasons we discuss below. We show these totals in the relevant treatment group over the course of the trial period (Table 42).

These two figures – total cost (£) and total flexibility (kWh) – allow us to obtain flexibility (kWh) per cost (£) – i.e., a simple benefit cost ratio that may be of interest to system operators procuring flexibility volumes. That said, this calculation does not attempt to quantify the full benefits of the procured flexibility to the energy system, nor to understand full societal welfare impacts, such as at least *some* of the cost representing a benefit to customers in the form of a transfer.[17]

Using this method, we find that the lower-consuming groups in terms of kWh flexibility *delivered* were higher-performing in terms of being lower cost in £/kWh terms. This

---

[15] We believe this is a safe assumption given the balanced attrition rates between treatment groups.
[16] There were 51 hours of turn-down events and 59 hours of turn-up events. We classified anti-symmetric events as turn-down events due to the great majority of Great Britain (all of England and Wales) turning *down* during anti-symmetric events.
[17] Hahn et al. (2024) and Hendren and Sprung-Keyser (2020) discuss this in more depth.
Hahn, R. W., Hendren, N., Metcalfe, R. D., & Sprung-Keyser, B. (2024). A welfare analysis of policies impacting climate change (No. w32728). *National Bureau of Economic Research.*
Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3), 1209-1318.

finding suggests a potential trade-off: to secure greater volumes of flexibility, a DSRSP or system operator may need to offer a higher £/kWh payment to sufficiently incentivise customer response.

Table 41: Cost of turn-down and turn-up flexibility using our RCT estimates of flexibility

Trial arm: Down	Average kW shift	# of customers	Total kWh	Remuneration	£/kWh
Low	-0.017	8,066	-6,969	£33,981.89	-£4.88
Medium	-0.020	8,098	-8,310	£48,125.82	-£5.79
High	-0.023	8,146	-9,497	£86,943.81	-£9.15
Low + CB	-0.027	8,046	-11,038		
Med + CB	-0.023	8,120	-9,566		
Trial arm: Up	**Average kW shift**	**# of customers**	**Total kWh**	**Remuneration**	**£/kWh**
Low	0.034	8,066	16,218	£5,317.11	£0.33
Medium	0.042	8,098	20,153	£8,308.82	£0.41
High	0.054	8,146	25,799	£15,018.12	£0.58
Low + CB	0.049	8,046	23,109		
Med + CB	0.053	8,120	25,362		

Table 42: Cost of *combined* turn-down and turn-up flexibility using our RCT estimates of flexibility

Trial arm	Total kWh (up and down combined)	Remuneration	£/kWh
Low	23,187	£39,299	£1.69
Medium	28,462	£56,435	£1.98
High	35,296	£101,962	£2.89
Low + CB	34,147	£135,060	£3.96
Med + CB	34,929	£154,863	£4.43

This approach results in higher costs per unit of flexibility compared to using remuneration per unit of settled flexibility.

Table 43: Cost of combined turn-down and turn-up flexibility in £/kWh based on settled flexibility

Trial Arm	Total kWh (up and down combined)	Remuneration	£/kWh[18]
Low	112,970	£40,194	£0.36
Medium	116,333	£57,025	£0.49
High	123,827	£102,467	£0.83
Low + CB	119,786	£135,060	£1.13
Med + CB	122,911	£154,863	£1.26

Settled flexibility (Table 43) refers to the consumption change measured by OVO on a per-customer basis to determine customer payments. OVO then settles this amount with NESO, the ultimate procurer of flexibility in this project (and potentially under business-as-usual conditions as well, although the procurer could also be a DSO depending on the specific flexibility requirement).

There are two key reasons why results differ depending on whether flexibility is measured using treatment-control comparisons (as in our analysis) or settled flexibility.

1. First, customer baselines may be inaccurate, leading to either over- or underestimation of settled flexibility relative to the flexibility estimated through experimental comparisons.
2. Second, OVO customers are remunerated for shifting in the correct direction as per P376 but are not penalised for shifting in the wrong direction. For example: if a customer shifts –1 kWh in the wrong direction in one settlement period, but +1.2 kWh in the correct direction in another, they are credited with a net shift of +0.2 kWh. Conversely, if a customer shifts -1.2 kWh in the wrong direction and then +1 kWh in the correct direction, they are credited with 0 kWh; their contribution is "clipped" at zero rather than being allowed to fall into negative values.

Total settlement claimed as part of the trial is based on these "correct direction" consumption changes, the same metric under which customers are credited.

[18] These £/kWh costs include the £1.75/kWh paid during critical peak events.

Aggregating the two examples above, the total settled consumption change is +0.2 kWh, even though the "unclipped" net change across both customers would be 0 kWh. This structure inflates settled flexibility relative to actual system-level impact – especially for demand turn-down, where the signal-to-noise ratio is smaller.

In Table 44 and Table 45, we examine the consumption change by treatment group using three different approaches: (1) treatment versus control group comparisons, which we consider the "ground truth" measure of flexibility by group; (2) comparisons of actual consumption against baseline values, where consumption changes in the incorrect direction can offset those in the correct direction; and (3) baseline comparisons where consumption changes in the incorrect direction are clipped to zero. The analysis indicates that most of the inflation arises from the clipping mechanism, as the estimates from approaches (1) and (2) are relatively similar, while the flexibility estimated under approach (3) is significantly higher.

It is also worth noting that the settlement approach used in CrowdFlex is similar to the approach used in live markets (such as the Demand Flexibility Service and Local Constraint Market), though the inflation factor in those markets is likely reduced because of opt-in requirements. CrowdFlex is able to quantify this inflation with a high degree of rigour using the randomised control group as a counterfactual portfolio-level consumption.

With these points in mind, OVO has planned to introduce an opt-in stage in its summer trial. It is expected that the introduction of an opt-in mechanism would reduce settled flexibility inflation, as a greater proportion of correct-direction P376 shifts would reflect deliberate customer action.

Table 44: Comparing measurements of flexibility per customer per hour: turn-down

Trial arm	1. Regression coefficient (vs. control)	2. Avg. P376 kW shift during (2)	3. Avg. P376 kW shift correct direction (3)	Inflation factor (measure #3 / measure #2)
Control		-0.00064	-0.14588	-
Low	-0.01694	-0.014484	-0.15266	10.5
Medium	-0.02012	-0.014254	-0.15204	10.7

Trial arm	1. Regression coefficient (vs. control)	2. Avg. P376 kW shift during (2)	3. Avg. P376 kW shift correct direction (3)	Inflation factor (measure #3 / measure #2)
High	-0.02286	-0.02126	-0.15718	7.4
Low + Consistency Bonus	-0.0269	-0.02158	-0.15562	7.2
Medium + Consistency Bonus	-0.0231	-0.02304	-0.15972	6.9
Treatment (pooled)	-0.02198	-0.018926	-0.15544	8.2

Table 45: Comparing measurements of flexibility per customer per hour: turn-up

Trial arm	1. Regression coefficient (vs. control)	2. Avg. P376 kW shift during (2)	3. Avg. P376 kW shift correct direction (3)	Inflation factor (measure #3 / measure #2)
Control		-0.006544	0.09858	-
Low	0.03408	0.0256	0.12976	5.1
Medium	0.04218	0.03546	0.13716	3.9
High	0.05368	0.04826	0.1504	3.1
Low + Consistency Bonus	0.04868	0.04262	0.1439	3.4
Medium + Consistency Bonus	0.05294	0.04366	0.14647	3.4
Treatment (pooled)	0.04634	0.03912	0.14154	3.6

It is worth noting that the ranking of treatment effectiveness based on measure (3) broadly aligns with that derived from measures (1) and (2) (Table 45). This alignment suggests that the *relative* cost-effectiveness of interventions is likely preserved in cost-benefit analyses, even if the absolute estimates are conservative due to the methodology used for settling flexibility.

4.6.2 Price elasticity

Finally, we examine how the estimated treatment effects translate into implied price elasticities by treatment group. Due to the opacity of per-kWh pricing under the Consistency Bonus groups, we restrict this analysis to the low, medium, and high incentive groups. For the purposes of this calculation, turn-down incentives are treated as an opportunity cost for non-participating customers and are therefore added to their typical unit rates. Turn-up incentives are treated as price discounts. We assume a baseline unit rate of £0.27/kWh (Table 46), approximating the rate of OVO Energy's Simpler Energy tariff (noting that exact rates vary slightly across regions of Great Britain). Under these assumptions, the treatments translate as follows:

Table 46: Treatments expressed in % change from typical OVO unit rate

Trial arm	Turn-down			Turn-up		
	Incentive	New "unit rate"	Increase from £0.27 per kWh	Incentive per kWh	New "unit rate"	Reduction from £0.27 per kWh
Low	£0.20	£0.47	74%	£0.10	£0.17	-37%
Medium	£0.50	£0.77	104%	£0.15	£0.12	-56%
High	£1.25	£1.52	179%	£0.25	£0.02	-93%

We then calculate the point price elasticity of demand implied by the percent change in quantity demand we identified earlier in this section (Table 47).

Table 47: Treatment effects expressed in point price elasticities

Trial arm	Turn-down			Turn-up		
	Price change (%)	Quantity change (%)	Price elasticity	Price change (%)	Quantity change (%)	Price elasticity
Low	74%	-3.18%	-0.04	-37%	8.15%	-0.22
Medium	104%	-3.78%	-0.04	-56%	10.08%	-0.18
High	179%	-4.29%	-0.02	-93%	12.83%	-0.14

In summary, we observe lower price elasticities in turn-down than turn-up events. Note that these results should be treated with caution. We have estimated elasticities based on some generalising assumptions. Further research is needed to construct robust price elasticities for demand response.

5. Discussion

This large-scale RCT provides strong causal evidence that financial incentives can drive demand flexibility among residential consumers. Across 76 events, with more than 55,000 participants, the intervention encouraged consumers to both reduce and increase electricity consumption during targeted time windows.

Turn-up events saw a stronger response (+11.1%) than turn-down events (−4.1%), along with greater price sensitivity to incentive levels. It is possible that consumers find it easier or more appealing to increase discretionary loads at key times than to delay or avoid usage, though in CSE's survey from the previous utilisation trial in summer 2024 participants report similar levels of ease in providing turn-up and turn-down.[19] The higher turn-up response may also be due to structural asymmetry between the types of event, where turn-up is theoretically unbounded whereas a customer cannot reduce demand past 0 kWh.

Lower incentives delivered more flexibility per pound spent, while higher incentives delivered greater volumes, pointing to a trade-off between maximising response and minimising cost. Full value assessments should include the potential long-term system and welfare benefits of demand flexibility, particularly its role in decarbonising the grid in line with the Government's 2030 target.

The trial found little evidence of demand displacement, potentially indicating load destruction or creation during events. That said, it should be noted that there may be displacement that is too diffuse to detect or happens outside the window around the event that we have investigated. This could be an area for more in-depth research in future.

Exploratory analysis offered early insights to refine demand response programs. Higher response was seen in households with on-site generation (turn-up only), adults over 65 (turn-up only), and participants in OVO's Power Move initiative. *Per-hour response* was greater in shorter-duration events, particularly in turn-up events. Notice period saw no clear trends and is an area for future investigation. These early insights suggest schemes could benefit from tailored design based on customer characteristics, technologies, and event configuration.

---

[19]*CrowdFlex Summer Trial Customer Survey Report*, Centre for Sustainable Energy, published on the ENA portal (05/03/2025). Available at: https://smarter.energynetworks.org/projects/10070764/

6. Appendix

6.1 Summary of treatment groups' pairwise comparisons and their statistical significance

Table 48: Summary of treatment groups' pairwise comparisons

Group 1	Group 2	Turn-up	Turn-down	Comparisons' statistical significance
Low + CB, Med + CB	Low, Med	Significant	Significant	Low + CB & Med + CB significantly outperformed Low & Med in both trials
Low + CB	Low	Significant	Significant	Low + CB significantly outperformed Low in both trials
High	Low	Significant	Significant	High significantly outperformed Low in both trials
Med + CB	Low	Significant	Significant	Med + CB significantly outperformed Low both trials
High	Low + CB	Not significant	Not significant	No significant differences were observed between the two treatments in either trial
High	Med + CB	Not significant	Not significant	No significant differences were observed between the two treatments in either trial
Low + CB	Med	Significant	Not significant	Low + CB significantly outperformed Med in **turn-up trial only**
Med + CB	Med	Significant	Not significant	Med + CB significantly outperformed Med in **turn-up trial only**
Med	Low	Significant	Not significant	Med significantly outperformed low in **turn-up trial only**
High	Med	Significant	Not significant	High significantly outperformed Med in **turn-up trial only**

6.2 Results of balance checks

Randomisation 1 (October 2024): Key findings

- No statistically significant differences were observed between trial groups in terms of Estimated Annual Consumption (EAC), Peak Electricity Use, and Off-Peak Electricity Use.
- The largest observed difference in mean values occurred between Cohort 3 and Cohort 5 (-1.093 kWh) and Cohort 3 and Cohort 6 (-1.095 kWh). However, these values do not exceed conventional thresholds for statistical concern.
- The smallest observed difference was between Cohort 5 and Cohort 6 (0.000 kWh), indicating near-identical baseline characteristics.
- The distribution of values remains balanced across groups, ensuring that randomisation was successful and that there is no indication of systematic bias in allocation.

Table 49: Result of pair wise T-tests (comparing EAC of each group) for randomisation 1

Group 1	Group 2	T-test result
1	2	0.168
1	3	0.805
1	4	0.404
1	5	-0.296
1	6	-0.296
2	3	0.645
2	4	0.239
2	5	-0.465
2	6	-0.466
3	4	-0.407
3	5	-1.093
3	6	-1.095
4	5	-0.698
4	6	-0.7

Group 1	Group 2	T-test result
5	6	0

Randomisation 2: Key findings

- No statistically significant differences were detected between trial groups in terms of Estimated Annual Consumption (EAC), Peak Electricity Use, and Off-Peak Electricity Use.
- The largest observed difference in mean values occurred between Cohort 1 and Cohort 4 (-0.919 kWh) and Cohort 4 and Cohort 5 (0.824 kWh). While larger than in the previous randomisation, these differences remain within acceptable limits for balance.
- The smallest observed difference was between Cohort 3 and Cohort 2 (0.000 kWh), indicating near-identical baseline characteristics.
- Despite minor variations, the distribution remains balanced, confirming that randomisation was successful, and no systematic bias is present.

Table 50: Result of pair wise T-tests (comparing EAC of each group) for randomisation 2

Group 1	Group 2	T-test result
1	2	-0.725
1	3	-0.724
1	4	-0.919
1	5	-0.093
1	6	-0.311
2	3	0
2	4	-0.193
2	5	-0.631
2	6	-0.411
3	4	-0.193
3	5	-0.629
3	6	-0.41
4	5	-0.824

Group 1	Group 2	T-test result
4	6	-0.604
5	6	-0.218

6.3 Primary and secondary regressions without controls

The primary and secondary analysis results presented in the main body of this report are based on regressions that include customer characteristics as covariates and day-time fixed effects (settlement period by day). As a robustness check, we also estimated a corresponding set of regressions without covariates. These yielded larger standard errors, but the point estimates remained broadly consistent with the covariate-adjusted models. The outputs from these unadjusted regressions are provided in tabular form in a supplementary appendix.

Table 51: Primary analysis - Treatment effects on hourly demand (kW) during turn-down events

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value
Control	0.5329			
Low	0.5166	-0.01574	0.006	0.0094
Medium	0.5126	-0.01972	0.006	0.0011
High	0.5116	-0.0208	0.006	0.00052
Low + Consistency Bonus	0.5096	-0.0226	0.006	0.00015
Medium + Consistency Bonus	0.5134	-0.01912	0.006	0.0016

Table 52: Primary analysis - Treatment effects on hourly demand (kW) during turn-up events

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value
Control	0.4184			
Low	0.4512	0.03386	0.0056	1.90E-09
Medium	0.4590	0.04082	0.0058	1.50E-12
High	0.4734	0.0554	0.0062	2.20E-16
Low + Consistency Bonus	0.4700	0.05216	0.0058	2.20E-16
Medium + Consistency Bonus	0.4724	0.05436	0.0058	2.20E-16

Table 53: Secondary analysis - Treatment effects (kW) outside of turn-down events

Trial Arm	Avg. kW outside turn-down events	Regression Coefficient (vs. Control)	Standard Error	P-value
Control	0.42			
Low	0.42	0.00	0.01	0.49
Medium	0.42	0.00	0.01	0.62
High	0.42	0.00	0.01	0.53
Low + Consistency Bonus	0.43	0.01	0.01	0.27
Medium + Consistency Bonus	0.43	0.01	0.01	0.19

Table 54: Secondary analysis - Treatment effects (kW) outside of turn-up events

Trial Arm	Avg. kW outside turn-up events	Regression Coefficient (vs. Control)	Standard Error	P-value
Control	0.43			
Low	0.43	0.00	0.01	0.63
Medium	0.43	0.00	0.01	0.99
High	0.43	0.00	0.01	0.70
Low + Consistency Bonus	0.43	0.00	0.01	0.40
Medium + Consistency Bonus	0.43	0.01	0.01	0.31

6.4 Turn-down analysis excluding critical pricing events

We re-ran our turn-down primary analysis regressions excluding the eight critical pricing events, where, as described above, incentive levels were £1.75/kWh across all groups. While our primary specification includes these events, we are also cautious that this inclusion could dilute treatment versus treatment differences (due to the fact that all treatments were treated the same in them).

We find largely similar results to our main analysis. Each arm's treatment effect is slightly lower than in our main analysis, which is to be expected given the larger effect sizes seen in critical pricing events. However, turn-down is still 3-5% of control group consumption, and the ranking *between* treatment arms remains the same.

Figure 15: Bar chart of treatment effects on hourly demand (kW) during turn-down events, excluding critical pricing events

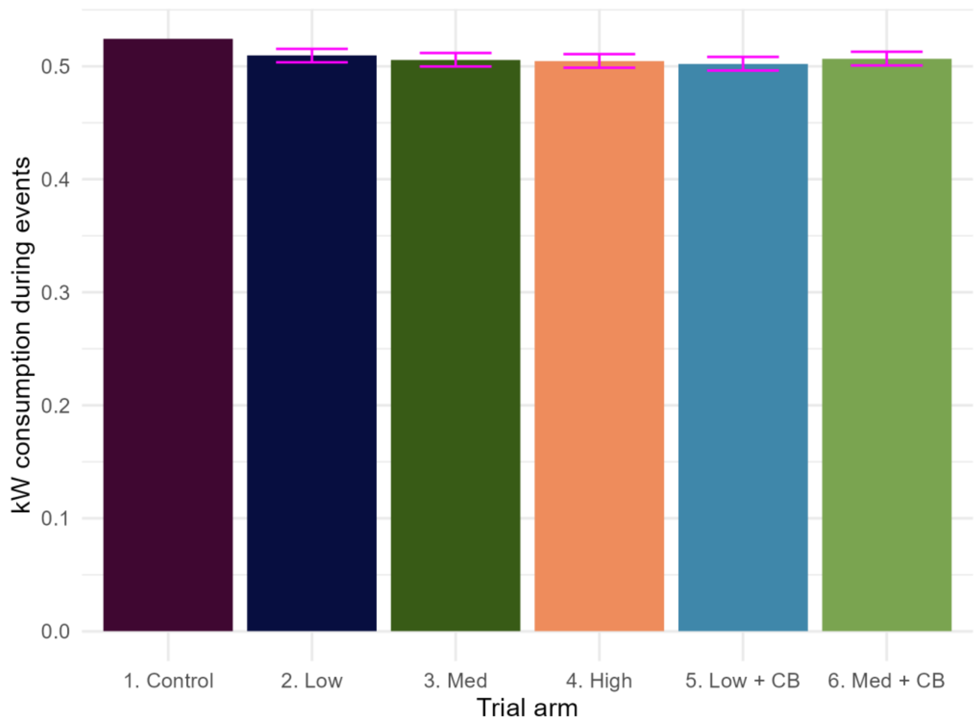


Figure 16: Coefficient plot of estimated treatments effects on hourly demand (kW) during turn-down events, excluding critical pricing events

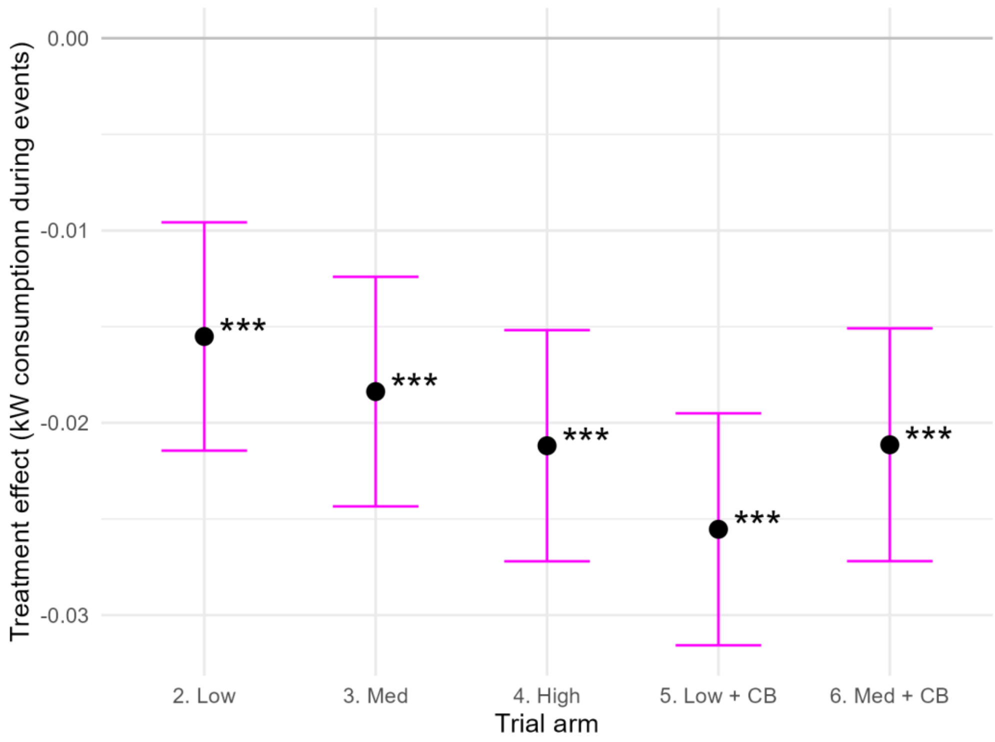


Table 55: Treatment effects on hourly demand (kW) during turn-down events (without critical price events)

Trial arm	Avg. kW demand	Regression coefficient (vs. control)	Standard error	P-value	% difference (regression coefficient vs. control)
Control	0.5772				
Low	0.5549	-0.01551	0.003	3.00E-07	-2.69%
Medium	0.5497	-0.018376	0.003	1.60E-09	-3.18%
High	0.5486	-0.02119	0.003	4.90E-12	-3.67%
Low + CB	0.5491	-0.025536	0.003	2.20E-16	-4.42%
Medium + CB	0.5481	-0.02114	0.003	7.70E-12	-3.66%

6.5 Event schedule

The table below provides a full overview of the scheduled demand response events, including details on timing, direction, and notice type. Each event is categorised by its date, EFA block, weekday type, and duration.

Table 56: Full event schedule

Event date	Event Start	EFA	Day type	Demand response direction	Notice period type	Duration type
08/10/2024	15:00:00	5	Weekday	Down	Short	Short
10/10/2024	12:00:00	4	Weekday	Up	Long	Short
11/10/2024	08:00:00	3	Weekday	Down	Long	Short
15/10/2024	12:00:00	4	Weekday	Anti	Long	Short
19/10/2024	12:00:00	4	Weekend	Up	Long	Short
22/10/2024	20:00:00	6	Weekday	Down	Short	Short
23/10/2024	12:00:00	4	Weekday	Down	Short	Short
24/10/2024	00:00:00	1	Weekday	Up	Long	Long
30/10/2024	20:00:00	6	Weekday	Up	Long	Short
05/11/2024	08:00:00	3	Weekday	Up	Long	Short
07/11/2024	00:00:00	1	Weekday	Down	Long	Short
10/11/2024	12:00:00	4	Weekend	Anti	Long	Long
11/11/2024	12:00:00	4	Weekday	Down	Short	Short
16/11/2024	20:00:00	6	Weekend	Down	Short	Short
17/11/2024	12:00:00	4	Weekend	Up	Short	Short

Event date	Event Start	EFA	Day type	Demand response direction	Notice period type	Duration type
20/11/2024	08:00:00	3	Weekday	Anti	Long	Short
21/11/2024	08:00:00	3	Weekday	Down	Long	Short
23/11/2024	20:00:00	6	Weekend	Anti	Long	Long
24/11/2024	00:00:00	1	Weekday	Up	Long	Long
29/11/2024	12:00:00	4	Weekday	Anti	Long	Long
01/12/2024	08:00:00	3	Weekend	Down	Long	Short
04/12/2024	08:00:00	3	Weekday	Down	Long	Short
04/12/2024	04:00:00	2	Weekday	Up	Long	Long
06/12/2024	12:00:00	4	Weekday	Up	Long	Short
07/12/2024	04:00:00	2	Weekend	Up	Long	Long
09/12/2024	17:00:00	5	Weekday	Down	Long	Short
11/12/2024	20:00:00	6	Weekday	Up	Long	Short
12/12/2024	17:00:00	5	Weekday	Down	Long	Long
14/12/2024	20:00:00	6	Weekend	Up	Long	Short
15/12/2024	08:00:00	3	Weekend	Down	Long	Short
17/12/2024	12:00:00	4	Weekday	Up	Long	Long
18/12/2024	17:00:00	5	Weekday	Down	Short	Short
19/12/2024	17:00:00	5	Weekday	Down	Short	Short
21/12/2024	17:00:00	5	Weekend	Down	Long	Short

Event date	Event Start	EFA	Day type	Demand response direction	Notice period type	Duration type
23/12/2024	12:00:00	4	Weekday	Up	Short	Short
03/01/2025	12:00:00	4	Weekday	Down	Short	Short
06/01/2025	08:00:00	3	Weekday	Down	Long	Short
07/01/2025	17:00:00	5	Weekday	Down	Short	Short
10/01/2025	20:00:00	6	Weekday	Up	Long	Long
11/01/2025	08:00:00	3	Weekend	Down	Long	Short
14/01/2025	17:00:00	5	Weekday	Down	Long	Short
15/01/2025	12:00:00	4	Weekday	Up	Long	Short
17/01/2025	00:00:00	1	Weekend	Up	Short	Short
20/01/2025	20:00:00	6	Weekday	Anti	Long	Short
21/01/2025	08:00:00	3	Weekday	Down	Long	Short
23/01/2025	17:00:00	5	Weekday	Down	Long	Short
24/01/2025	17:00:00	5	Weekday	Down	Long	Short
26/01/2025	20:00:00	6	Weekend	Up	Long	Long
29/01/2025	17:00:00	5	Weekday	Critical Down	Short	Short
30/01/2025	12:00:00	4	Weekday	Anti	Long	Short
02/02/2025	08:00:00	3	Weekend	Down	Long	Short
04/02/2025	17:00:00	5	Weekday	Critical Down	Short	Short

Public

Event date	Event Start	EFA	Day type	Demand response direction	Notice period type	Duration type
05/02/2025	04:00:00	2	Weekday	Up	Long	Short
07/02/2025	12:00:00	4	Weekday	Down	Long	Short
11/02/2025	08:00:00	3	Weekday	Up	Long	Short
13/02/2025	17:00:00	5	Weekday	Critical Down	Short	Short
15/02/2025	12:00:00	4	Weekend	Down	Long	Long
17/02/2025	12:00:00	4	Weekday	Up	Short	Short
21/02/2025	17:00:00	5	Weekday	Critical Down	Short	Short
24/02/2025	17:00:00	5	Weekday	Critical Down	Short	Short
26/02/2025	00:00:00	1	Weekday	Up	Short	Short
01/03/2025	17:00:00	5	Weekend	Up	Short	Long
05/03/2025	17:00:00	5	Weekday	Down	Short	Short
08/03/2025	17:00:00	5	Weekend	Critical Down	Short	Short
09/03/2025	20:00:00	6	Weekend	Up	Short	Long
23/02/2025	20:00:00	6	Weekend	Down	Short	Long
10/03/2025	17:00:00	5	Weekday	Critical Down	Short	Short
12/03/2025	20:00:00	6	Weekday	Down	Long	Short

Event date	Event Start	EFA	Day type	Demand response direction	Notice period type	Duration type
14/03/2025	20:00:00	6	Weekday	Up	Short	Short
17/03/2025	17:00:00	5	Weekday	Down	Short	Long
18/03/2025	00:00:00	1	Weekday	Up	Short	Short
20/03/2025	17:00:00	5	Weekday	Up	Short	Long
25/03/2025	12:00:00	4	Weekday	Up	Short	Long
27/03/2025	17:00:00	5	Weekday	Down	Short	Short
28/03/2025	17:00:00	5	Weekday	Down	Short	Short
31/03/2025	17:00:00	5	Weekday	Critical Down	Short	Short