Public

CrowdFlex report: Availability Trial

Summer 2025

Submitted: December 2025

NESO
National Energy
System Operator

Contents

Executive summary

CrowdFlex is a NESO led innovation project, funded by Ofgem's Strategic Innovation Fund (SIF), which is investigating the potential of domestic flexibility to help operate the grid. CrowdFlex is aiming to establish domestic flexibility as a reliable energy and grid management resource by identifying the technology capability, understanding the statistical nature of flexibility and aligning NESO and DSO requirements. Through large-scale randomised control consumer trials, CrowdFlex is collecting data to develop demand and consumer flexibility prediction models using common APIs. NESO is delivering CrowdFlex with a consortium of industry partners: OVO, Ohme, Centre for Net Zero, ERM, AWS, National Grid Electricity Distribution, Scottish and Southern Electricity Networks, and supported by Smart Grid Consultancy, CGI, Smith Institute and Centre for Sustainable Energy.

As the transition to a decarbonised electricity system accelerates, managing electric vehicle charging demand will become increasingly important for maintaining grid stability. Demand side response programmes provide a mechanism for aligning household energy use with system needs by encouraging customers to shift when and how they charge their vehicles. This field trial examined how availability payments and behavioural interventions can incentivise EV demand flexibility, rewarding customers for being plugged in and available when events are called by grid operators.

The summer availability payments trial built on findings from the previous winter phase, extending the analysis assessing how financial incentives and behavioural prompts could improve not only plug-in frequency but also the quantity of flexibility delivered during demand-side response events. It included 33,060 participants (22,469 Ohme and 10,591 OVO participants) across 60 availability events, making it among the largest programmes of field trials to date investigating EV charging flexibility in the UK. To identify the additional impact of availability payments, over and above the automated dispatch provided by demand response providers, two controls groups were used for both OVO and Ohme trials:

- **Non-dispatched Control group**, which was excluded completely from dispatch during turn-up and turn-down events (but was dispatched for non-CrowdFlex reasons, such as in response to wholesale prices).
- **Dispatched Control group**, which was dispatched for CrowdFlex events automatically, but did not receive any availability payments or interventions to attempt to increase the flexibility provided.

OVO and Ohme operate distinct commercial models that may influence trial outcomes. As a retail energy supplier, OVO offers smart charging as an add-on for customers to obtain a lower-cost unit rate for charging, in exchange for allowing OVO to optimise charging load based on wholesale and other markets. In contrast, Ohme operates as a smart charging provider across multiple retailers, meaning their customer base spans a diverse range of tariffs with varying levels of pre-existing optimisation. These structural differences result in distinct customer relationships, touchpoints, and baseline behaviours. Consequently, the observed differences in percentage changes relative to each one's control groups may reflect these different operating environments and baseline saturations rather than a simple measure of relative provider effectiveness.

Executive Summary Table 1: Key results

OVO		
Primary analysis: % change in demand vs. Dispatched Control (95% confidence interval)		
	Turn-up	**Turn-down**
Non-dispatched control (no availability payment)	**-27.11%** (-31.27%, -22.94%)	**+31.46%** (20.76%, 42.16%)
Escalating payments	**+21.07%** (16.05%, 26.09%)	**0.00%** (-10.25%, 10.25%)
Escalating payments + ready-by-time nudge	**+21.14%** (16.09%, 26.19%)	**-5.59%** (-15.60%, 4.43%)
Escalating payments + ready-by-time payment	**+20.00%** (15.00%, 25.01%)	**-8.37%** (-18.32%, 1.58%)
Secondary analysis: % change in plug-in vs. Dispatched Control (95% confidence interval)		
	Plug-in sessions per week	
Non-dispatched control (no availability payment)	**3.74%** (-0.13%, 7.61%)	
Escalating payments	**+23.2%** (19.11%, 27.34%)	
Escalating payments + ready-by-time nudge	**+25.2%** (21.09%, 29.34%)	
Escalating payments + ready-by-time payment	**+25.2%** (21.10%, 29.33%)	
Ohme		
Primary analysis: % change in demand vs. Dispatched Control (95% confidence interval)		
	Turn-up	**Turn-down**
Non-dispatched control (no availability payment)	**-9.23%** (-18.01%, -0.46%)	**+92.01%** (65.49%, 118.54%)
Standard payments	**+28.97%** (21.78%, 36.16%)	**-14.14%** (-2.95%, 31.22%)
Standard payments + prompt	**+35.45%** (28.22%, 42.69%)	**-4.21%** (-21.06%, 12.65%)
Enhanced payments	**+28.84%** (21.64%, 36.03%)	**-2.56%** (-14.26%, 19.39%)
Enhanced payments + prompt	**+36.41%** (29.15%, 43.67%)	**-11.61%** (-28.31%, 5.09%)
Secondary analysis: % change in plug-in vs. Dispatched Control (95% confidence interval)		
	Plug-in sessions per week	
Non-dispatched control (no availability payment)	**-0.82%** (-8.38%, 6.75%)	
Standard payments	**+32.7%** (26.56%, 38.91%)	
Standard payments + prompt	**+34.7%** (28.58%, 40.88%)	
Enhanced payments	**+31.8%** (25.72%, 37.93%)	
Enhanced payments + prompt	**+36.6%** (30.46%, 42.74%)	

Key findings

1. **Automated dispatch of EVs delivered significant flexibility**
Across both OVO and Ohme trials, automated dispatch of EVs during events produced large, statistically significant shifts in consumption for turn-up and turn-down events, aligned with system balancing needs.

2. **Availability payments substantially increased plug-in frequency**.
All payment structures led to significant rises in weekly plug-in sessions and hours compared with control groups, confirming that financial incentives effectively encouraged customers to make their vehicles available for flexibility. However, how these payments were structured appeared to make relatively little difference. Standard escalating payment structures, without additional remunerated prompts/nudges, also offered lower £/kWh rates for flexibility delivery, suggesting they may be more cost-effective.

3. **Increased availability translated into measurable turn-up, but not turn-down**
Participants receiving availability payments delivered significant demand shifts during turn-up events. For OVO, these payments increased turn-up by roughly 20% relative to the Dispatched Control group (or 27% compared to the Non-dispatched Control, as shown in Table 1 below), while these did little to increase turn-down effect sizes, which were 0-8% compared to the Dispatched Control. For Ohme, turn-up rose by 29-36% relative to the Dispatched Control group, while again turn-down was generally unaffected by increased availability.

4. **Prompts to change customers' app settings had mixed impacts on flexibility response.**
For OVO, RBT nudges and payments successfully shifted customers' chosen Ready-by-Time, but this behavioural change did not translate into significant improvements in turn-up or turn-down flexibility beyond the effect of Escalating Payments. We hypothesise that although customers adjusted their RBTs as intended, the degree of relaxation was too small to materially expand the scope for additional flexibility delivery. Ohme's prompts seemed to increase flexibility delivered in turn-up and even turn-down events, but imbalances between the groups make us cautious in interpreting these results as causal.

5. **Time-of-day effects reflected customer availability and event scheduling.**
The largest turn-up responses occurred overnight and early morning, aligning with typical EV charging patterns, while turn-down responsiveness was more limited during evening peaks – periods when many customers were already avoiding charging.

6. **There were some notable variations in how different customers responded.**
Incentives appeared more effective for vehicles with larger batteries, which delivered not only higher absolute but even higher proportional increases in kW demand during turn-up events. At the same time, lower-consuming households (with annual electricity consumption below 7,300 kWh) showed larger percentage increases in

plug-in hours, suggesting a stronger behavioural response from these customers. While this analysis is non-causal, it may imply different responses based on vehicle type or household profile, for providers to consider in optimising participation.

Overall, the summer 2025 trials reaffirmed and extended insights from the previous winter: automated dispatch delivers reliable flexibility, incentives meaningfully shape when customers plug in their EV, but additional flexibility delivered depends on times that events are called. Integrating availability payments into mainstream commercial flexibility products – aligned with event timing, vehicle mix, and consumer characteristics – could unlock substantial additional flexibility at low cost, further supporting the transition to a renewables-based grid. While straightforward and predictable payments clearly deliver additional flexibility, more complex and incremental incentive structures may deliver diminishing returns

1. Introduction

1.1 Background and rationale

Managing EV charging demand is increasingly important for electricity system stability as EV adoption accelerates. Demand side response (DSR) programmes offer a mechanism to align EV charging behaviour with grid needs by encouraging customers to shift when and how they charge their vehicles. CrowdFlex, a programme combining model development with randomised and matched controlled trials, seeks to provide rigorous evidence on the design and impact of availability payments to enhance demand flexibility. This, in turn, supports the operation of a renewables-based electricity grid and the UK's broader decarbonisation objective.

During CrowdFlex's winter 2024 trials, we observed that escalating payment structures were effective at increasing plug-in rates compared to flat payments. Behavioural interventions, such as informational prompts, also showed potential for influencing customer behaviour but were not directly targeted at maximising flexibility delivered (i.e. kWh shifted) during DSR events.

This report presents findings from the summer availability trials, which ran between July and September 2025. Building on the winter results, these trials explored how both financial incentives and behavioural prompts could improve not only plug-in frequency but also the volume of flexibility delivered. A key question was whether such interventions could effectively shift charging behaviour in ways that generate measurable grid benefits, particularly during periods of high renewable generation or peak demand.

The trials were delivered by two Demand Side Response Service Providers (DSRSPs): OVO and Ohme. While both investigated how availability payments and behavioural interventions influence EV charging behaviour, their research questions differed slightly. This reflects the distinct business priorities and operational models of the participating organisations. Ohme focused on how varying incentive sizes and non-financial prompts influenced general plug-in behaviour and kW demand across a broad EV customer base. OVO, meanwhile, examined how customers could be encouraged - through both messaging and financial rewards - to adjust specific charging parameters such as their "ready-by-time," (RBT) which directly impacts the DSRSP's ability to call upon flexible capacity. These different angles allowed the trials to test complementary behavioural levers and operational strategies, while contributing to the shared goal of understanding how availability payments can enhance system flexibility through EV charging.

Given these differences, this report is structured into two sections: one focused on OVO's trial and the other on Ohme's. This structure allows the findings from each trial to be presented clearly and in context, ensuring the results are interpreted appropriately for the distinct research aims and operational models of each partner.

2. Trial design

2.1 Event parameters and variability

The trial ran between 07/07/2025 - 28/09/2025 and included both turn-up and turn-down DSR events, scheduled at varied times and durations to reflect the diversity of conditions typically experienced on the electricity grid during the summer period. The event design aimed to simulate realistic operational scenarios under which flexibility services might be called, ensuring that the trial outcomes would be relevant and applicable to real-world grid management.

Table 1: Key dates for Summer availability trial

Tasks	Start	End
Summer availability payments trial	07/07/2025	28/09/2025

The events defined our primary analysis, in which we examined treatment effects on consumption during the events (separately for turn-up and turn-down). They also served as the underlying dataset for the Modelling workstream, which uses the same events to predict demand and flexibility outturn, and to assess prediction accuracy.

2.1.1 Turn-up events

A total of 40 turn-up events were conducted, distributed across multiple times of day to capture a range of operational contexts, including daytime, evening, and overnight hours. The most frequent start times for turn-up events were clustered between 11:00 and 14:00, as well as during overnight periods between 23:00 and 04:00, reflecting periods when surplus renewable generation or low system demand may occur.

Event duration ranged from 30 minutes to 2 hours. This range of durations was chosen to test flexibility across short and extended periods, mirroring potential grid needs during times of excess generation and grid constraints in summer.

Turn-up events varied in duration as follows:

- 30 minutes: 11 events
- 1 hour: 17 events
- 1.5 hours: 5 events
- 2 hours: 7 events

2.1.2 Turn-down events

A total of 20 turn-down events were executed, scheduled primarily during morning and evening peaks when demand reduction is more likely to be valuable for grid balancing. The most frequent start times for turn-down events were between 06:00 and 08:00, and between 17:00 and 20:00, reflecting typical periods of higher electricity demand.

Durations for turn-down events were distributed as follows:

- 30 minutes: 6 events
- 1 hour: 7 events
- 1.5 hours: 5 events
- 2 hours: 2 events

Table 2: Event distribution

Hour	Turn-up	Turn-down
00:00	3	0
01:00	3	0
02:00	3	0
03:00	4	0
04:00	3	0
06:00	0	2
07:00	0	1
07:30	0	2
08:00	0	4
11:00	5	0
12:00	5	0
13:00	5	0
14:00	5	0
17:00	0	1
17:30	0	2
18:00	0	2
18:30	0	1
19:00	0	3
20:00	0	2
23:00	4	0
Total	**40**	**20**

2.2 Control group structure

Matched control groups were used to establish a baseline for comparison, helping to identify whether observed changes in charging behaviour were a result of the trial interventions or would have occurred anyway. As in the previous Winter trial, both OVO and Ohme included control groups to distinguish typical charging patterns from those influenced by availability payments and behavioural prompts.

To ensure the effects of the availability payments and behavioural interventions could be rigorously evaluated, both OVO and Ohme trials incorporated two distinct control groups. These groups allowed the analysis to separate normal, automated behaviour from any changes directly attributable to the treatments.

- **Non-dispatched Control:** Customers in this group were excluded from CrowdFlex dispatch during both turn-up and turn-down events, providing a baseline measure of natural charging behaviour. However, these customers could still be dispatched for other operational reasons unrelated to the trial (e.g. wholesale price optimisation).
- **Dispatched Control:** Customers were automatically dispatched during CrowdFlex events as part of normal business-as-usual operation, they but did not receive availability payments or behavioural prompts. This group therefore represents standard flexibility delivery without additional incentives.

By comparing the treatment groups to both control groups, the trial was able to isolate the incremental impact of availability payments and behavioural interventions over and above the baseline level of flexibility already achieved through automated dispatch in response to other system incentives and constraints (e.g., wholesale prices and customer preferences).

To create these groups, OVO and Ohme both used matching to create a pool of matched control customers; they then randomised within that pool to create the Non-dispatched Control and Dispatched Control groups.

2.3 Sample sizes and power

To ensure the trial would yield dependable results, power calculations were performed using data collected from the previous Winter 2024 Availability trial. This analysis confirmed that the Summer 2025 trial design was well-powered to detect meaningful differences in participant behaviour, such as changes in energy delivered or how often people plugged in.

The final analysis included 28,753 participants. While the OVO sample experienced a 22.14% drop-out rate, primarily because participants left OVO or dropped the *Charge Anytime* add-on, the study remained well-powered. The Ohme sample had a higher participant count with a much lower drop-out rate of 5.27%, possibly reflecting the relative ease of switching energy supplier versus Charge Point Operator.

Details of each DSRSP's sample size are provided in their respective methodology sections.

2.4 Data collection

Data were collected through telemetry systems integrated with EV chargers, providing the necessary data regarding plug-in frequency and electricity consumption.

- Primary outcome: This measure focused on energy consumption (half-hourly kWh, converted to kW for ease of interpretation), used to evaluate the extent to which higher plug-in frequencies resulted in increased flexibility capacity for demand turn-up and turn-down events. Note that both OVO and Ohme used consumption data from the charger itself, not from customers' smart meters.
- Secondary outcomes: Plug-in frequency was measured in two ways: "qualifying" plug-in sessions per week (qualifying sessions were defined as being plugged in for at least 6 hours during a day, from 12 pm to 12 pm the following day, for OVO; and for at least 6 hours during a day, from 6am to 6am, for Ohme) and hours plugged in per week.

Data collection was continuous throughout the trial period, capturing variations across different Electricity Forward Agreement (EFA) blocks and trial arms.

2.5 Outcome measures

The trial assessed how financial incentives and behavioural signals affected EV availability and the delivery of flexibility. The outcomes were as follows, structured by level of priority:

Table 3: Key analysis outcome measures

Outcome	Definition	Measurement Method
Primary: Consumption (kWh) – for the purpose of assessing kWh flexibility response	Total kWh consumed during called event periods – where turn-up and turn-down events are analysed separately, thus implying two separate outcomes / analyses	kWh consumed during scheduled event windows (dispatch periods)
Secondary: Plug-in frequency	Number of qualifying plug-ins per customer per week	Charging telemetry data, filtered for sessions >6 hours
Exploratory: Flexibility capacity	Weekly plug-in duration × state of charge (SoC) at arrival	Telemetry data (session start/end times and SoC, where available)

2.6 Statistical methods

The analysis was conducted on balanced panel datasets generated from participant-level daily data covering 7 July to 28 September 2025. Each dataset included trial arm identifiers, participant characteristics (battery size, customers' estimated annual consumption (EAC) of electricity from their smart meter, for OVO, and from their charger, for Ohme EAC), and event metadata (date, week, weekday/weekend, time-of-day, event type).

Analysts estimated treatment effects using regression models.[1] For each primary and secondary outcome, regression models were run separately for *turn-up* and *turn-down* events.

2.7 DSRSP dispatch strategy

The DSRSPs employed somewhat different dispatch strategies during the trial.

- Ohme executed the NESO dispatch API without any optimisation around the event, following standard dispatch instructions.
- OVO also used the NESO dispatch API but applied additional optimisation during the four-hour notice period. OVO's optimiser did not explicitly aim to avoid or increase charging before an event. Instead, for turn-up events, it re-optimised customers' charging schedules to maximise the amount of charging that could occur during the turn-up event, subject to each customer's RBT and state-of-charge. In practice, this often meant reducing charging before a turn-up event to avoid reaching a high state-of-charge too early, but in other cases, such as overnight events, it could still mean charging both before and after the event. For turn-down events, the optimiser treated the event period as a high-cost time, re-optimising charging around it so that the largest feasible reduction in demand could be delivered during the event itself. This could involve charging before or after the event depending on what best satisfied the customer's RBT while maximising the achievable reduction during the turn-down window.

All else being equal, we would thus expect slightly higher impacts of dispatch from OVO than from Ohme. On the other hand, as OVO already optimises against wholesale signals for its *Charge Anytime* customers, they might have already accessed much of the available latent flexibility. In contrast, Ohme's diverse customer base, specifically those with no prior optimisation, potentially offered greater capacity for incremental load shifting, which could have yielded higher relative impacts during events.

[1] Unless otherwise specified, models included week fixed effects to control for temporal shocks and clustered SEs at the participant level to account for repeated measures.

3. OVO - Research questions and hypotheses

Below, we set out the core research questions and accompanying hypotheses that underpin analysis of OVO's trial.

3.1 Primary research question: interventions' effect on kW demand

1. Effect of dispatch: What is the impact of dispatch on kWh, comparing the Dispatched Control group to the non-dispatched one?
2. Effect of plug-in interventions: What are the impacts of escalating plug-in frequency payments, messages to relax ready-by-time (RBT, i.e., the user-specified deadline by which a customer requires their EV to reach the desired state of charge), and messages plus payments to relax RBT, on kW demand during DSR events, when compared to the Dispatched Control group?

3.1.1 Primary research question: Hypotheses

- Impact of dispatch: Customers in the Dispatched Control group will deliver significantly more kWh flexibility during DSR events than those in the Non-dispatched Control group.
- Impact of escalating payments: Offering an escalating payment will lead to additional kWh flexibility delivery compared to the Dispatched Control group, further widening the gap relative to the non-dispatched group.
- Impact of escalating + RBT nudge: Providing a ready-by-time message will also increase kWh flexibility delivery relative to the Dispatched Control group.
- Impact of escalating + RBT payment: The combined treatment (standard escalating payments + ready-by-time payment) will result in the highest kWh flexibility delivery of all groups, though the increase beyond the Dispatched Control group will be smaller than the sum of the individual effects of escalating payments and the behavioural prompt.

3.2 Secondary research questions: interventions' effect on plug-in hours per week

1. What are the effects of escalating plug-in frequency payments, incentives to relax ready-by-time, and the combination of the two, on plug-in hours per week during DSR events, when compared to the Dispatched Control group?

3.2.1 Secondary research question: Hypotheses

- No impact of dispatch: Customers in the Dispatched Control group will deliver the same level of weekly plug-ins and plugged-in hours as those in the Non-dispatched Control group, on average.
- Impact of escalating payment: Offering escalating payments will lead to higher weekly plug-ins compared to the Dispatched Control group, further widening the gap relative to the non-dispatched group.

- Impact of escalating payments + RBT nudge: Providing a behavioural prompt may increase weekly plug-ins relative to escalating payments only, given that relaxing one’s ready-by-time may cause longer plug-ins.
- Impact of escalating payments + RBT payment: The combined treatment (standard escalating payment + ready-by-time payment) will also increase weekly plug-ins compared to escalating payments only, given that relaxing one’s ready-by-time may cause longer plug-ins.

3.3 Exploratory research questions

Exploratory research questions examined treatments’ impacts on customers ‘dispatchable availability’ (state of charge * hours plugged in), weekly charging, and accuracy of P376 baseline; differential treatment effects by subgroup; and the extent to which the nudges/prompts to change app settings were successful in doing so.

4. OVO Methodology

4.1 Trial design, eligibility, and interventions

This trial tested five groups to examine how different payment structures and messages affect when and how long customers plug in their electric vehicles for flexibility services. The primary outcome was the total flexibility (kWh) delivered per participant per week during DSR events, analysed separately for turn-down and turn-up events. In groups receiving a message, the purpose of the message was to encourage participants to amend their ready-by time, enabling greater flexibility.

For the three treatment groups, escalating payments were offered to encourage customers to plug their EVs in. Customers were paid £1/week for ≥3 plug-ins, £2/week for ≥4 plug-ins, and £3/week for ≥5 plug-ins.

The “ready-by-time” interventions tested whether asking customers to extend their ”ready-by time" (RBT) adding specific time targets – either as a message, or with an additional reward[2] – increased the usefulness of their availability. Dispatch itself is automated: if a customer’s EV is plugged in during an event window, their vehicle can be used to provide flexibility without the customer needing to take any action at the time. Note that dispatch is part of a general optimisation process that prioritises CrowdFlex response but also includes other price signals, particularly wholesale prices by settlement period (half-hour) of the day.

Table 4: OVO trial arms

Trial Arm	Trial Arm Name	Standard Escalating Payment	Ready-by-Time Prompt	Ready-by-Time Reward
Control 1	Non-dispatched Control	No	No	No
Control 2	Dispatched Control	No	No	No
Treatment group 1	Escalating Payments	Yes	No	No
Treatment group 2	Escalating + RBT Nudge	Yes	Yes	No
Treatment group 3	Escalating + RBT Payment	Yes	Yes	Yes

Table 5: OVO sample size

Trial arm	# Start of trial	# End of trial	Total drop-out (%)
Non-dispatch control	2,661	2,110	20.71%
Dispatch control	2,671	2,197	17.75%

[2] OVO sent two prompts per week to each participant, inviting them to modify their ready-by-time to a specified increase in ready-by-time. If a participant acted on the prompt and adjusted their ready-by-time accordingly, they earned £1 per instance. This meant participants could earn up to £2 per week by responding to both prompts.

Trial arm	# Start of trial	# End of trial	Total drop-out (%)
Escalating payments	2,690	2,118	21.26%
Escalating + RBT nudge	2,695	2,072	23.12%
Escalating + RBT payment	2,686	2,094	22.04%
Total	13,403	10,591	22.14%

4.2 Eligibility and recruitment

To be eligible for participation in this trial, customers were required to be enrolled in OVO's Charge Anytime add-on, which enables flexible charging of electric vehicles. Eligible customers were subsequently invited to participate in the CrowdFlex trial and were required to provide explicit opt-in consent.

4.3 Randomisation and group allocation

The trial used a combination of matching and randomisation. Customers who signed up to the trial were randomly assigned to one of the three Treatment groups. OVO then used matching to create two control groups, selecting matches from a pool of customers who had the *Charge Anytime* add-on but had not signed up to the trial, based on their similarity to the signed-up customers. These matched customers were then randomly assigned to one of two Control groups.

Randomisation of treatment groups: Randomisation is important for ensuring that any differences observed between groups can be reliably attributed to the interventions under study, rather than to pre-existing differences or other external influences. Randomly allocating participants to different trial arms helped create groups that are similar on both measured and unmeasured characteristics, providing a robust foundation for drawing conclusions about the effectiveness of different incentives in influencing electricity consumption behaviour.

Randomisation in the trial was conducted with stratification by Grid Supply Point (GSP). Stratification ensured balanced representation of geographic areas across trial arms, accounting for potential regional variation in grid conditions and customer characteristics. Participants were individually randomised into one of the three treatment groups using simple random allocation within each GSP stratum, following a 1:1:1 ratio. This design aimed to enhance comparability and minimise potential confounding.

Before analysing trial outcomes, OVO conducted balance checks to assess whether randomisation resulted in comparable groups across key baseline characteristics. Specifically, distributions were examined for:

- Region (via GSP group)
- Tenure with OVO (in years)
- Historic electricity consumption (kWh in June)
- Average weekly plug-in hours during June (past 3 months)
- Median “ready-by” hour (typical time participants finish charging)

Statistical tests indicated no significant differences between groups across these variables, suggesting successful randomisation and well-balanced trial arms. Results are detailed in the appendix.

Matching: In addition to the randomised groups, matched control groups were created using nearest-neighbour matching without replacement, based on variables including region, average plug-in hours, and electricity usage from the charger itself. This multivariate matching process minimised baseline differences between trial and control customers, helping to isolate the effects of the interventions.
While matching cannot ensure internal validity to the same extent as randomisation, the combination of randomisation, stratification, and matched controls, together with accompanying balance checks, provides a strong basis for attributing observed differences in outcomes to the interventions under study.

Randomisation of control groups: Finally, OVO conducted a simple randomisation of its matched control group pool into Dispatched and Non-dispatched Control groups. Results of the balance checks and matching can be found in the appendix.

4.4 Sample exclusions and cohort definitions

In late June 2025, OVO identified approximately 1,000 recruited participants whose "ready-by" times (RBTs) were either undefined or occurred too late in the day to support meaningful intervention. These participants were therefore excluded from the main randomisation process, as they could not be allocated to time-sensitive treatment arms (Groups 2 and 3).

This left a core sample of approximately 8,071 participants with valid RBTs, who were individually randomised into one of the three treatment groups. All primary analyses will focus on this core sample and their corresponding matched control groups. Outcomes associated with Group 1 will also exclude the 1,000 participants with invalid or late RBTs to maintain consistency across groups.

To construct an appropriate counterfactual, matched control groups were drawn from the wider population of Charge Anytime customers who were eligible but did not enrol in the trial. Two distinct cohorts were defined:

- Core cohort control group: Matched to the 8,071 participants with valid RBTs
- Late/unspecified RBT cohort control group: Matched to the 1,000 participants excluded from randomisation

Each matched control group was then randomly split into dispatched and non-dispatched subgroups, yielding a total of four matched control groups:

- Core cohort, dispatched
- Core cohort, non-dispatched
- Late/unspecified/irregular cohort, dispatched
- Late/unspecified/irregular cohort, non-dispatched

This structure allowed for clean identification of dispatch effects, while preserving consistency within and across cohorts.

4.5 Ready-by-time prompt schedule

Ready-by-time prompts were delivered to customers on a weekly basis throughout the trial period. The table below summarises the number of prompts issued each week (Weeks 28–40).

In most weeks, two prompts were sent. Occasional variation occurred, with as few as one prompt (Weeks 33, 39, and 40) and as many as three prompts (Week 34). In total, 24 prompts were delivered across the 13-week period.

This consistent scheduling provided regular exposure to the intervention while allowing some flexibility in timing. A more in-depth prompt schedule is provided in the appendix.

Table 6: Ready-by-time monthly prompt schedule

Month	# of prompts per month
July	7
August	9
September	8
Grand Total	24

5. OVO Results

5.1 Primary Analysis: Interventions' effect on kW demand during trial events

5.1.1 Turn-up events

The Dispatched Control group demand was 0.569 kW on average during turn-up events over the course of the trial. The Non-dispatched Control group demand was significantly lower, delivering 0.412 kW, which represents a significant −27.11% reduction compared to the Dispatched Control (or, equivalently, a 38.1% increase in the Dispatched Control group's demand versus the Non-dispatched Control group).

All three treatment arms showed a significant increase ($P<0.001$) in kW demand over the Dispatched Control:

- The Escalating Payments group (0.699 kW) recorded a 21.07% increase.
- The Escalating + RBT Nudge group (0.709 kW) recorded a 21.14% increase.
- The Escalating + RBT Payment group (0.695 kW) recorded a 20% increase.

These effects, when pooled, indicate an approximately 70% increase in demand in the treatment groups compared to the Non-dispatched Control group. In other words, dispatch had a very meaningful impact on demand, but the treatments, likely by causing more plug-in overall, increased turn-up above and beyond this impact of dispatch.

There was no statistically significant difference in performance found between the three individual treatment arms, suggesting that the RBT elements did not substantially enhance the turn-up kW demand effect achieved by the basic Escalating Payments structure.

Figure 1: Average kW demand by treatment arm during turn-up events

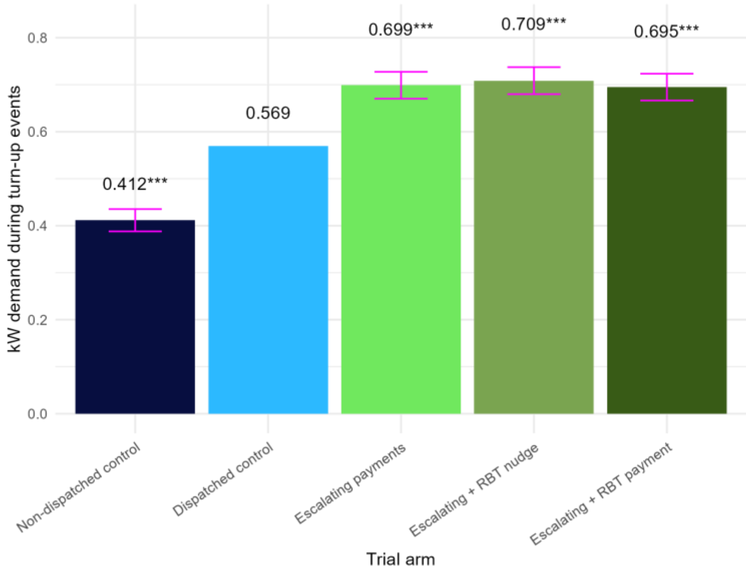
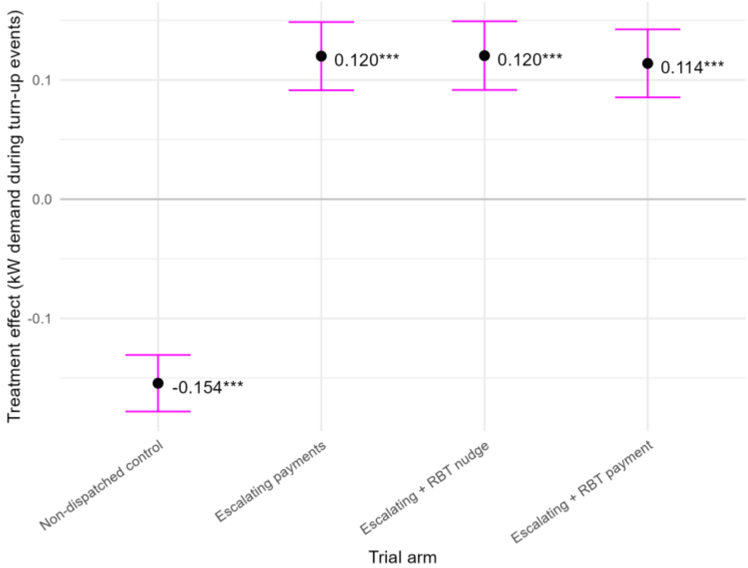


Figure 2: Estimated treatment effects (coefficients) on kW consumption during turn-up events (relative to Dispatched Control)



5.1.2 Turn-down events

The Non-dispatched Control group recorded the highest average demand at 0.158 kW, which was a statistically significant increase of 31.46% compared to the Dispatched Control group (0.121 kW).

The Dispatched Control group and all treatment groups showed a significant demand reduction during trial events. However, the difference between the treatment groups' reductions and the Dispatched Control was not significant.

- The Escalating Payments group (0.123 kW) showed a difference of 0% (not significant).[3]
- The Escalating + RBT Nudge group (0.119 kW) showed a non-significant reduction of −5.59%.
- The Escalating + RBT Payment group (0.114 kW) showed a non-significant reduction of −8.37%.

[3] The bar chart does not exactly match this coefficient because bar heights are not adjusted by covariates. We rely on the covariate-adjusted coefficients, as shown in the coefficient plots, in reporting differences between groups.

Figure 3: Average kW demand during turn-down events

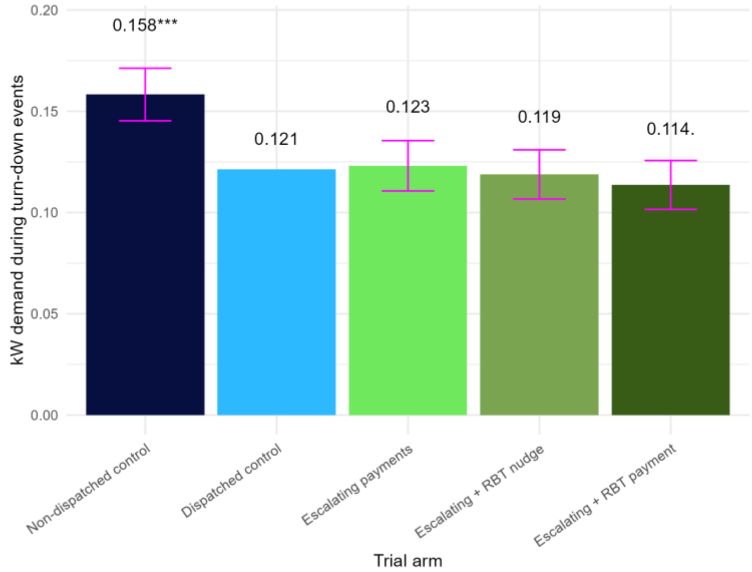
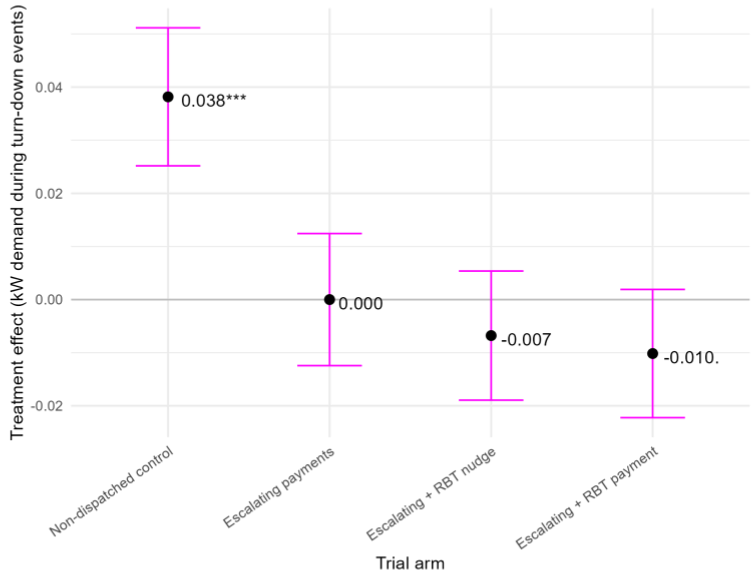


Figure 4: Estimated treatment effects (coefficients) on kW consumption during turn-down events (relative to Dispatched Control)



These results indicate that the treatments did not increase turn-down capacity above and beyond the impact of dispatch – even though they did increase plug-in, as discussed in secondary analysis. It is interesting to consider why the treatments improved turn-up but not turn-down capacity. We hypothesise two potential reasons:

1. Turn-down events tended to occur at times of day with less additional plug-in among the treatment groups than during turn-up events. (See Section 5.4.2.1 for plug-in by time of day.)
2. Extra plug-in may indirectly – perhaps too indirectly to detect, given our sample size – help turn-down, by providing the DSRSP with more alternative charging hours. By

contrast, it much more directly, and therefore perhaps more substantively, helps turn-up, by giving the DSRSP more vehicles plugged in the 1-2-hour window when it is trying to increase charging.

In addition, it is worth noting that the impact of dispatch was itself smaller for turn-down than for turn-up – while dispatch did cause turn-down, the effect size was smaller than for turn-up (0.038 kW turn-down, versus 0.154 kW turn-up). We have two hypotheses for this pattern:

1. The trial examined explicit flexibility – above and beyond implicit flexibility already being provided by OVO in optimising vehicles' charging demand against wholesale prices. By *implicit* flexibility, we refer to demand shaping against wholesale prices and other market signals. This is important to keep in mind; OVO typically already reduced demand during times when turn-down events were called because these tended to feature higher wholesale energy prices. This dynamic is evidenced by lower demand in the Non-dispatched Control during turn-down events (0.158) compared to turn-up events (0.412). We reason that because demand has already been minimised in response to wholesale energy price signals, an additional *explicit* trial instruction leaves little further demand to be reduced.
2. By design, demand turn-up can be delivered by "saving up" charging for later in the session, whereas demand turn-down requires charging to be spread thinly across the entire session. For example, with 10 kWh to be charged over a 15-hour plug-in, turn-up could be achieved by delaying charging until the end (delivering ~7 kW response for most of the session), while turn-down would limit response to the much lower average rate (~0.8 kW).

5.2 Secondary Analysis

5.2.1 Interventions' effect on plug-in sessions per week

The average number of weekly plug-in sessions was significantly higher across all three treatment arms compared to the control groups. The Non-dispatched Control group (2.33 sessions) recorded 3.74% more sessions than the Dispatched Control (2.25 sessions), though this difference was not statistically significant (P= 0.0582). We would not have expected increased plug-in between these two groups; the difference here may be statistical noise.

All treatment arms showed a significant increase over the Dispatched Control: the Escalating Payments group (2.81 sessions) recorded a 23.22% increase in sessions; the Escalating + RBT Nudge group (2.88 sessions) recorded a 25.21% increase; and the Escalating + RBT Payment group (2.86 sessions) recorded a 25.22% increase. The overall effect of the Pooled Treatments was significant when compared to the controls.

There was no statistically significant difference found when comparing the three treatment arms to each other, suggesting that the addition of the RBT elements did not substantially enhance the effect of the basic Escalating Payments structure. This is somewhat to be expected, as these RBT elements did not relate to plug-in frequency, as they were instead an effort to provide OVO more scope to use the provided plug-in for demand response.

Figure 5: Average weekly plug-in sessions by trial arm

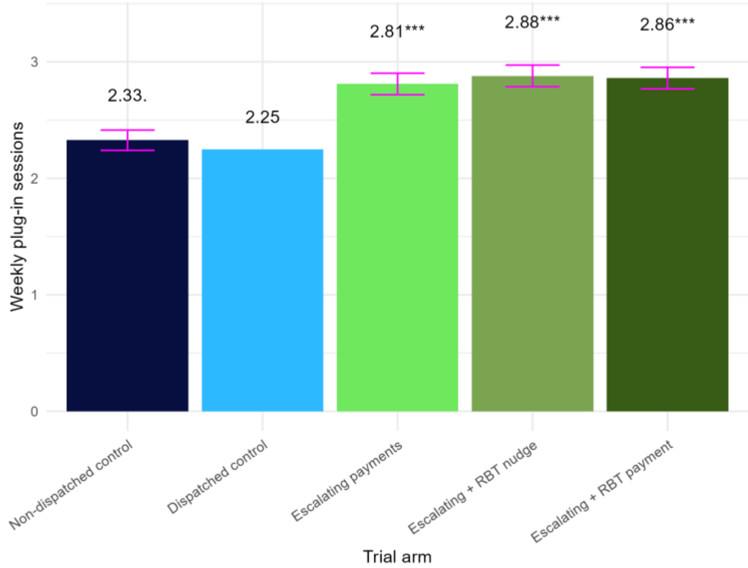
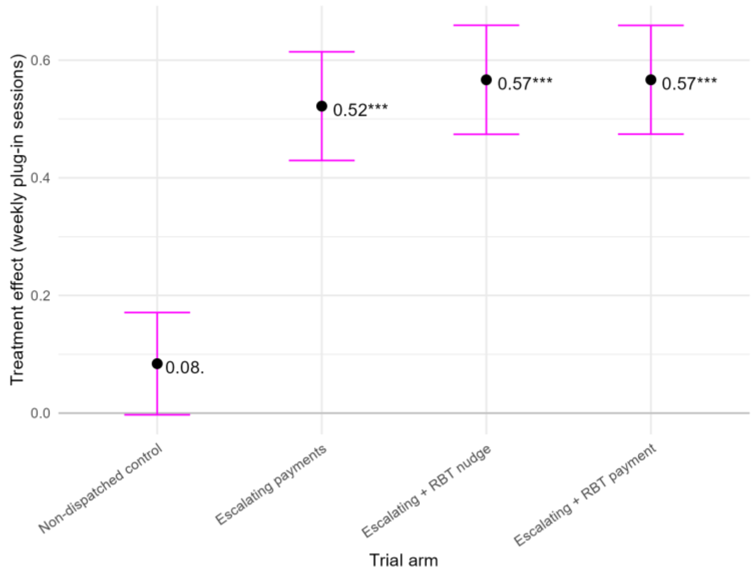


Figure 6: Estimated treatment effects (coefficients) on plug-in sessions (relative to Dispatched Control)



5.2.2 Interventions' effect on plug-in hours per week

The analysis of average weekly plug-in hours showed that all three treatment arms resulted in a significantly higher number of hours compared to the control groups. The Dispatched Control group had an average of 36.7 hours, while the Non-dispatched Control group averaged 37.8 hours, representing a 3.26% difference that was not statistically significant (P=0.13). In contrast, all incentive-based groups showed a significant increase (P<0.001):

the Escalating Payments group averaged 47.0 hours, a 26.42% increase; the Escalating + RBT Nudge group averaged 48.3 hours, a 29.04% increase; and the Escalating + RBT Payment group averaged 47.9 hours, a 28.79% increase. The overall picture of trial arms' plug-in in terms of hours plugged mirrored the picture in terms of plug-in *sessions* very closely.

Figure 7: Treatment effects on average weekly plug-in hours

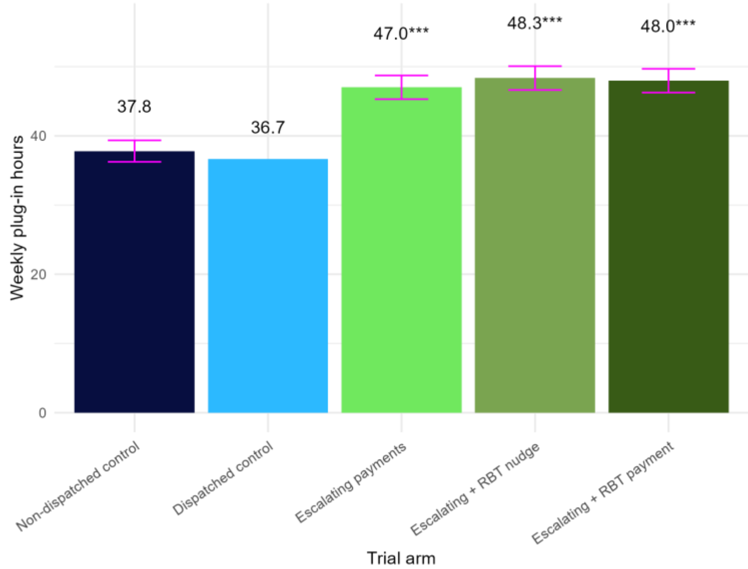
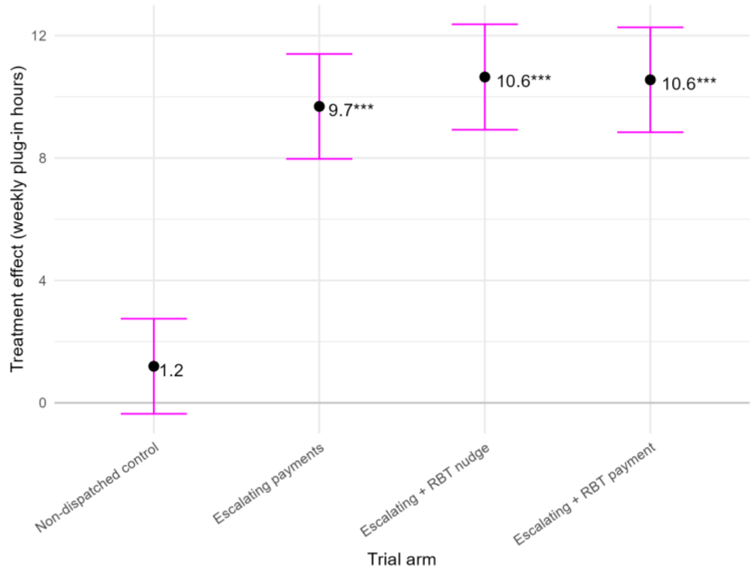


Figure 8: Estimated treatment effects (coefficients) on plug-in hours



5.3 Exploratory analysis

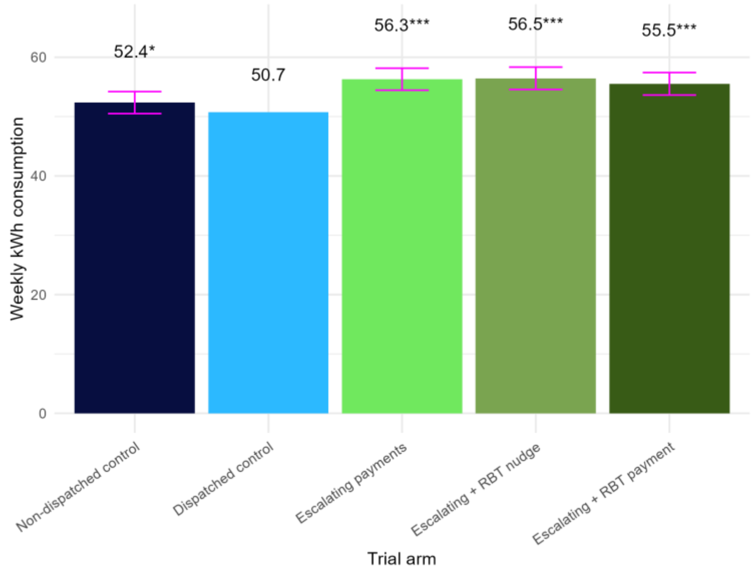
5.3.1 Interventions' effect on weekly consumption

The analysis of average weekly kWh consumption showed a significant increase for all three treatment arms compared to the Dispatched Control, and a significant difference between the two control groups.

The Non-dispatched Control group (52.36 kWh) recorded 3.80% more consumption than the Dispatched Control (50.73 kWh). This difference was statistically significant (P<0.05). We are not certain for the reason here. It may be that dispatch overall reduced charging at home (perhaps compensated by more charging away from home). For example, a turn-up event at 04:00 might cause OVO to delay or reduce charging 00:00 to 04:00; some customers might take their car off charge at 07:00 not fully charged; and this could cause a reduction in overall home charging. However, there are also some indications that there was a mild imbalance between the groups despite the randomisation (due to chance) leading the Non-dispatched Control to have slightly higher consumers than the Dispatched Control.

All treatment arms showed a significant increase (P<0.001) over the Dispatched Control: the Escalating Payments group (56.30 kWh) recorded a 9.01% increase; the Escalating + RBT Nudge group (56.45 kWh) recorded a 7.40% increase; and the Escalating + RBT Payment group (55.53 kWh) recorded a 6.96% increase. The overall effect of the Pooled Treatments was significant when compared to the pooled controls. However, there was no statistically significant difference found among the three individual treatment arms, suggesting that the RBT elements did not increase the effect of the basic Escalating Payments structure on weekly kWh consumption. We hypothesise that the treatments led to demand creation at home, where the mechanism would be that more time plugged in ends up causing more charging overall, at least at home (and perhaps offset by less charging away from home, though there may also be more driving overall).

Figure 9: Treatment effects on average weekly consumption



5.3.2 Event studies: dynamic kW response around flexibility events

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

5.3.2.1 Turn-up events - Treatment vs. Dispatched Control

This comparison measures the incremental gain from the financial incentive above the baseline represented here by the Dispatched Control. The financial incentive provided a statistically significant, incremental lift in kW demand:

- **Pre- and Post-Event Periods:** In the SPs before (−12 to −1) and after (1 to 12) the event, there is a pattern of a small amount of higher consumption, likely related to the slightly higher overall consumption caused by plugging in more (in general) that is also observed in analysis of weekly consumption.
- **During Event Period (e1 to e4):** A clear, statistically significant spike was observed. The incentive treatments added an incremental 0.09 to 0.14 kW demand *above* the Dispatched Control, demonstrating that the financial incentive reinforced the technical dispatch signal by making more vehicles available.

Figure 10: Differential kW response during turn-up events: Treatment vs. Dispatched Control (settlement period-series)

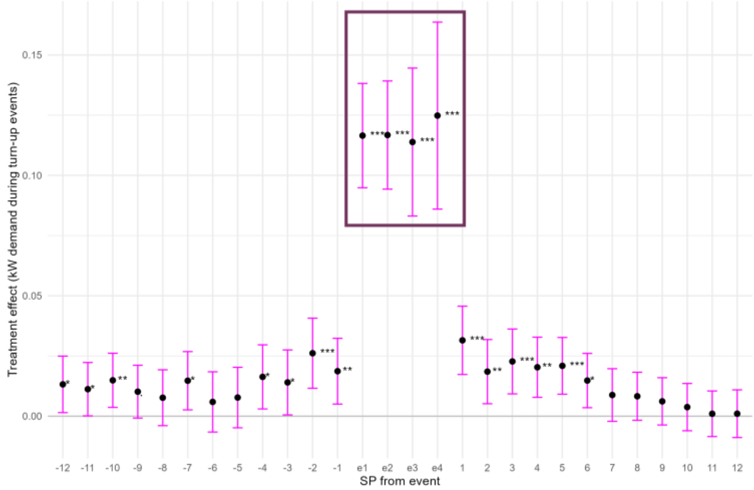
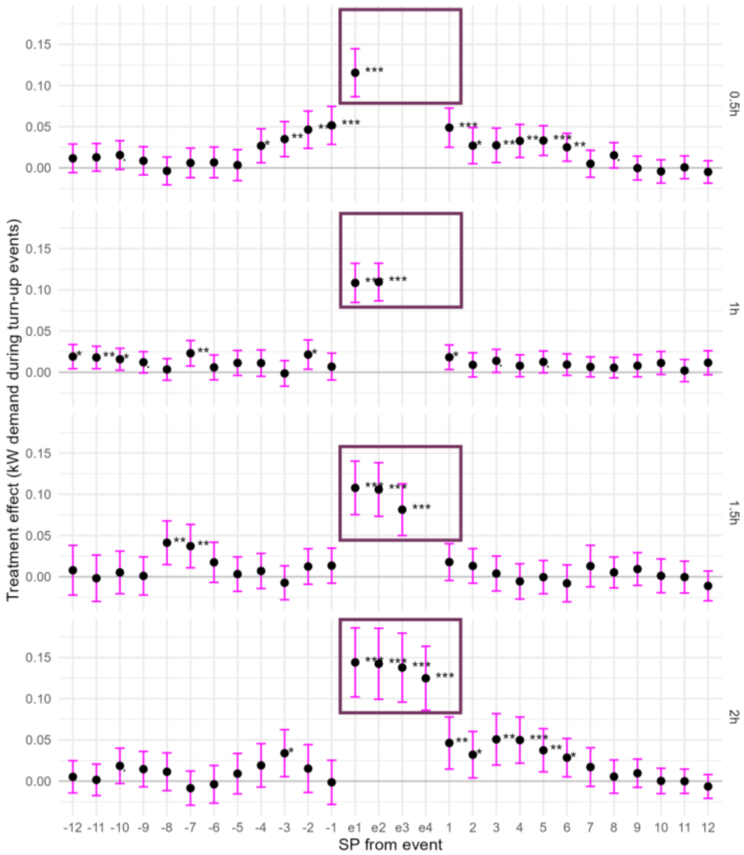


Figure 11: Differential kW response during turn-up events: Treatment vs. Dispatched Control (settlement period-series) split by event duration



5.3.2.2 Turn-up events - Treatment vs. Non-dispatched Control

This comparison measures the total impact of the intervention (incentive plus dispatch signal) relative to a completely passive baseline.

- **Pre- and Post-Event Periods:** There did not appear to be anticipatory displacement, but there is some evidence of post-event displacement. As in other work on flexibility, the displacement is smaller than the in-event response, suggesting some demand creation; however, we caution that there may be further diffuse displacement outside the event study period.
- **During Event Period (e1 to e4):** A large, sharp, and statistically significant spike was present. The Treatment group delivered between approximately 0.20 and 0.32 kW more than the Non-dispatched Control during these four SPs. This difference indicates that the combination of the dispatch signal and the incentive was highly effective in driving kW demand compared to the passive control group.

Figure 12: Differential kW response during turn-up events: Treatment vs. Non-dispatched Control (settlement period-series)

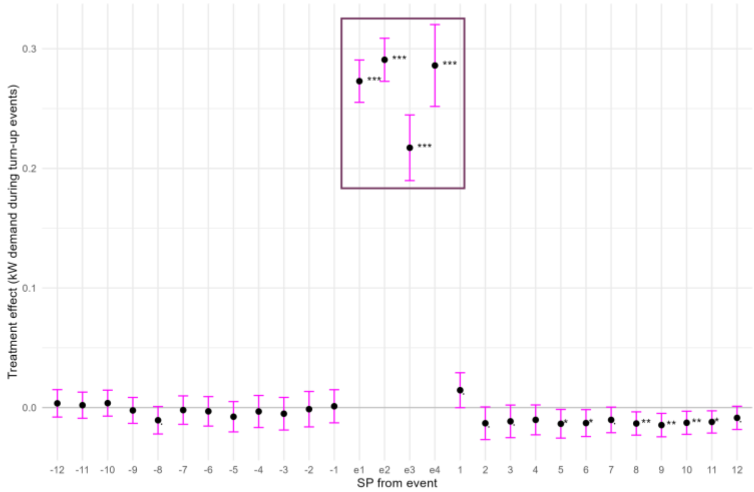
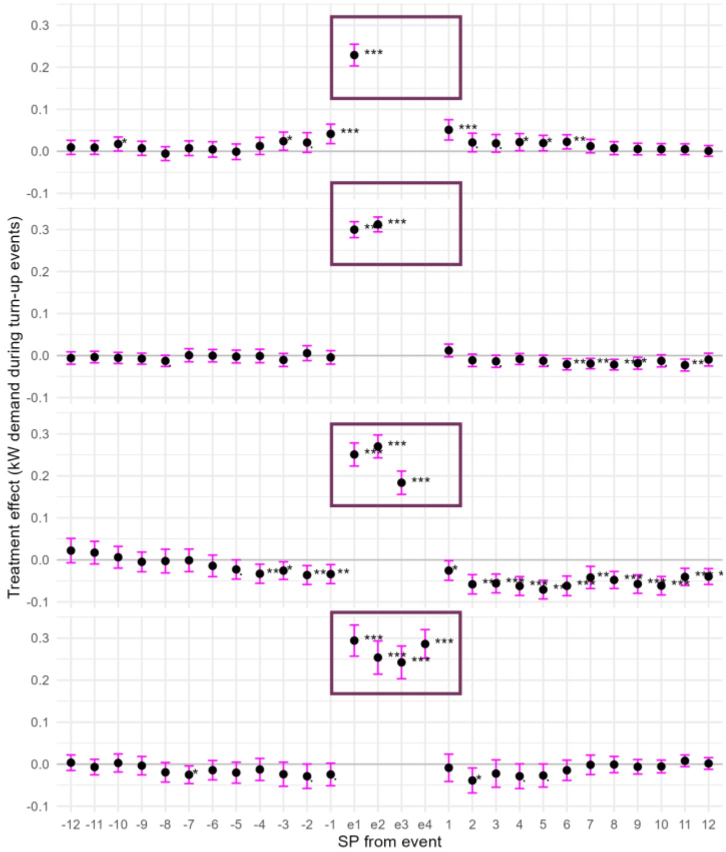


Figure 13: Differential kW response during turn-up events: Treatment vs. Non-dispatched Control (settlement period-series) split by event duration



5.3.2.3 Turn-down events - Treatment vs. Dispatched Control

The event study presents a complex picture. There are a few hypotheses for the pattern we see:

- We see what appears to be a reduction in consumption caused by treatment (and the greater availability it provided OVO). At the same time, the structurally higher consumption caused by the higher plug-in seems to have cancelled out this turn-down. From a grid perspective, insofar as these opposing effects largely offset each other, the result is then a negligible net impact.
- Alternatively, it may be that turn-down events happened at times when OVO was already reducing demand (due to them typically happening during peak periods). Insofar as the extra demand caused by extra plug-in was concentrated in non-peak periods, we might see higher consumption in treatments *around* but not during events. Again, from a grid perspective, the turn-down against the control group is negligible *during* events, as discussed in primary analysis.

Specifically:

- Pre-Event Period (−12 to −1 SPs): There appears to be higher consumption before the event in the Treatment group than the Dispatched Control, especially in the period 6-3 hours before the event. We speculate that this comes from the treatment groups' higher overall consumption caused by higher overall plug-in.

- During Event Period (e1 to e4 SPs): The coefficients mostly hover around the zero line and show some slight negative movement (desired reduction) in e1 and e4, dropping to -0.009 kW and –0.007 kW respectively. However, the confidence intervals for all four event SPs clearly cross the zero line. This signifies that the financial incentive did not add a statistically robust, measurable reduction in kW demand beyond the capability already achieved by simply dispatching the control group. However, as mentioned, this impact may obscure two mutually contradicting forces: real turn-down, on the one hand, against more plug-in and thus more consumption, on the other.
- Post-Event Period (1 to 12 SPs): Coefficients are again positive. As with the pre-event period, we speculate that this comes from the treatment groups' higher overall consumption caused by higher overall plug-in.

Figure 14: Differential kW response during turn-down events: Treatment vs. Dispatched Control (settlement period time-series)

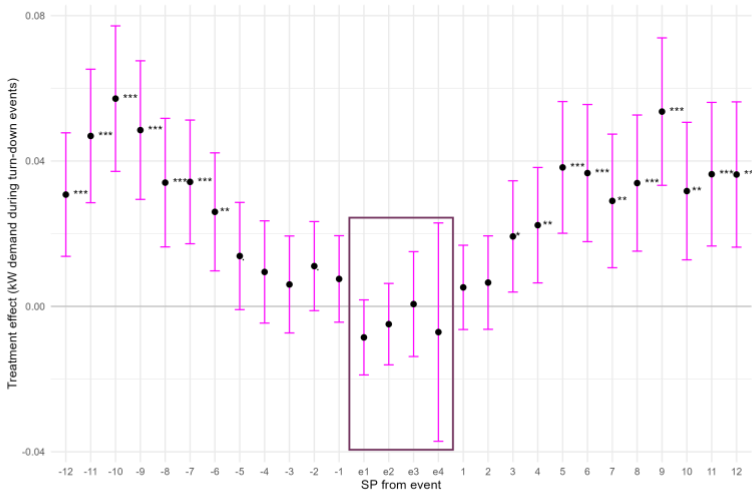
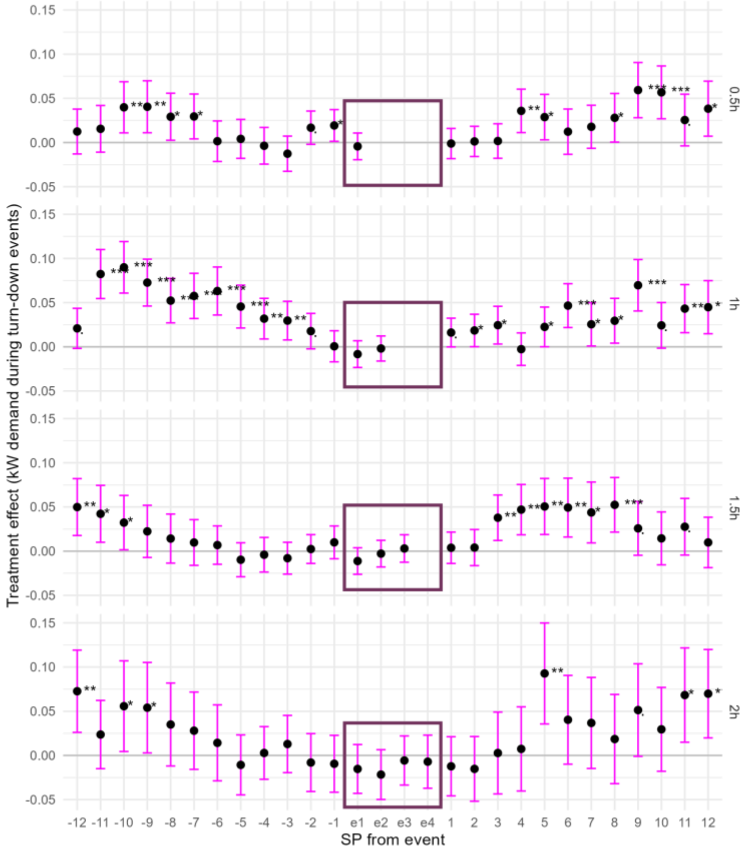


Figure 15: Differential kW response during turn-down events: Treatment vs. Dispatched Control (settlement period time-series) split by duration



5.3.2.4 Turn-down events - Treatment vs. Non-dispatched Control

The comparison against the passive control shows a clear movement towards the desired demand reduction during the event.

- **Pre- and post-event periods:** Coefficients throughout the six hours before (−12 to −1 SPs) and after (1 to 12 SPs) the event are somewhat positive, consistent with higher plug-in and consumption in the Pooled Treatment group than the Non-dispatched Control Group.[4]
- **During event period (e1 to e4 SPs):** A clear negative trend is present, indicating that the Treatment group consumed less energy than the passive control. The reduction was greatest in the final settlement period of events – though note that the greater impacts in the final settlement period may be an artifact of the fact that longer events tended to happen at different times than shorter events.

[4] The Dispatched Control group had slightly lower consumption than the Non-dispatched Control group. We believe this is due to random chance.

Figure 16: Differential kW response during turn-down events: Treatment vs. Non-dispatched Control (settlement period time-series)

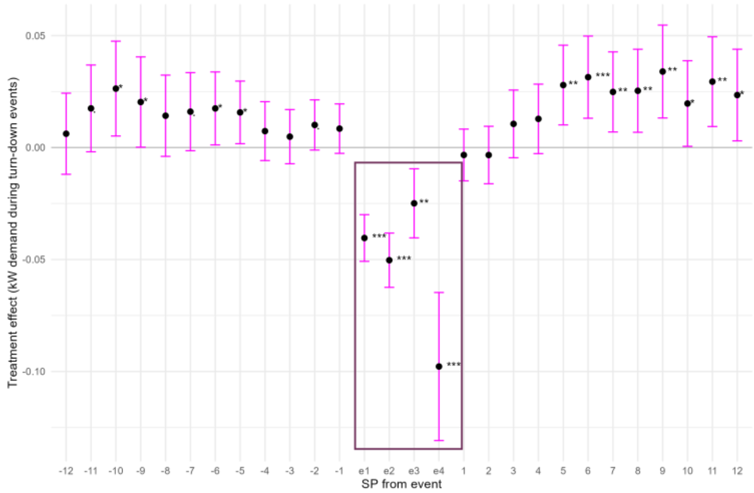
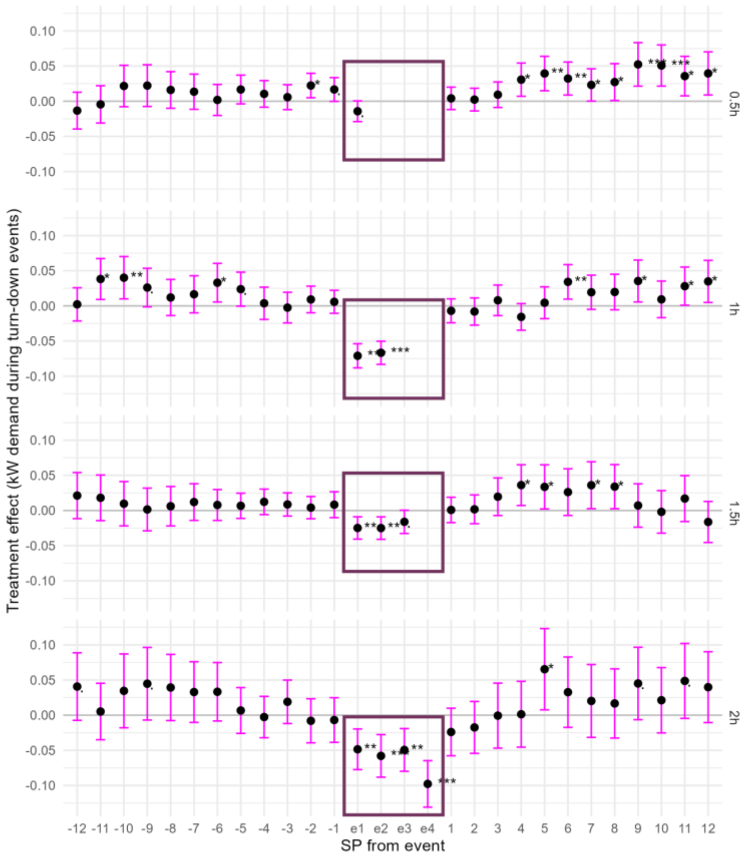


Figure 17: Differential kW response during turn-down events: Treatment vs. Non-dispatched Control (settlement period time-series) split by duration



5.3.3 Portfolio-level P376 consumption shift vs. average treatment effects

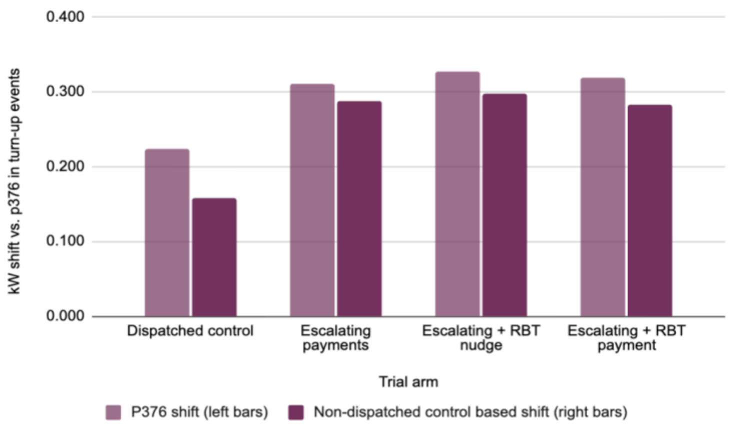
This analysis evaluated the kW shift achieved by each trial arm using two alternative calculations: first, the difference between each arm's demand and that of the Non-dispatched Control; and second, the difference between each arm's demand and its P376 Baseline. P376 is a Balancing and Settlement Code (BSC) modification that defines a baseline derived from meter data against which metered event demand is compared to calculate demand shift. Essentially, DSRSPs compare actual consumption to recent historical consumption in the same settlement periods to estimate shift, especially where DSRSPs are remunerating response per kWh. This trial did not use P376 baselines in this way, as participants were rewarded based on plug-in behaviour rather than consumption shift. However, we have compared actual shift (compared to the matched control group) to the P376-baseline-based shift to inform current and future flexibility programmes that may rely on P376 baselines.

We regard the first calculation as a near ground truth, although its reliability depends on the quality of the matching used to construct the Non-dispatched Control group. The analysis therefore provides a way to assess how well demand shift estimates based on the P376 Baseline perform, at the portfolio level.
So, in summary, the analysis compared, for both Turn-up and Turn-down events:
1. Trial arm average demand minus the trial arm's own average P376 baseline.
2. Trial Arm average demand minus the Non-dispatched Control group's average demand.

5.3.3.1 Turn-up events

The P376 method appeared to slightly overestimate demand turn-up (by 8-13% in the treatment groups, but by 42% in the Dispatched Control). The slight overestimation may be due to the P376 underestimating demand, although the differences may also be due to group imbalances.

Figure 18: Portfolio-level P376 consumption shift vs. average treatment effects

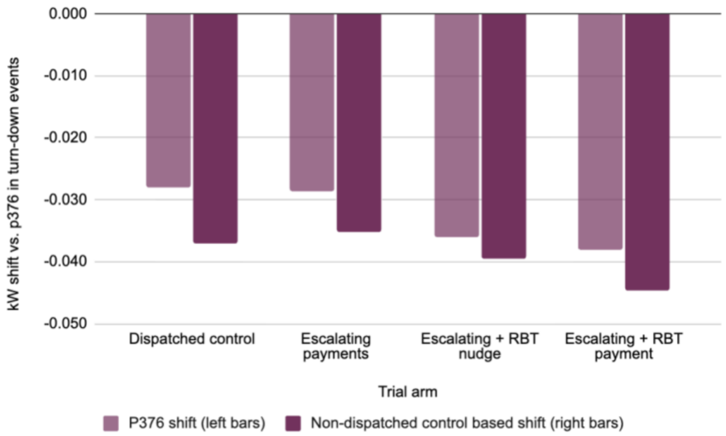


It is not entirely clear why the concordance appears stronger for the treatment groups than for the Dispatched Control. The increased plug-in observed in the treatment groups may have improved the performance of the P376 baseline as a predictor of their expected response. However, again note that the difference may also be due to slight imbalances between Treatment and Control.

5.3.3.2 Turn-down events

For turn-down, the P376-based shift appeared to slightly underestimate the true demand reduction when using the Non-dispatched Control as the counterfactual demand, though – by 9-19% in the treatment groups and 24% in the Dispatched Control. Although the P376-based shift was an underestimate for turn-down but an overestimate for turn-up, these results likely stem from the same underlying reason: the historical P376 baseline may have been underestimating expected demand during the summer event windows. For turn-down, an underestimated baseline leads to an understated negative shift; for turn-up, it leads to an overstated positive shift

Figure 19: Portfolio-level P376 consumption shift vs. average treatment effects



5.3.4 Relationship between plug-in rate and demand shift

A descriptive, event-level graphical analysis was conducted to examine how differences in plug-in behaviour between the treatment and Dispatched Control groups related to differences in demand shift achieved.

The dataset was filtered to include only periods when participants were dispatched, and participants were categorised as either *treatment* (those receiving a payment structure) or *Dispatched Control* (those without a payment structure).

For each event and group, two measures were calculated:

- The average plug-in rate, defined as the proportion of participants connected to a charge point during the event; and
- The average demand shifted by treatment, measured as the mean charging consumption per participant (kW).

These metrics were aggregated to the event × group level. Differences between the treatment and control groups were then computed for each event as follows:

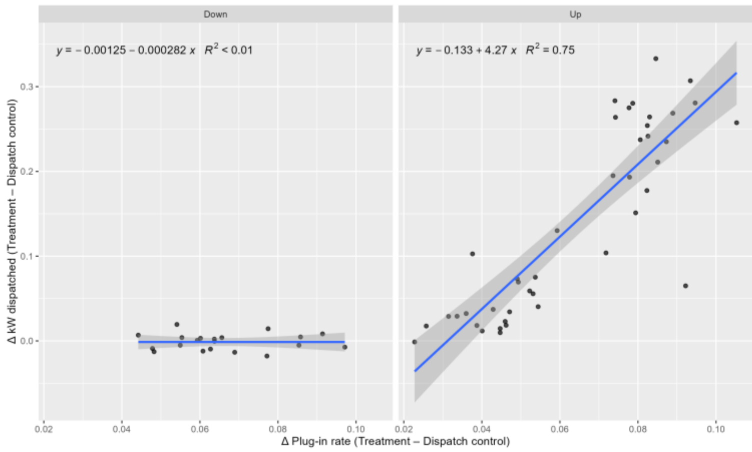
- Δ Plug-in rate = plug-in rate (treatment) – plug-in rate (control)
- Δ kW dispatched = average demand (treatment) – average demand (control)

Each point in the resulting plot represents an individual event, showing how the change in plug-in rate between groups related to the change in dispatched demand. Separate panels were produced for turn-up and turn-down events, with fitted linear trend lines and 95% confidence intervals to illustrate the overall association.

Across turn-up events, a strong positive association was observed between changes in plug-in rates and changes in average demand. Events with higher plug-in rates in the treatment group relative to the control group also recorded higher levels of charging delivered. This relationship indicates that, within the observed data, differences in plug-in behaviour correlated with differences in demand shifted during turn-up events.

Across turn-down events, there was no clear association between changes in plug-in rates and changes in average demand shifted between the treatment and control groups. Variation in the proportion of participants plugged in did not correspond to consistent differences in energy reductions across events.

Figure 20: Relationship between differences in plug-in rate and average demand shifted across events



5.4 Subgroup analyses

In order to examine heterogeneity of treatment effects depending on customer or event type, we undertook subgroup analysis where we examined the extent to which treatment effects varied by factors related to customers or the events themselves. To do so, we compared the Pooled Treatment group (i.e., all three treatment groups pooled together) with the Dispatched Control group. We used the Dispatched Control group as a reference to better isolate the impact of the treatments on behaviour, above and beyond the impact of dispatch.

5.4.1 Subgroup analyses of event kW response

5.4.1.1 Time of day

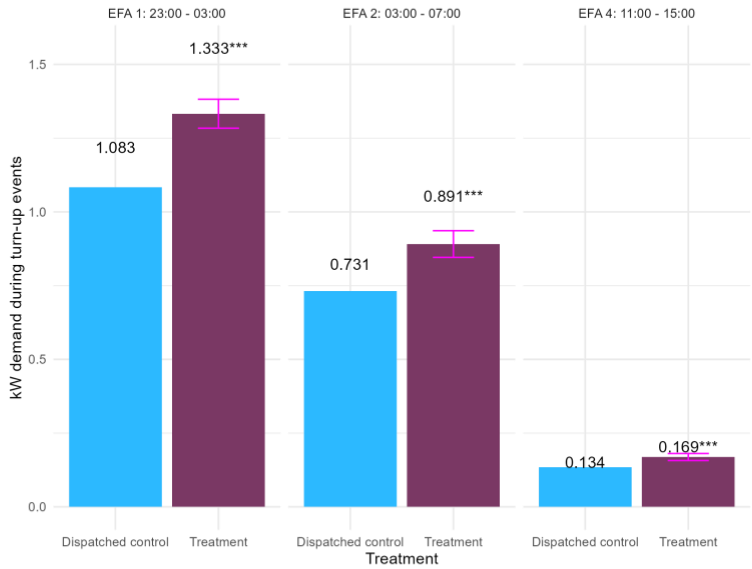
The analysis of average kW demand during flexibility events compared the Dispatched Control group with the Pooled Treatment group across specific EFA time blocks, revealing differential efficacy for turn-up and turn-down events.

Turn-up events
The Pooled Treatment group demonstrated a statistically significant increase in kW demand in all three time blocks in which turn-up events were called, indicating effectiveness in fulfilling turn-up requests.

The Treatment group's average kW demand significantly exceeded that of the Dispatched Control (P<0.001) during:

- EFA 1 (23:00 - 03:00): 1.333 kW vs. 1.083 kW, a 20.35% increase.
- EFA 2 (03:00 - 07:00): 0.891 kW vs. 0.731 kW, a 19.54% increase.
- EFA 4 (11:00 - 15:00): 0.169 kW vs. 0.134 kW, a 23.48% increase.

Figure 21: Average kW demand during turn-up events by time block (EFA)



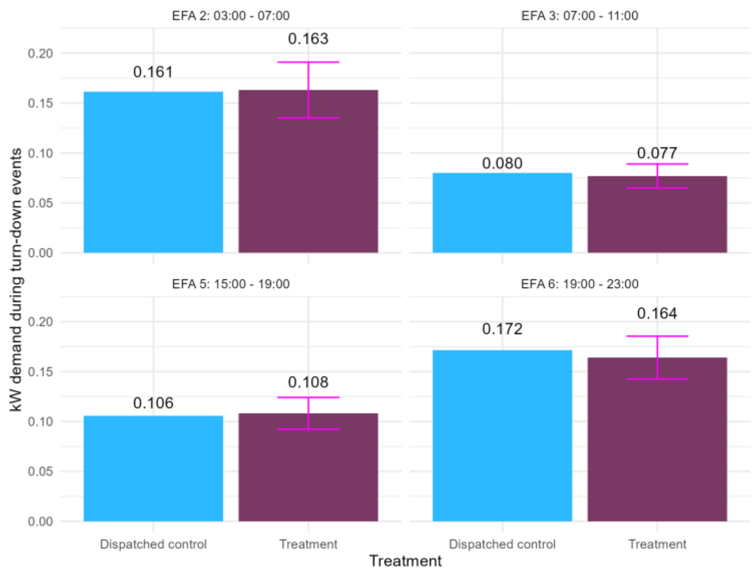
Turn-down events
In contrast, the Pooled Treatment group demonstrated no statistically significant difference in its ability to reduce kW demand during turn-down events when compared to the Dispatched Control group in any of the four time blocks when turn-down events were called:

- EFA 2 (03:00 - 07:00): The response was a non-significant reduction of −2.74% (P=0.757).
- EFA 3 (07:00 - 11:00): The response was a non-significant reduction of −6.04% (P=0.431).
- EFA 5 (15:00 - 19:00): The difference was non-significant at −0.26% (P=0.9730).

- EFA 6 (19:00 - 23:00): The response was a non-significant reduction of −6.71% (P=0.228).

These results showed that the escalating availability payments were successful in improving demand response during turn-up events, but they did not provide a statistically significant improvement in demand reduction during turn-down events beyond the response achieved through automated dispatch itself.

Figure 22: Average kW demand during turn-down events by time block (EFA)



5.4.1.2 Weekday/weekend

The analysis of average kW demand during flexibility events, split by weekday and weekend, compared the Dispatched Control group with the Pooled Treatment group. We found similar dynamics on weekends as on weekdays: a significant increase in kW demand during turn-up events (on both weekends and weekdays) but no statistically significant effect on turn-down events (on either day type).

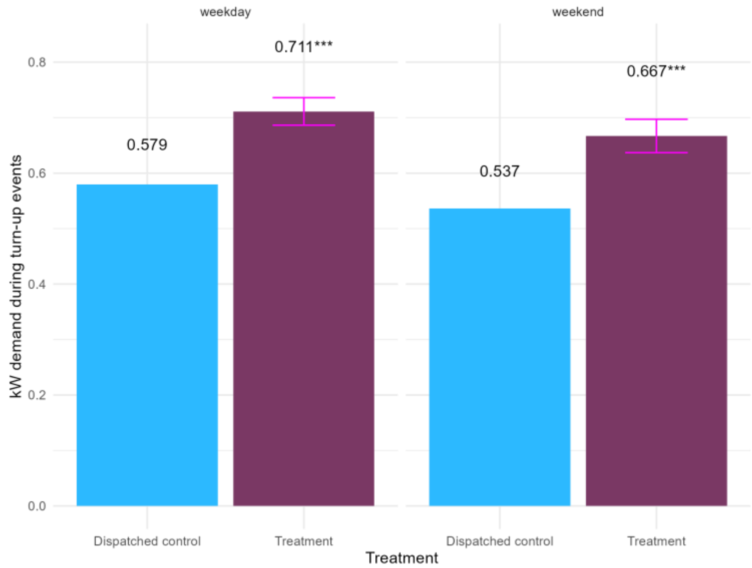
Turn-up events

The Pooled Treatment group resulted in a statistically significant increase in kW demand on both weekdays and weekends (P<0.001):

- Weekdays: The Treatment group delivered 0.711 kW versus 0.579 kW for the Dispatched Control, a 20.10% increase.
- Weekends: The Treatment group delivered 0.667 kW versus 0.537 kW for the Dispatched Control, a 22.01% increase.

In summary, the performance increase was broadly consistent whether on weekdays or weekends.

Figure 23: Average kW demand during turn-up events by day type

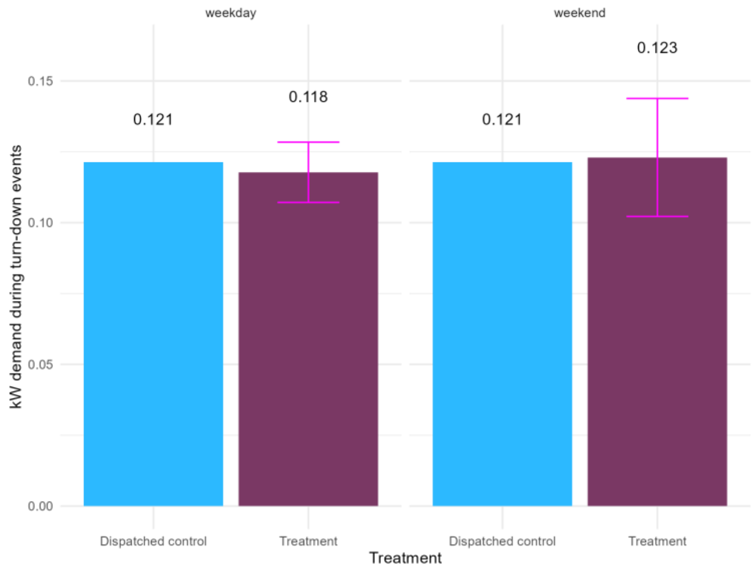


Turn-down events

For turn-down events, the Pooled Treatment group showed no statistically significant difference in kW demand when compared to the Dispatched Control on either day type (P>0.25 for both):

- Weekdays: The Treatment group's demand (0.118 kW) showed a non-significant reduction of −5.14% (P=0.251).
- Weekends: The Treatment group's demand (0.123 kW) showed a non-significant reduction of −1.66% (P=0.850).

Figure 24: Average kW demand during turn-down events by day type



5.4.1.3 EV battery size

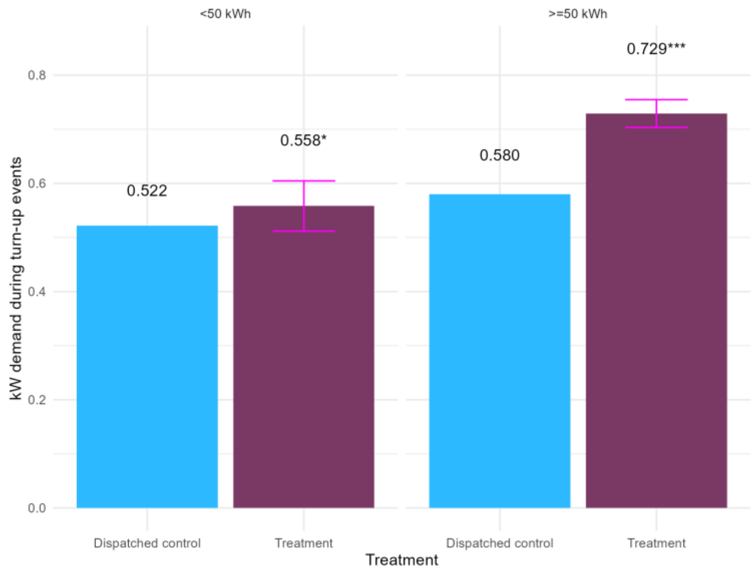
The analysis of average kW demand during flexibility events, split by vehicle battery size (< 50 kWh vs. =>50 kWh), compared the Dispatched Control group with the Pooled Treatment group. We found that the incentive significantly increased kW demand during turn-up events for both segments, but the effect was substantially greater for vehicles with larger batteries. As with other subgroups, no statistically significant effect was found on turn-down events for either battery size segment.

Turn-up events
The pooled treatment group resulted in a statistically significant increase in kW demand for both battery size segments:

- **Batteries < 50 kWh:** The Treatment group delivered 0.558 kW versus 0.522 kW for the control. This 9.08% increase was statistically significant (P=0.0454).
- **Batteries =>50 kWh:** The Treatment group delivered 0.729 kW versus 0.580 kW for the control. This 22.15% increase was significant (P<0.001).

The incentive was more effective at increasing kW demand for vehicles with larger batteries, achieving a percentage increase that was more than double that of the smaller battery segment

Figure 25: Average kW demand during turn-up events by vehicle battery size

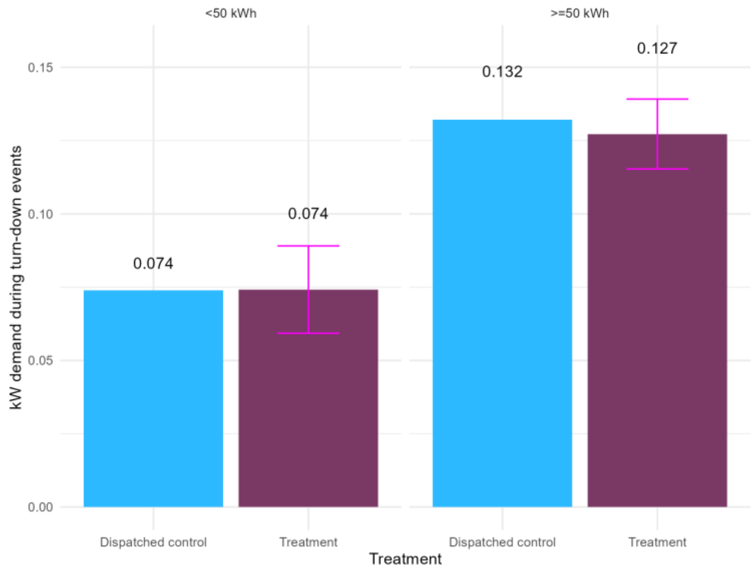


Turn-down events
The pooled treatment group showed no statistically significant effect on demand reduction in either battery segment when compared to the Dispatched Control:

1. **Batteries < 50 kWh:** The Treatment group's demand (0.074 kW) showed a negligible, non-significant change of 1.02% compared to the control (P=0.9208).
2. **Batteries =>50 kWh:** The Treatment group's demand (0.127 kW) showed a small, non-significant reduction of −6.84% compared to the control (P=0.1372).

Although neither turn-down result was significant, the larger battery segment demonstrated a numerically better demand reduction (−6.84%) than the smaller battery segment (1.02%). This minor difference aligns with the turn-up results, suggesting that vehicle battery size might be a facilitating factor for both increasing and decreasing kW response, even when the incremental effect of the incentive is not statistically established.

Figure 26: Average kW demand during turn-down events by vehicle battery size



5.4.1.4 EAC

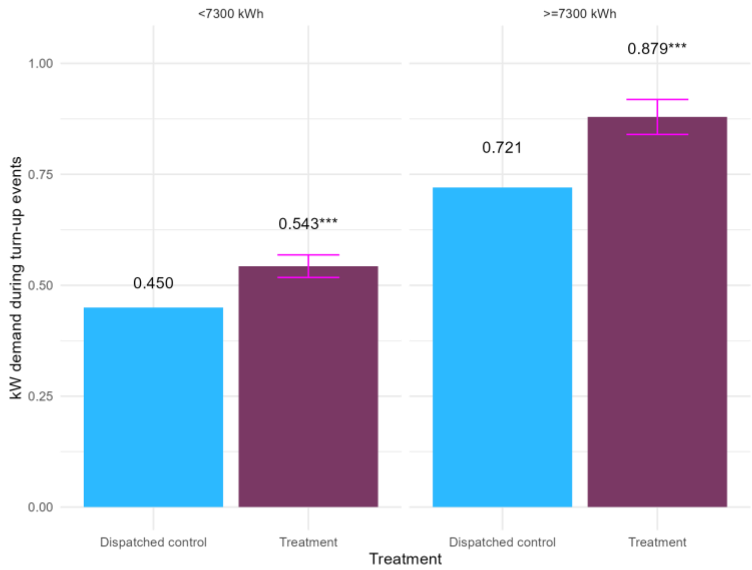
The analysis of average kW demand during flexibility events, split by EAC, compared the Dispatched Control with the Pooled Treatment group, showing similar effects for turn-up but potentially differential effects for turn-down events based on consumption level.

Turn-up events
The Pooled Treatment group delivered a statistically significant increase in kW demand across both household consumption segments (P<0.001), indicating strong performance regardless of baseline energy use:

- Low Consumption (<7300 kWh): The Treatment group delivered 0.543 kW versus 0.450 kW for the Dispatched Control, an 18.30% increase.
- High Consumption (≥7300 kWh): The Treatment group delivered 0.879 kW versus 0.721 kW for the Dispatched Control, a 20.90% increase.

The proportional increase in kW demand was similar in marginally greater in the high-consumption segment.

Figure 27: Average kW demand during turn-up events by household consumption (EAC) segment



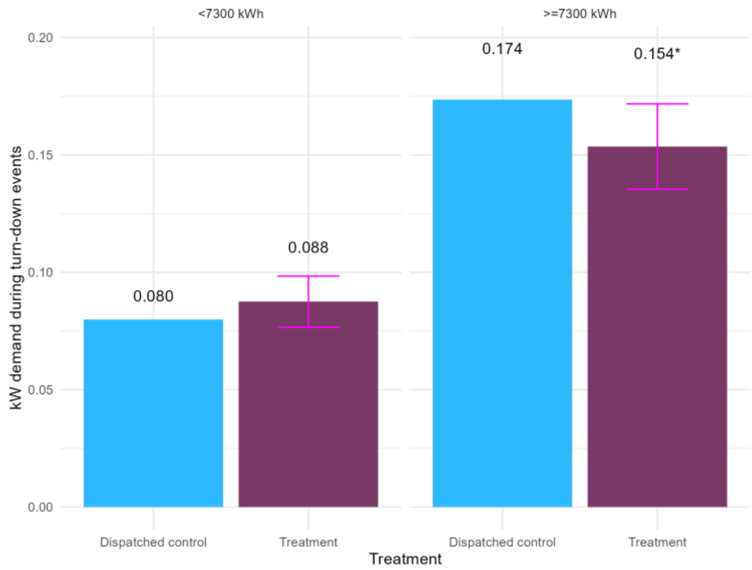
Turn-down events

The effect of the Pooled Treatment on reducing kW demand during turn-down events varied significantly by consumption segment:

- Low Consumption (<7300 kWh): The Treatment group's demand (0.088 kW) showed a non-significant increase of 7.07% compared to the control (0.080 kW) (P=0.3094).
- High Consumption (≥7300 kWh): The Treatment group's demand (0.154 kW) showed a statistically significant reduction of −11.83% compared to the control (0.174 kW) (P=0.027).

Given the general pattern that treatments did not have a significant on turn-down above and beyond the impact of dispatch, we should be cautious about over-interpreting this heterogeneity of turn-down by EAC. We hypothesise, however, that larger-consumption customers may have more charging needs; where they plug in more, they may give OVO an important opportunity to turn-down relative to the Dispatched Control in a way that is not the case for low-EAC customers.

Figure 28: Average kW demand during turn-down events by household consumption (EAC) segment



5.4.1.5 Event duration

The analysis of average kW demand during flexibility events, split by event duration (≤1 hour vs. >1 hour), compared the Dispatched Control with the Pooled Treatment group, again showed similar results regardless of event type: a significant increase in kW demand for turn-up events of both lengths, and no significant effect on turn-down events of either length.

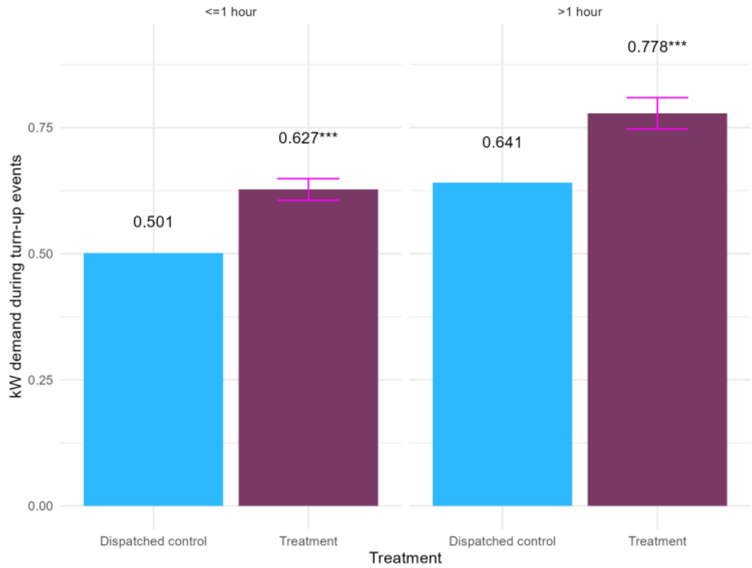
Turn-up events

The Pooled Treatment group achieved a statistically significant increase in kW demand for both short and long events ($P<0.001$):

- Short Events (≤1 hour): The Treatment group delivered 0.627 kW versus 0.501 kW for the Dispatched Control, a 22.04% increase.
- Long Events (>1 hour): The Treatment group delivered 0.778 kW versus 0.641 kW for the Dispatched Control, a 19.27% increase.

The proportional increase in kW demand was slightly greater for shorter events, though the absolute kW delivered per hour was higher during longer events; with that said, the treatment effects were broadly quite similar to each other across event lengths.

Figure 29: Average kW demand during turn-up events by event duration

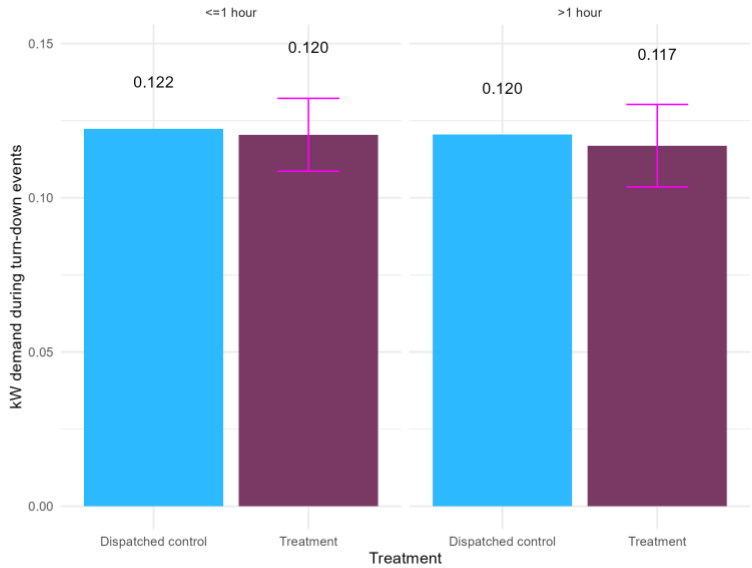


Turn-down events

For turn-down events, the Pooled Treatment group showed no statistically significant difference in kW demand when compared to the Dispatched Control for either duration (P>0.33 for both):

- Short Events (≤1 hour): The Treatment group's demand (0.120 kW) showed a non-significant reduction of −3.75% compared to the control (0.122 kW) (P=0.446).
- Long Events (>1 hour): The Treatment group's demand (0.117 kW) showed a non-significant reduction of −5.43% compared to the control (0.120 kW) (P=0.338).

Figure 30: Average kW demand during turn-down events by event duration



5.4.2 Subgroup analyses - plug-in hours

5.4.2.1 Time of day

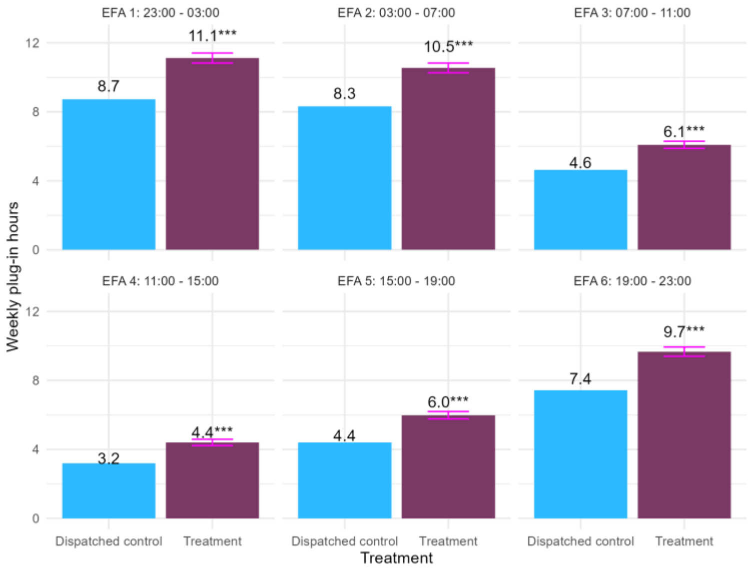
This analysis compared the Dispatched Control group with the Pooled Treatment groups across six fixed time blocks (EFA) to assess how the incentive affected the timing of weekly plug-in hours.

The results show that the Pooled Treatment group recorded a significant increase (P<0.001) in weekly plug-in hours compared to the Dispatched Control in all six time blocks. The effect was consistent across the day:

- EFA 1 (23:00 - 03:00): Treatment hours averaged 11.1 compared to 8.7 for the control.
- EFA 2 (03:00 - 07:00): Treatment hours averaged 10.5 compared to 8.3 for the control.
- EFA 3 (07:00 - 11:00): Treatment hours averaged 6.1 compared to 4.6 for the control.
- EFA 4 (11:00 - 15:00): Treatment hours averaged 4.4 compared to 3.2 for the control.
- EFA 5 (15:00 - 19:00): Treatment hours averaged 6.0 compared to 4.4 for the control.
- EFA 6 (19:00 - 23:00): Treatment hours averaged 9.7 compared to 7.4 for the control.

These findings confirmed that the incentive structure successfully stimulated a statistically significant increase in plug-in behaviour across all measured time segments, including periods that are typically considered on-peak (EFA 3, 5, 6), as well as off-peak (EFA 1, 4).

Figure 31: Treatment effects on weekly plug-in hours by time of day



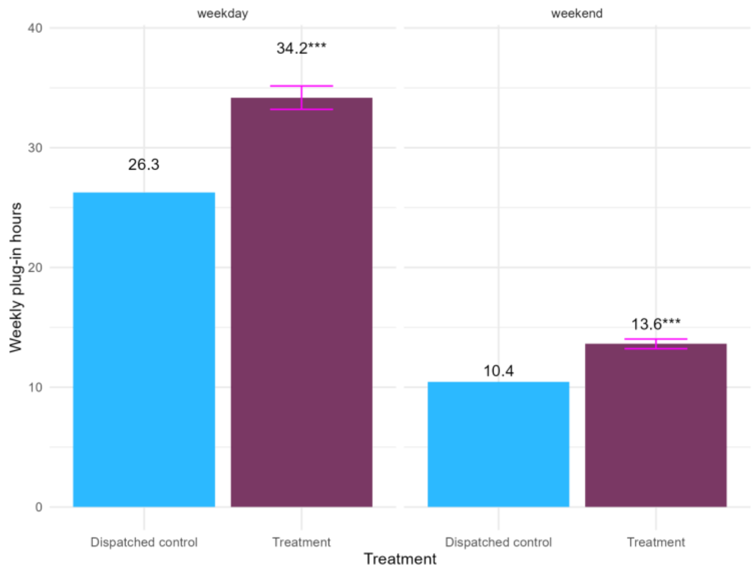
5.4.2.2 Weekly plug-in hours by day type (Weekday vs. Weekend)

The analysis of weekly plug-in hours when split by day type showed that the Pooled Treatment group resulted in a significant increase (P<0.001) in plug-in hours for both weekdays and weekends compared to the Dispatched Control group.

For weekdays, the treatment group averaged 34.2 hours compared to 26.3 hours for the control, representing a 27.32% increase in connected time.
For weekends, the treatment group averaged 13.6 hours compared to 10.4 hours for the control, which is a 28.35% increase in connected time.

These results indicated that the incentive was effective at stimulating a significant and similarly sized proportional increase in weekly plug-in hours on both weekdays and weekends.

Figure 32: Treatment effects on weekly plug-in hours by day type



5.4.2.3 EV battery size

The analysis of weekly plug-in hours split by EV battery capacity showed that the Pooled Treatment group resulted in a significant increase (P<0.001) in plug-in hours for both groups, but with a larger proportional effect on vehicles with larger batteries.[5]

For vehicles with smaller batteries (<50 kWh), the treatment group averaged 66.9 hours compared to 56.4 hours for the Dispatched Control, representing a 19.80% increase in

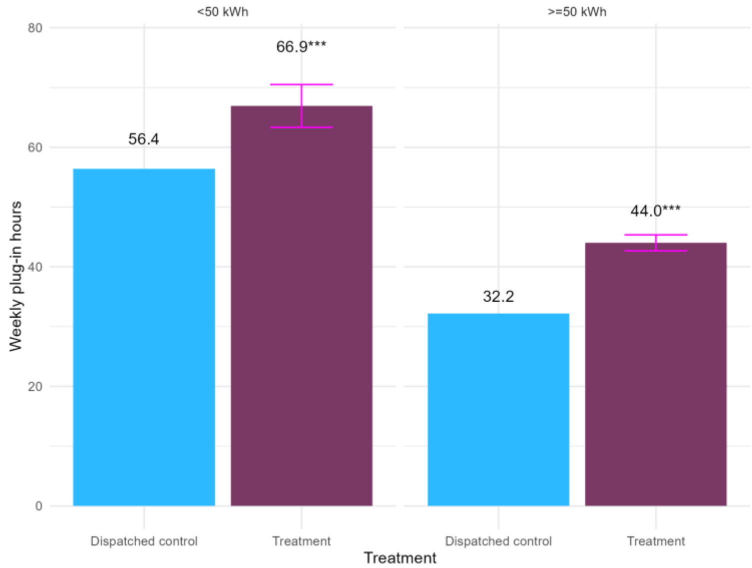
[5] We used 50 kWh as a threshold based on analysis that the EV battery average capacity is "around 50 kWh" in the UK (Etxandi-Santolaya, Casals, and Corchero, 2023), who cite: UK Department for Transport, n.d. Vehicle licensing statistics data files [WWW Document]. GOV.UK. URL https://www.gov.uk/government/statistical-data-sets/vehicle-licensing-statistics-data-files.

connected time.

For vehicles with larger batteries (≥50 kWh), the treatment group averaged 44.0 hours compared to 32.2 hours for the Dispatched Control, representing a 33.26% increase in connected time.

While the absolute difference in hours was similar (around 10-11 hours), the incentive structure was disproportionately effective in promoting plug-in behaviour for vehicles with larger batteries, achieving a percentage increase that was approximately one-third greater than that observed for smaller-battery vehicles.

Figure 33: Treatment effects on weekly plug-in hours by EV battery size



5.4.2.4 EAC

The analysis of weekly plug-in hours split by household estimated annual consumption (EAC)[6] showed that the Pooled Treatment group ('Treatment') resulted in a significant increase (P<0.001) in plug-in hours for both high and relatively lower consumption households, but the effect was proportionally greater for lower-consumption households.

For households with relatively lower consumption (<7300 kWh), the treatment group averaged 43.7 hours compared to 31.6 hours for the Dispatched Control, representing a substantial 36.03% increase in connected time.

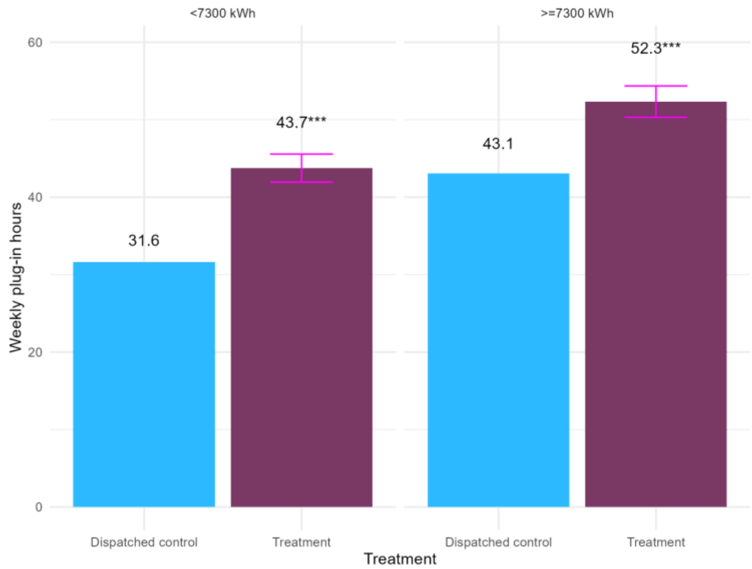
For households with higher consumption (≥7300 kWh), the treatment group averaged 52.3 hours compared to 43.1 hours for the Dispatched Control, representing a 20.10% increase

[6] Energy suppliers use EAC (Estimated Annual Consumption) to estimate each property's expected yearly electricity usage in kWh. This figure is calculated using past meter readings and historical data to predict future consumption. The median EAC in this trial – i.e., among OVO customers with Charge Anytime who had signed up to CrowdFlex for the summer 2025 trial – was 7,300 kWh.

in connected time.

While the higher consumption group maintained a higher absolute plug-in average, the incentive structure was more effective in stimulating plug-in behaviour in the lower consumption group, resulting in a percentage increase that was approximately one-half greater than that observed for higher consumption households.

Figure 34: Treatment effects on weekly plug-in hours by household electricity consumption



5.5 Effect of RBT prompts

In this section, we examine whether the Escalating + RBT Nudge and Escalating + RBT Payment arms led customers to push back their Ready-by-Times (RBTs), as intended. Earlier sections assessed whether these arms produced different flexibility outcomes from the simpler Escalating Payments arm, finding little difference in delivered flexibility. One possible explanation is that customers did not in fact relax their RBTs. However, we find evidence of a moderate but meaningful behavioural response among customers in the Nudge and Payment groups, who did shift their RBTs later.

Customers in these groups were significantly more likely to delay their RBTs than those in the control group. The RBT Payment group achieved the highest overall shift rate at 27.2%, followed by the RBT Nudge group at 23.6%, compared with a background 12.7% rate among participants in the non-RBT treatment group (who still received availability payments but no nudges or payments to change their RBTs).

The likelihood of a successful shift declined as the requested delay increased. For a 1-hour requested shift, success rates were 32% (Reward), 29% (No Reward), and 14% (Control). For 2-hour shifts, the corresponding figures were 20%, 16%, and 10%; for 3-hour shifts, 15%, 13%, and 10%. Across all shift durations, the Reward group consistently achieved the highest success rates, demonstrating the clear motivational effect of a modest financial incentive.

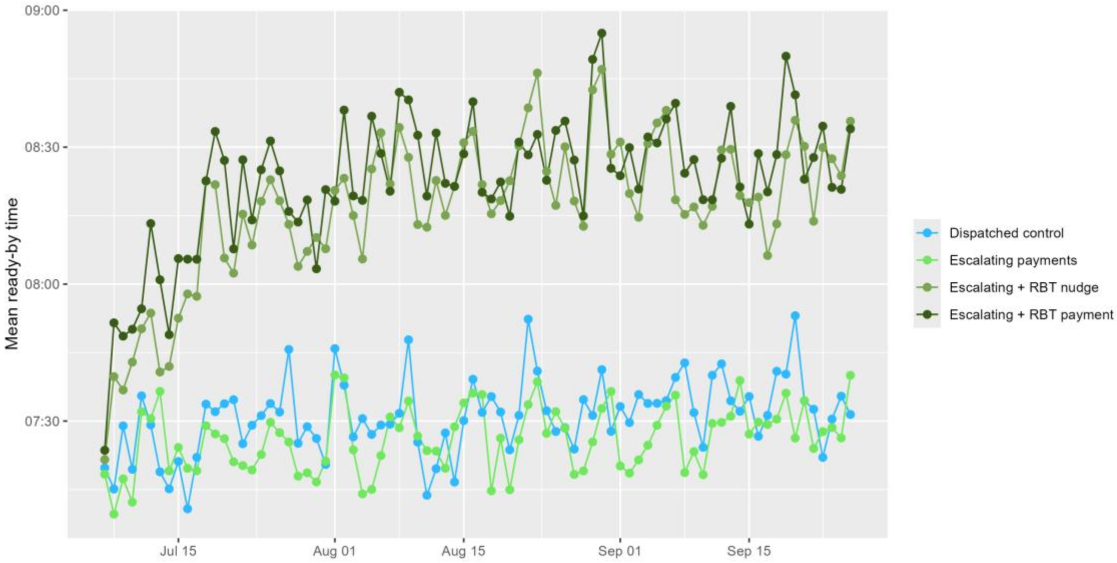
Notably, the average share of participants successfully shifting their RBT (approximately 25% across the Escalating + RBT Nudge and Escalating + RBT Payment groups) was substantially lower than the 72.8% of respondents in OVO's earlier winter "temperature check" survey who reported being willing to delay by at least one hour. This discrepancy likely reflects a gap between stated and revealed preferences.

We next examined the mean daily RBT across trial arms to assess which interventions most effectively achieved this behavioural shift and whether the change persisted over time. The chart below shows the mean RBT by day, treatment group, and Dispatched Control.

The Dispatched Control (light blue line) and Escalating Payments (light green line) groups exhibited similar patterns, setting earlier RBTs that clustered around 07:30–07:45. This confirms that the escalating payment incentive alone, without an explicit RBT prompt, did not materially change ready-by-time behaviour relative to the baseline control.

By contrast, the Escalating + RBT Nudge (dark green line) and Escalating + RBT Payment (darkest green line) groups consistently set later RBTs, typically between 08:15 and 08:30. This represents an average shift of roughly 45–60 minutes, effectively extending the window for flexible load management. However, the addition of the RBT payment did not appear to yield a materially greater shift than the nudge alone.

Figure 35: Daily mean RBT by trial arm - Effect of escalating payments and RBT prompt



5.6 Cost per kWh of flexibility

First, we examined the cost per kWh of flexibility achieved by the treatment groups above and beyond the Dispatched Control. For this definition of cost per kWh, we focused solely on turn-up because the turn-down coefficients were all near zero and statistically insignificant, meaning the treatments achieved virtually no additional turn-down compared to the Dispatched Control. We found similar costs per kWh for the Escalating Payments and Escalating + RBT Nudge groups; however, the Escalating + RBT Payment group was somewhat costlier per kWh.

Table 7: Cost per kWh of additional flexibility by treatment group – versus Dispatched Control group

Trial arm	Total participants (average trial start, end)	Coefficient (turn-up)	Event hours (turn-up)	Total kWh shifted (turn-up)	Total remuneration	£/kWh
Dispatched control	2434		44			
Non-dispatched Control	2386	-0.154				
Escalating payments	2404	0.120		12,686.85	£29,781.00	£2.35
Escalating + RBT nudge	2390	0.120		12,657.54	£30,555.00	£2.41
Escalating + RBT payment	2384	0.114		11,945.91	£35,331.00	£2.96

It is important to note that £/kWh cost metrics are sensitive to trial design features such as event frequency, duration, and timing. As such, care should be taken when making comparisons across trials or projects. For example, the higher event frequency during the winter availability trial likely contributed to a lower equivalent £/kWh figure (around £0.67/kWh). However, this difference does not necessarily reflect a lower procurement cost or greater cost-effectiveness, since the metric is largely driven by how intensively events were scheduled rather than by fundamental differences in the efficiency or value of the flexibility being procured.

In addition, note that in the table above, we have costed the *extra* flexibility from treatment *over and above dispatch*. The £ per kWh compared to the Non-dispatched Control is lower – we have shown it in the table below. This table, then, is the cost of total flexibility, *including the impact of dispatch*. In this way, it looks at the cost of a more expansive definition of flexibility – where we credit the treatment groups for dispatch as well as extra plug-in.

Table 8: Cost per kWh of total flexibility by treatment group – versus Non-dispatched Control group

Trial arm	Total participants (average trial start, end)	Coefficient UP	Event hours UP	Total kWh UP	Total remuneration	£/kWh
Dispatched control	2434		44			
Non-dispatched control	2386	-0.154				
Escalating payments	2404	0.274		29,010.31	£29,781.00	£1.03
Escalating + RBT nudge	2390	0.275		28,885.93	£30,555.00	£1.06
Escalating + RBT payment	2384	0.268		28,133.56	£35,331.00	£1.26

Note that the reduction in demand during turn-down events compared to the *Non-dispatched Control* was statistically significant. When we include those ”turn-down kWh” of flexibility, the cost per kWh decreases further to £0.96, £0.98, and £1.15 £ per total kWh of flexibility.

Another relevant cost metric is the cost per extra hour of plug-in. We found relatively similar costs per hour of plug-in between treatment groups.

Table 9: Cost per hour of extra plug-in by treatment group – versus Dispatched Control group

Trial arm	Total participants (average trial start, end)	Coefficient plug-in hours (per week)	Coefficient plug-in hours (per week) - full sample	Coefficient plug-in hours (total)	Total remuneration	£/hour of extra plug-in
Dispatched control	2434					
Non-dispatched control	2386	1.195	2,851	33,808		
Escalating payments	2404	9.686	23,285	276,095	£29,781.00	£0.11
Escalating + RBT nudge	2390	10.647	25,446	301,721	£30,555.00	£0.10
Escalating + RBT payment	2384	10.555	25,163	298,363	£35,331.00	£0.12

6. Ohme research questions and hypotheses

Below, we set out the core research questions and accompanying hypotheses that underpin Ohme's analysis. These questions were tailored to this trial to explore how participants respond to different incentive structures, and various behavioural prompts.

6.1 Primary research question: interventions' effect on kW demand

1. **Effect of dispatch**: What is the impact of dispatch on kW demand, comparing the Dispatched Control group to the non-dispatched one?
2. **Main impact of payment:** Does the Enhanced Escalating Payments structure lead to a statistically significant difference in kW demand during DSR events compared to the standard structure?
3. **Main impact of prompt:** Does the inclusion of the behavioural signal prompt lead to a statistically significant difference in kW demand during DSR events compared to no prompt?
4. **Combination impacts:** Does the impact of the behavioural prompt on kW demand during DSR events differ depending on the Payment Level received (and vice versa)?

6.1.1 Primary research question: Hypotheses

We hypothesised that:

- **Impact of dispatch:** Customers in the Dispatched Control group will deliver significantly more flexibility (in terms of kW demand) during DSR events than those in the Non-dispatched Control group.
- **Impact of enhanced escalating payments:** Offering a higher incentive (enhanced escalating payments) will lead to additional flexibility delivery compared to the Dispatched Control group, further widening the gap relative to the non-dispatched group.
- **Impact of behavioural prompt:** Providing a behavioural prompt will also increase flexibility delivery relative to the Dispatched Control group.
- **Impact of the combination of enhanced escalating payments and behavioural prompt:** The combined treatment will result in the highest flexibility delivery of all groups, though the increase beyond the Dispatched Control group will be smaller than the sum of the individual effects of enhanced escalating payments and the behavioural prompt.

6.2 Secondary research questions: interventions' effect on plug-in hours per week

1. **Main effect of payment:** Does the Enhanced Escalating Payments structure lead to a statistically significant difference in plug-in hours per week compared to the standard structure?
2. **Main effect of prompt:** Does the inclusion of the behavioural signal prompt lead to a statistically significant difference in plug-in hours per week compared to no prompt?

3. **Interaction effect:** Does the effect of the behavioural prompt on plug-in hours per week differ depending on the Payment Level received (and vice versa)?

6.2.1 Secondary research question: Hypotheses

- **No impact of dispatch:** Customers in the Dispatched Control group will deliver the same level of weekly plug-ins and plugged-in hours as those in the Non-dispatched Control group, on average.
- **Impact of enhanced escalating payments:** Offering a higher incentive (enhanced escalating payments) will lead to higher weekly plug-ins compared to the Dispatched Control group, further widening the gap relative to the non-dispatched group.
- **Impact of behavioural prompt:** Providing a behavioural prompt will also increase weekly plug-ins relative to the Dispatched Control group.
- **Impact of the combination of enhanced escalating payments and behavioural prompt:** The combined treatment will result in the highest number of weekly plug-ins of all groups, though the increase beyond the Dispatched Control group will be smaller than the sum of the individual effects of enhanced escalating payments and the behavioural prompt.

6.3 Exploratory research questions

Exploratory research questions examined treatments' impacts on customers 'dispatchable availability' (state of charge * hours plugged in), weekly charging, and accuracy of P376 baseline; differential treatment effects by subgroup; and the extent to which the nudges/prompts to change app settings were successful in doing so.

7.0 Ohme methodology

7.1 Trial design, eligibility, and interventions

This trial employed a 2x2 factorial randomised controlled design, testing the impact of two variables:

- **Payment level:** Standard Escalating vs. Enhanced Escalating
- **Behavioural prompt presence:** No vs. Yes (non-monetary behavioural signals)

Participants were randomly assigned to one of the four treatment combinations. This design enabled us to estimate both the main effects of each intervention and any interaction effects between payment levels and behavioural prompts.

The primary outcome was the total demand (kWh/hour, expressed in kW) per participant per during DSR events, analysed separately for turn-down and turn-up events. Comparing groups on this outcome provides a measure of demand response per event.

Table 10: Ohme trial arms

Arm Name	Payment Structure	Behavioural Prompt (No Reward)	Dispatch status	Communications
Control 1	N/A	No	FALSE	N/A
Control 2	N/A	No	TRUE	N/A
Treatment group 1	Standard Escalating	No	TRUE	Receives monthly summary emails only
Treatment group 2	Standard Escalating + Prompt	Yes	TRUE	Receives monthly summary emails & weekly behavioural signals/education
Treatment group 3	Enhanced Escalating	No	TRUE	Receives monthly summary emails only
Treatment group 4	Enhanced Escalating + Prompt	Yes	TRUE	Receives monthly summary emails & weekly behavioural signals/education

Table 11: Ohme sample size

Trial arm	# Start of trial	# End of trial	Total drop-out (%)
Non-dispatched Control	2500	2500	0.00%
Dispatched Control	2500	2499	0.04%
Standard Escalating	4480	4142	7.54%
Standard Escalating + Prompt	4995	4596	7.99%
Enhanced Escalating	4511	4161	7.76%

Trial arm	# Start of trial	# End of trial	Total drop-out (%)
Enhanced Escalating + Prompt	4985	4571	8.30%
Total	23971	22469	5.27%

Monthly behavioural prompts: In Treatment Groups 2 and 4, each month focused on testing a specific behavioural prompt designed to influence customer engagement with smart charging behaviours. While the behavioural focus changed monthly, the underlying trial structure and treatment arms remained consistent.

- **July**: *Dynamic Charging* – This goal encouraged users to toggle Dynamic Charging on in the app. This enables chargers to optimise prioritising greenest times and cheapest rates, whilst still meeting customers charging goals. It also requires the user to set a charge target (percentage) and ready-by time; these aspects introduced the key elements for the next two months' nudges.
- **August:** *Ready-by Time* – The focus was on encouraging users to set and extend their 'Ready-by time' to a later time (e.g., from 6:00 AM to 8:00 AM). This gives the charger more flexibility to charge during off-peak, lower-price periods.
- **September**: *Charge percentage* – The goal was to encourage users to set their charge target (the percentage they want their car to reach) to only what they need. The email suggests making sure customers ‘Charge to %’ isn’t set higher than the in-vehicle limits, to help with setting smart schedules.

7.2 Eligibility, recruitment, and randomisation

To participate in the Ohme trial, individuals were required to:

- Own an electric vehicle (EV)
- Have an Ohme charger

These criteria ensured a clear separation between the OVO and Ohme trial populations, while maintaining the technical capability needed for automated smart charging during demand-side response events.

Participants who had signed up for the Summer 2025 trial (both newly recruited and participants who had previously participated in previous Availability CrowdFlex trials) were distributed into the four treatment groups via a random allocation process. The randomisation process allocated the participants by sequentially sampling a specific number of rows for each treatment combination

7.3 Customer matching and control group allocation

Ohme utilised a two-stage methodological approach to ensure a statistically sound trial design: Customer Matching via K-Nearest Neighbours (KNN) sampling and a random treatment group allocation. This process established a robust control group and randomly allocated the recruited customers exclusively to the various treatment cohorts for accurate comparison of outcomes.

To create a valid control group comprising customers who had never been contacted for the CrowdFlex programme, K-Nearest Neighbours (KNN) sampling was implemented, matching the recruited CrowdFlex customers to an un-contacted control base dataset.

7.3.1 KNN matching methodology

The matching relied on comparing users based on two primary, normalised features:

1. The number of distinct hours a user was plugged in.
2. Total energy consumption in kWh.

The KNN implementation matched each treatment user to their closest control user using Euclidean distance, and deduplication ensured each control user was unique. The target control group size was 2,500 users per control group. If the matched sample fell short, the required shortfall was addressed by randomly sampling additional users from the control base.

Once the matched control group was established, the participants were randomly split into two groups: a Non-Dispatch control group and a Dispatch control group.

7.3.2 Control group adjustments

During analyses of the Ohme trial, imbalances between the Control and Treatment groups, were identified, primarily driven by a subset of customers within the original Control group. These methodological issues were addressed through data cleaning and the exclusion of customers with inconsistent or unmatched consumption data. This resulted in the control groups being reduced to 1028 (Non-dispatched Control group) and 1059 (Dispatched Control group). While the resulting revised Control group provides a more balanced basis for comparison, results should be interpreted cautiously as some residual differences may remain.

8. Ohme results

8.1 Primary analysis: Interventions' effect on kW demand during trial events

The primary analysis addressed the core research questions concerning how escalating payment structures and behavioural prompts influenced the amount of flexibility (kWh) delivered during turn-up and turn-down DSR events. The main effects of payment and prompt, their combination, and the baseline impact of automated dispatch were specifically examined.

Note on control group composition and interpretation of results:

- Initial analyses identified some imbalances between the Control and Treatment groups, mainly driven by a subset of customers within the original Control group. These issues were addressed through data cleaning and the exclusion of customers with inconsistent or unmatched consumption data.
- While the revised Control group provides a more balanced and reliable basis for comparison than the original matched Control group, some residual differences may remain. Results should therefore be interpreted with this in mind, and differences between groups should not be taken as strictly causal.
- We also discuss lingering observable differences between Control and Treatment, and even between the randomised treatment groups themselves, in Section 8.4.5 Analysis of Ohme's prompts.

Note that we have also excluded the first 12 events from analysis (those occurring before 24 July, 2025), as Ohme was not yet dispatching the Dispatched Control group during those events.

8.1.1 Effect of treatment on kW demand during trial events

8.1.1.1 Effect of treatment on kW demand during turn-up events

The analysis examined the effect of four individual treatment arms on average kW demand during turn-up grid events, comparing them to the Dispatched Control.

The Non-dispatched Control registered an average of 0.432 kW demand per hour, while the Dispatched control was higher at 0.476 kW (this difference was statistically significant).

All four treatment arms resulted in a significant increase in kW demand during turn-up events when compared to the Non-dispatched Control group.

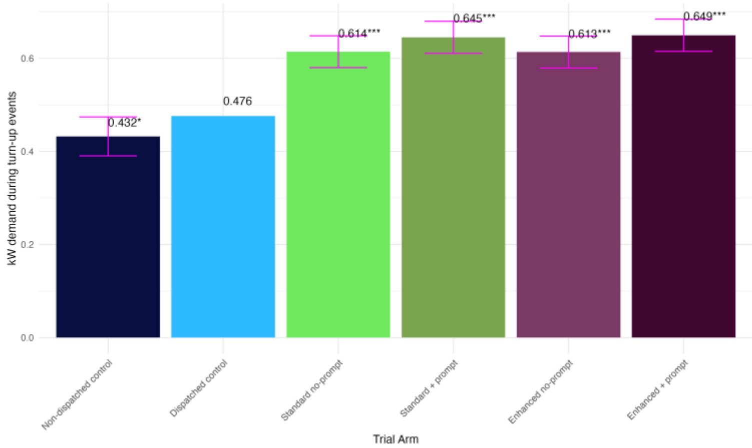
The effects of the treated arms were as follows:

- Enhanced + prompt delivered the highest average at 0.649 kW per hour, representing a 36.41% difference from the Non-dispatched Control. This effect was significant ($P<0.001$).

- Standard + prompt delivered 0.645 kW (35.45% difference), which was also significant (P<0.001).
- Standard no-prompt delivered 0.614 kW, representing a 28.97% increase in demand, a significant result (P<0.001).
- Enhanced no-prompt delivered 0.613 kW (28.84% difference), which was also significant (P<0.001).

The incentives to plug in more – combined with dispatch – increased kW demand relative to both Control group. All four treatment arms successfully increased kW demand to a similar degree, with increases ranging from 28% to 36% compared to the Dispatched Control.

Figure 36: Average kW demand by treatment arm during turn-up events



A series of pairwise comparisons was conducted to test whether differences in incentive strength or messaging influenced average demand during turn-up events.

- Average weekly demand was very similar between the standard (0.553 kW) and enhanced (0.555 kW) incentive groups, with no statistically significant difference (P=0.787), indicating that the enhanced structure did not increase demand relative to the standard version.
- In contrast, participants who received a prompt recorded significantly higher average demand (0.568 kW) than those without a prompt (0.541 kW), a 4.9% increase that was statistically significant (P<0.001).

Table 12: Pairwise treatment comparisons: Impact of prompts and incentive tiers on kW demand during turn-up events

Comparison	Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Standard vs Enhanced	Standard	0.553				
	Enhanced	0.555	0.002	0.007	0.7871	0.33%
No Prompt vs Prompt	No Prompt	0.541				
	Prompt	0.568	0.027	0.007	<0.001***	4.93%

8.1.1.2 Effect of treatment on kW demand during turn-down events

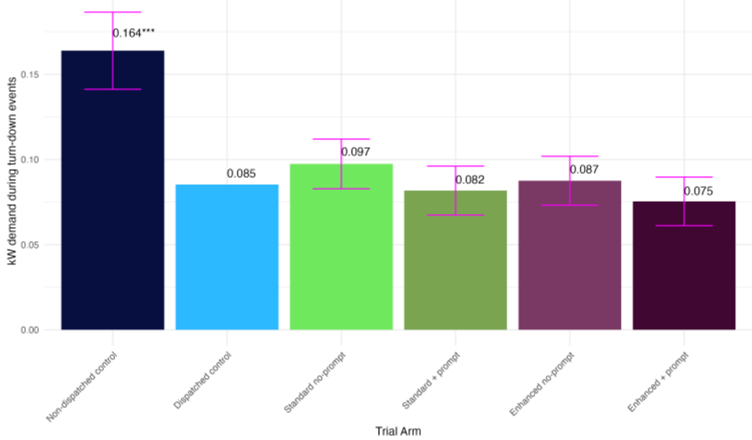
The effects of the treatment groups during turn-down events were varied, and the differences in demand compared to the Dispatched Control (0.085 kW) were not statistically significant (P -values ranging from 0.1050 to 0.7652).

The Non-dispatched Control exhibited the highest average demand at 0.164 kW per hour, which was significantly higher than the Dispatched Control's average (P <0.001), indicating that the act of automating dispatch successfully reduced demand.

Demand for the treatment arms relative to the Dispatched Control indicted no significant differences.

Two of the four treatment groups resulted in a slight reduction in kW demand compared to the Dispatched Control, but none of these reductions achieved statistical significance. In other words, the incentive structures and prompts did not lead to turn-down above and beyond automating dispatch.

Figure 37: Average kW demand by treatment arm during turn-down events



A series of pairwise comparisons was also run for turn-down events.

- Average demand was slightly lower in the enhanced group (0.075 kW) compared with the standard group (0.080 kW), a 6.9% reduction that was statistically significant (*P*=0.035).
- Participants who received a prompt recorded lower average demand (0.072 kW) than those without a prompt (0.083 kW), a 13.9% decrease that was highly significant (*P*<0.001).

In summary, the prompts led to higher demand in turn-up events and lower demand in turn-down events. We believe mechanism for this effect is likely customers setting ready-by times and/or target charge percentages that better enabled Ohme to reduce charging during turn-down events. We discuss the extent to which customers responded to the prompt in Section 8.4.5.

Table 13: Pairwise treatment comparisons: Impact of prompts and incentive tiers on kW demand during turn-down events

Comparison	Trial Arm	10	Coefficient	SE	P-value	% difference
Standard vs Enhanced	Standard	0.080				
	Enhanced	0.075	-0.006	0.003	0.0351*	-6.85%
No Prompt vs Prompt	No Prompt	0.083				
	Prompt	0.072	-0.012	0.003	<0.001***	-13.85%

8.2 Secondary Analysis: Interventions' effect on plug-in hours during the Summer 2025 trial

The secondary analysis assessed how the interventions – payment level and behavioural prompt – influenced plug-in behaviour, namely weekly plug-in sessions and total weekly plug-in hours.

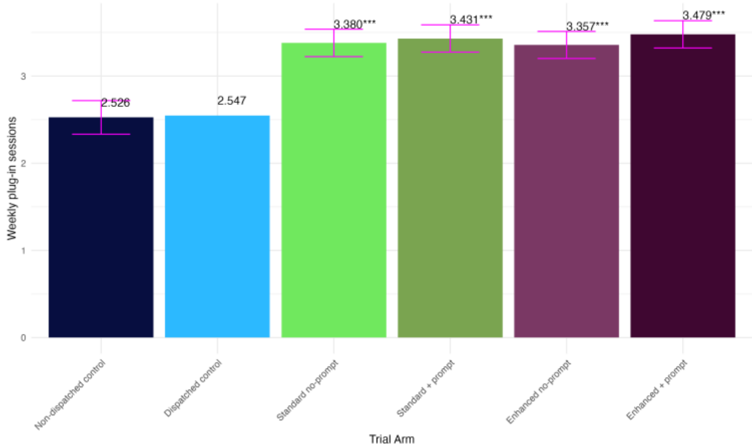
8.2.1 Interventions' effect on plug-in sessions per week

The analysis of average weekly plug-in sessions found that all four treatment arms showed a statistically significant increase in sessions compared to the Dispatched Control (P<0.001 for all).

- The Dispatched Control averaged 2.547 sessions per week.
- The Non-dispatched Control averaged 2.526 sessions, which was a non-significant 0.82% difference (P=0.8322) compared to the Dispatched Control (the lack of difference is what would be expected).
- All four treatment arms achieved significant increases ranging from 32% to 37% compared to the Dispatched Control.

The comparison confirmed that the incentive structures were effective at increasing the frequency of plug-in sessions, but there was no detectable improvement from enhanced (versus standard) payments.

Figure 38: Average weekly plug-in sessions by trial arm



Pairwise comparisons showed no significant difference in average plug-in sessions between the Standard and Enhanced incentive groups (*P*=0.584).

However, customers who received a prompt recorded significantly more plug-in sessions than those without a prompt (+2.1%, *P*=0.045), indicating that prompts modestly increased engagement with plug-in behaviour.

Table 14: Pairwise treatment comparisons - Impact of prompts and incentive tiers on plug-in sessions

Comparison	Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Standard vs Enhanced	Standard	3.402				
	Enhanced	3.421	0.019	0.035	0.5839	0.56%
No Prompt vs Prompt	No Prompt	3.377				
	Prompt	3.447	0.070	0.035	0.0448*	2.07%

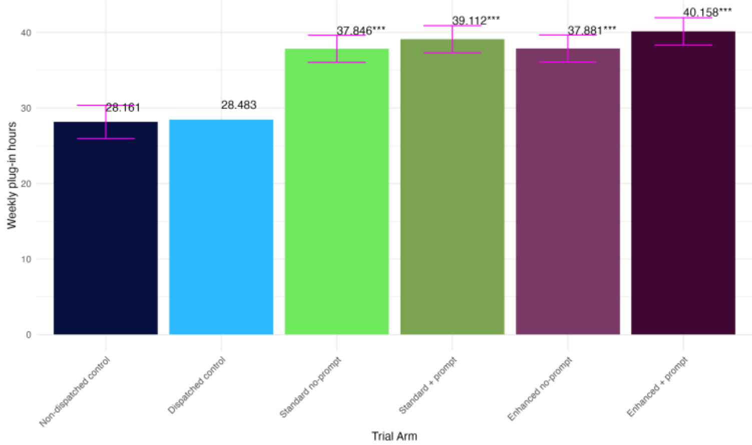
8.2.2 Interventions' effect on plug-hours per week

As with plug-in sessions, the analysis of average weekly plug-in hours found that all four treatment arms showed a statistically significant increase in hours compared to the Dispatched Control (P<0.001 for all).

- The Dispatched Control averaged 28.483 hours per week.
- The Non-dispatched Control averaged 28.161 hours, which was a non-significant -1.13% difference (P=0.7742) compared to the Dispatched Control.
- All four treatment arms achieved increases ranging from 33% to 40.1% compared to the Dispatched Control. The highest average was observed in the Enhanced + prompt arm at 40.158 hours.

The results demonstrated that the incentives were also effective at increasing the total duration of plug-in time – if anything, their effect on plugged in hours was higher than their impact on plug-in sessions per week. This suggests that customers were genuinely plugging in more in response to the treatments, rather than simply trying to achieve a specific number of sessions to achieve a higher availability payment without actually plugging in more hours. However, again, there was no detectable improvement from the enhanced payments.

Figure 39: Average weekly plug-in hours by trial arm



Pairwise comparisons were used to assess whether differences in incentive strength or messaging affected weekly plug-in hours across treatment groups.

Average plug-in duration was similar between the Standard (38.68 hours) and Enhanced (39.17 hours) incentive groups, a 1.3% increase that was not statistically significant (*P*=0.247). By contrast, customers who received a prompt recorded significantly higher plug-in hours than those who did not (+4.3%, *P*<0.001), averaging 39.74 versus 38.11 hours per week.

These findings indicate that behavioural prompts, rather than the level of financial incentive, were more effective in encouraging customers to remain plugged in for longer each week.

Table 15: Pairwise treatment comparisons - Impact of prompts and incentive tiers on plug-in hours

Comparison	Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Standard vs Enhanced	Standard	38.678				
	Enhanced	39.173	0.494	0.427	0.2469	1.28%
No Prompt vs Prompt	No Prompt	38.107				
	Prompt	39.744	1.637	0.427	<0.001***	4.30%

8.2.3.3 Comparison to Turn-up and Turn-down kW impact

The results for weekly plug-in hours reflected the findings from the kW demand analysis concerning the effect of the prompt.

- The Prompt element had a significant impact on both plug-in duration (increasing it by 4.30%) and kW performance (increasing demand during turn-up and decreasing demand during turn-down).
    - In Ohme's case, the prompts in this trial did seem to increase plug-in time. This was also the case in their winter trial. However, the winter trial prompts explicitly reminded customers to plug in more. This trial's prompts asked for changes such as extending RBT and enabling dynamic charging. The fact that plug-in duration increased in both cases suggests that prompts drove greater plug-in – at least for Ohme users – even when the instruction wasn't related to plugging in more.
- Conversely, the Standard versus Enhanced incentive tier comparison showed no statistically significant difference in weekly plug-in hours (P = 0.2469). This broadly aligned with the kW findings, where the difference between Standard and Enhanced incentive tiers was non-significant for turn-up (P = 0.7871) but minor and statistically significant for turn-down (−6.85%, P = 0.0351). We believe the difference in turn-down between the groups may reflect small imbalances between them, rather than being causal, especially given the lack of difference in plug-in.

8.3 Exploratory Analysis

8.3.1 Interventions' effect on state of charge

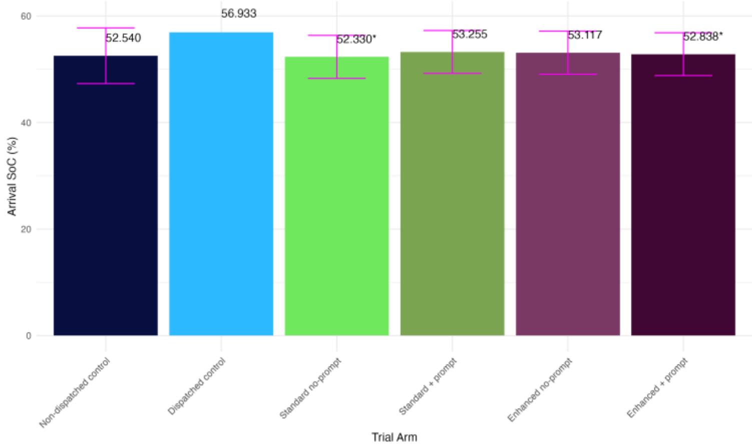
The exploratory research questions examined treatments' impacts on customers' 'dispatchable availability,' calculated using the state of charge (SoC) and hours plugged in with SoC below 80%. In this context, arrival SoC is derived from the SoC of the vehicle at the start of each charging session. This value represents the charge level available in the vehicle's battery when the customer initiates a charging session, providing an indication of their readiness to charge and their potential dispatchable capacity.

8.3.1.1 Interventions' effect on arrival state of charge

We had expected the SoC to be higher in the treatment groups, a sort of cannibalisation of treatment effect where the extra plug-in would eventually cease to be useful due to the vehicle already being highly charged. However, this was not what we found.

- The Dispatched Control group recorded the highest average SoC upon arrival at 56.93%.
- The Non-dispatched Control group arrived with a SoC of 52.549%. This difference of –7.72% compared to the Dispatched Control was not statistically significant (P=0.1411).
- All four treatment arms recorded a lower average SoC compared to the Dispatched Control, ranging between 52-53%. Both the Standard no-prompt and the Enhanced + prompt recorded minor, but statistically significantly lower arrival SoC (-8.10%, P=0.0251; -7.19%, P=0.0459 respectively).

Figure 40: Average weekly arrival SoC by trial arm



Analyses were also run comparing the incentive treatment groups to each other to assess the aggregated effect of the prompt and incentive tiers on average weekly arrival SoC.

There was no statistically significant difference found between the Standard and Enhanced incentive tiers (P=0.8136). The Enhanced group's average SoC was 53.856%, 0.22% higher than the Standard group's 53.743% average. Overall, the type of incentive (Standard versus Enhanced) and the presence of a prompt showed no significant impact on a customer's State of Charge upon plugging in their vehicle for the week. Similarly, there was no statistically significant difference in arrival SoC between the Prompt group and the No Prompt group (P=0.8207). The Prompt group's average SoC was 53.847%, 0.22% higher than the No Prompt group's 53.747% average.

Table 16: Pairwise treatment comparisons - Impact of prompts and incentive tiers on arrival Soc

Comparison	Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Standard vs Enhanced	Standard	53.743				
	Enhanced	53.856	0.125	0.532	0.8136	0.23%
No Prompt vs Prompt	No Prompt	53.747				
	Prompt	53.847	0.121	0.533	0.8207	0.22%

8.3.2 Impact of treatment on weekly consumption

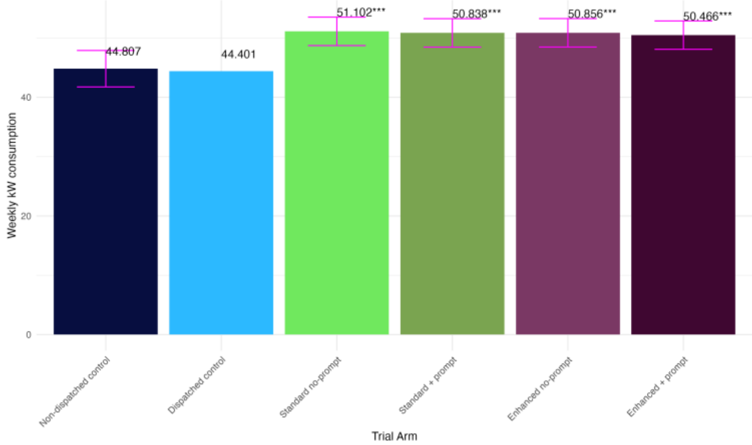
This exploratory analysis assessed the overall impact of the treatment arms on average weekly kWh consumption, comparing all incentive arms to the Dispatched Control.
The analysis found that all four incentive-based treatment arms resulted in a statistically significant increase in weekly kWh consumption compared to the Dispatched Control (P<0.001 for all). No significant differences were observed between the two control groups (P=0.7952)

All four treatment arms showed significant increases in weekly consumption:

- Standard no-prompt: 51.102 kWh, a 15.09% increase.
- Standard + prompt: 50.838 kWh, a 14.50% increase.
- Enhanced no-prompt: 50.856 kWh, a 14.54% increase.
- Enhanced + prompt: 50.466 kWh, a 13.66% increase.

Differences in weekly consumption between the four individual treatment arms were small and not statistically significant.

Figure 41: Treatment effects on average weekly consumption



This analysis mirrored OVO's analysis of availability payments' impact on weekly consumption. Our hypothesis is, similarly, that the treatments led to demand creation at home. The mechanism would be that more time plugged in ends up causing more charging overall, at least at home (perhaps offset by less charging away from home, though there may also be more driving overall).

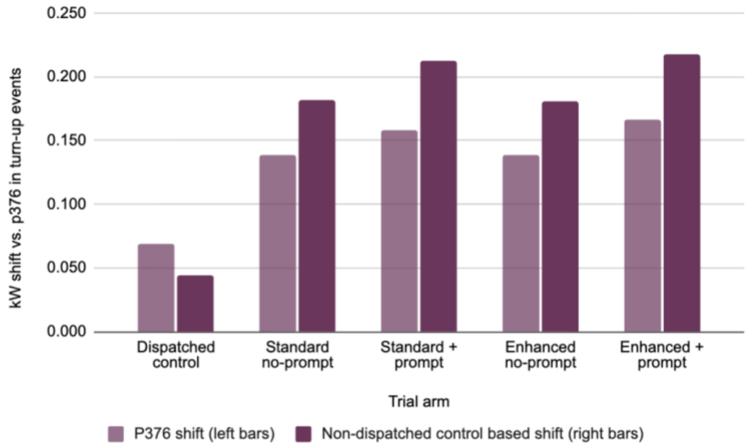
8.3.3 Portfolio-level P376 consumption shift vs. average treatment effects

As with OVO, we compared demand shifts using two reference points: each trial arm's own P376 baseline, and the Non-dispatched Control group's actual demand. The latter *theoretically* provides a near ground truth, but this is subject to the quality of matching. In Ohme's analysis, there are some concerns about how well the control groups were matched to the treatment groups. With this in mind, we advise caution in interpreting these results with strong confidence.

8.3.3.1 Turn-up events

We found that the P376-based calculation overstated demand shifts for the Dispatched Control group but understated them for the treatment groups. This inconsistent picture, combined with the aforementioned concerns about matching, makes it unclear whether the P376 baseline systematically underestimated expected demand (thereby overstating turn-up) or overestimated it (thereby understating turn-up).

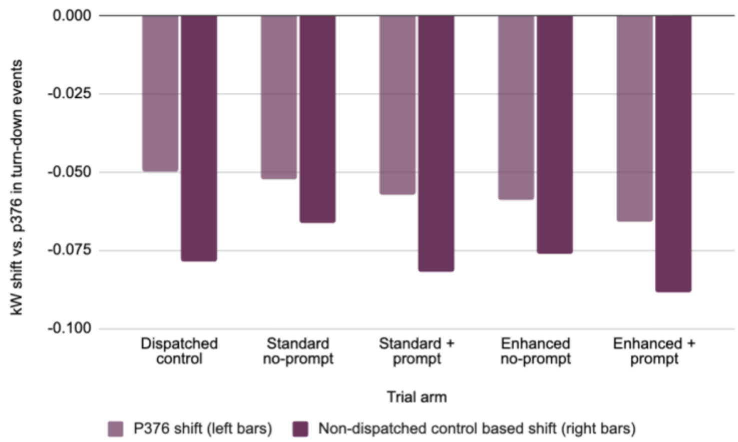
Figure 42: Portfolio-level P376 consumption shift vs. average treatment effects – Turn-up



8.3.3.2 Turn-down events

For turn-down events, the pattern was more consistent across the Dispatched Control and treatment groups: the P376-based estimates appeared to underestimate turn-down, implying that expected demand was overestimated. However, as before, we advise caution in interpreting these findings, given concerns about the quality of the matching between control and treatment groups.

Figure 43: Portfolio-level P376 consumption shift vs. average treatment effects – Turn-down

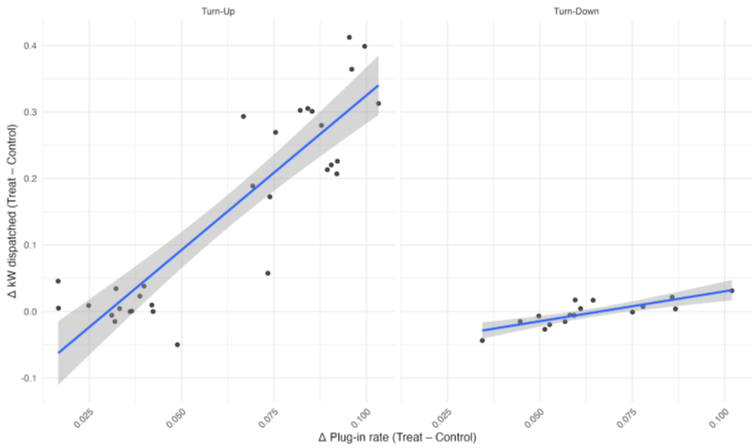


8.3.4 Relationship between differences in plug-in rate and average demand shifted across events

As with OVO, a descriptive, event-level graphical analysis was conducted to examine how differences in plug-in behaviour between the treatment and Dispatched Control groups related to differences in power dispatched.

Similar to the results we saw with OVO, Ohme's data show a clear positive correlation between plug-in rate and kW shift up, during turn-up events, but not down, during turn-down events. This echoes the primary and secondary analysis showing that the treatments increased plug-in, and increased turn-up capacity but not turn-down capacity (above and beyond automated dispatch).

Figure 44: Relationship between differences in plug-in rate and average demand shifted across events



8.3.5 Cost per kWh of flexibility

The cost per kWh of flexibility was higher under the Enhanced Escalating payment structures compared with the Standard Escalating ones. The Standard Escalating groups achieved flexibility between £1.60/kWh - £1.81/kWh, while the Enhanced Escalating group required £2.53/kWh - £2.92/kWh — a roughly 60% increase in cost per unit of flexibility.

Intuitively, as no statistically significant differences were observed in kW demand between the standard and enhanced groups, the higher incentive levels did not yield additional measurable flexibility. This suggests reduced value for money from the enhanced payments relative to the standard tier.

As we did with OVO's results, we focused solely on turn-up because the turn-down coefficients were all near zero and statistically insignificant, meaning the treatments achieved virtually no additional turn-down compared to the Dispatched Control.

Table 17: Cost per kWh of additional flexibility by treatment group - versus Dispatched Control group

Trial arm	Total participants (average trial start, end)	Coefficient UP	Event hours UP	Total kWh UP	Total remuneration	£/kWh
Dispatched control	2500		32			
Non-dispatched control	2500	-0.044	32			
Standard Escalating	4311	0.138	32	19,029.77	£34,410.00	£1.81
Standard Escalating + Prompt	4796	0.169	32	25,911.77	£41,407.00	£1.60

Trial arm	Total participants (average trial start, end)	Coefficient UP	Event hours UP	Total kWh UP	Total remuneration	£/kWh
Enhanced Escalating	4336	0.137	32	19,052.89	£55,555.00	£2.92
Enhanced Escalating + Prompt	4778	0.173	32	26,510.48	£66,987.00	£2.53

As discussed in the cost per kWh calculations for OVO's trial, the £/kWh cost metric is sensitive to trial design features such as event frequency, duration, and timing. Comparisons across trials should therefore be treated with caution, as differences may reflect scheduling intensity rather than true differences in cost-effectiveness.

In addition, note that the above table calculated the incremental flexibility delivered by treatment groups over and above the *Dispatched Control*. Comparing to the *Non-dispatched Control* yielded a lower £/kWh figure, as shown in the table below. This table, then, is the cost of total flexibility, *including the impact of dispatch* (though, as always, note that we assume the Control groups are good counterfactuals for the treatment groups, despite the concerns about matching mentioned previously). In this way, it looks at the cost of a more expansive definition of flexibility – where we credit the treatment groups for dispatch as well as extra plug-in.

Table 18: Cost per kWh of total flexibility by treatment group - versus Non-dispatched Control group

Trial arm	Total participants (average trial start, end)	Coefficient UP	Event hours UP	Total kWh UP	Total remuneration	£/kWh
Dispatched control	2500		32			
Non-dispatched control	2500	-0.044	32			
Standard Escalating	4311	0.182	32	25,096.32	£34,410.00	£1.37
Standard Escalating + Prompt	4796	0.213	32	32,660.83	£41,407.00	£1.27
Enhanced Escalating	4336	0.181	32	25,154.63	£55,555.00	£2.21
Enhanced Escalating + Prompt	4778	0.217	32	33,234.20	£66,987.00	£2.02

When we include the turn-down kWh compared to the Non-dispatched Control, the costs per kWh of flexibility fall further to £1.15, £1.05, £1.80, £1.66.

Finally, in examining cost per extra hour of plug-in, the Standard arms were more cost-effective than the Enhanced arms. This reflects our secondary analysis, showing little difference in plug-in hours between treatments, despite the higher cost of the Enhanced payments.

Table 19: Cost per hour of extra plug-in by treatment group – Vs Dispatched control group

Trial arm	Total participants (average trial start, end)	Coefficient plug-in hours (per week)	Coefficient plug-in hours (per week) - full sample	Coefficient plug-in hours (total)	Total remuneration	£/hour of extra plug-in
Dispatched control	2500					
Non-dispatched control	2500	-0.323				
Standard no-prompt	4311	9.363	40,364	478,600	£34,410.00	£0.07
Standard + prompt	4796	10.629	50,977	604,439	£41,407.00	£0.07
Enhanced no-prompt	4336	9.398	40,750	483,179	£55,555.00	£0.11
Enhanced + prompt	4778	11.674	55,780	661,388	£66,987.00	£0.10

8.4 Subgroup analyses of event kW response

This section examined whether the treatment effects on event kW response varied depending on event characteristics (e.g., time of day, weekday/weekend) or customer characteristics (e.g., EV battery size, EAC). These subgroup results are exploratory. The trial was not specifically designed or powered to detect differences between these subgroups, so findings should be treated with caution. The patterns observed are indicative and intended to inform hypotheses for future studies rather than definitive evidence of subgroup effects.

During turn-up events, several subgroups showed statistically significant increases in average demand, suggesting that the incentive to remain plugged in for longer led to higher availability for dispatch.

During turn-down events, no statistically significant differences were observed across any subgroup, indicating that the treatment had no detectable effect.

Please note that in tables 14 and 15, SE refers to the Standard Error, and that the % difference is the difference between the ‘Coefficient’ and the control average (Avg.).

8.5.1 Battery size

Turn-up events: Average demand during flexibility events was significantly higher for both smaller (≤50 kWh) and larger (>50 kWh) battery groups under treatment (+0.13–0.16 kW, ≈ +32%, $P<0.001$).

Turn-down events: No meaningful differences were detected. Smaller-battery customers showed a small decrease (–0.009 kW, –13%), and larger-battery customers showed a negligible increase (+0.002 kW, +2%).

Interpretation: Incentives to remain plugged in resulted in higher demand during turn-up events for both battery sizes, consistent with greater plug-in availability when dispatch was active. No significant changes occurred during turn-down events, suggesting no systematic difference in baseline demand between groups.

8.5.2 EACC (for Ohme, this is estimated annual consumption of electricity from the charger)

Turn-up events: Average demand increased significantly for lower-consuming households (+0.15 kW, +32%) and rose moderately for higher-consuming households (+0.30 kW, +23%).

Turn-down events: Average demand was largely unchanged for lower-consuming participants (+0.003 kW, +4%) but declined for higher-consuming ones (–0.13 kW, –27%).

Interpretation: Higher-consuming customers showed larger treatment effects in absolute terms, but smaller effects relative to their baseline demand.

8.5.3 Time of day

Turn-up events: Statistically significant increases in demand were observed in early flexibility windows — EFA1 (+0.28 kW, +34%) and EFA2 (+0.24 kW, +36%). Later periods showed no significant effects (EFA4: ns).

Turn-down events: Differences were minimal and not significant across all windows (within ±0.01 kW, ns).

Interpretation: Significant increases in average demand during early-day turn-up events suggest the incentive successfully increased plug-in availability when dispatch was active. The lack of significant differences in turn-down periods likely reflects the broader pattern of minimal turn-down among Ohme Treatment groups compared to the Dispatched Control.

8.5.4 Day type

Turn-up events: Average demand was significantly higher for both weekdays (+0.15 kW, +32%) and weekends (+0.16 kW, +35%).

Turn-down events: No significant differences were found (Weekday: –0.001 kW, ns; Weekend: +0.006 kW, ns).

Interpretation: Incentives to stay plugged in led to consistent and statistically significant increases in event-period demand across both weekdays and weekends during turn-up events, reflecting more availability for automated dispatch. No significant changes occurred during turn-down events.

Table 20: Average demand (kW) during turn-up events by subgroup

Subgroup	Control avg.	Treatment avg.	Coefficient	% difference	kW difference
Battery ≤50kWh	0.385	0.51	0.125***	32.47%	0.125
Battery >50kWh	0.505	0.664	0.159***	31.49%	0.159
EACC ≤7300	0.462	0.608	0.146***	31.60%	0.146
EACC >7300	1.301	1.597	0.296	22.75%	0.296
EFA1	0.812	1.086	0.274***	33.75%	0.275
EFA2	0.567	0.773	0.206***	36.31%	0.206
EFA4	0.092	0.098	0.007	7.62%	0.007
Weekday	0.478	0.629	0.152***	31.80%	0.152
Weekend	0.471	0.636	0.164***	34.82%	0.164

Table 21: Average demand (kW) during turn-down events by subgroup

Subgroup	Control avg.	Treatment avg.	Coefficient	co-efficient vs	kW difference
Battery ≤50kWh	0.071	0.061	-0.009	-12.68%	-0.009
Battery >50kWh	0.09	0.092	0.002	2.22%	0.002
EACC ≤7300	0.076	0.079	0.003	3.95%	0.003
EACC >7300	0.492	0.359	-0.133	-27.03%	-0.133
EFA2	0.082	0.101	0.019	23.39%	0.019
EFA3	0.053	0.047	-0.006	-11.20%	-0.006
EFA5	0.076	0.072	-0.004	-5.60%	-0.004
EFA6	0.126	0.128	0.002	1.93%	0.002
Weekday	0.088	0.087	-0.001	-1.14%	-0.001
Weekend	0.074	0.08	0.006	8.11%	0.006

8.4.5 Analysis of Ohme's prompts

As noted previously, the subject of Ohme's prompts changed each month. In July, Ohme encouraged users to turn Dynamic Charging "on" in their app; in August, to set and extend their 'Ready-by time' (RBT) to a later time (e.g., from 6:00 AM to 8:00 AM); and in

September, to set their charge target (the percentage they want their car to reach) to only what they need.

In the following graphs, we show rates of customer behaviour on the y-axis, first split by trial group, then focusing on treatment groups only and comparing the two non-prompted treatment groups (pooled) versus the two prompted treatment groups (pooled). In the first two cases, the desired effect is a higher outcome on the y-axis: *higher* rate of dynamic charging "on", and a *later* average RBT. In the case of charge target, a *lower* average charge target is desirable from the perspective of maximising Ohme's ability to charge flexibly.

For dynamic charging (July's encouragement target), there appears to be a moderate but notable effect of the nudge in causing more customers to turn "on" dynamic charging.

Figure 45: Dynamic charging "on" rate, by trial group

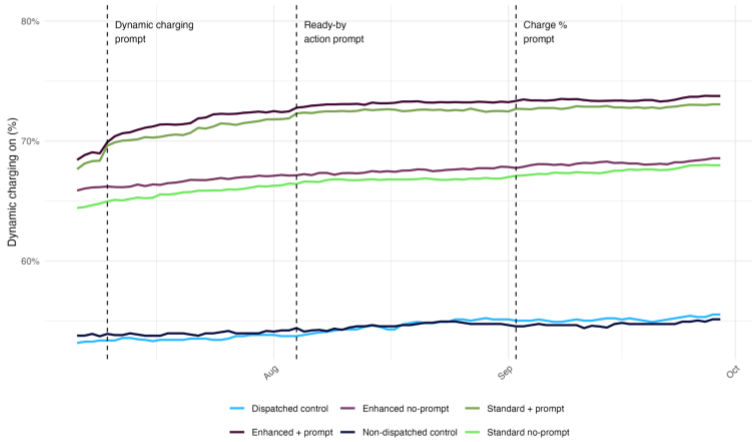
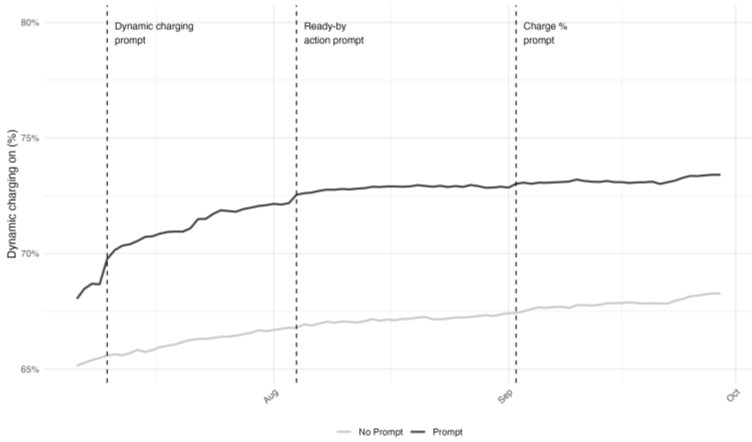


Figure 46: Dynamic charging "on" rate, by Prompt vs No Prompt (among treatment customers)



The impact of the RBT nudge in August is murkier. We do not see a clear impact in August, but there is an uptick in the RBT time in September. We are not certain why we see this pattern.

Figure 47: RBT time, by trial group

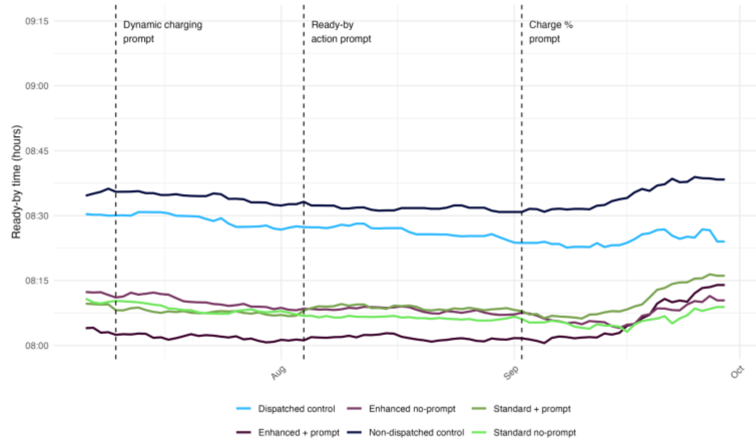
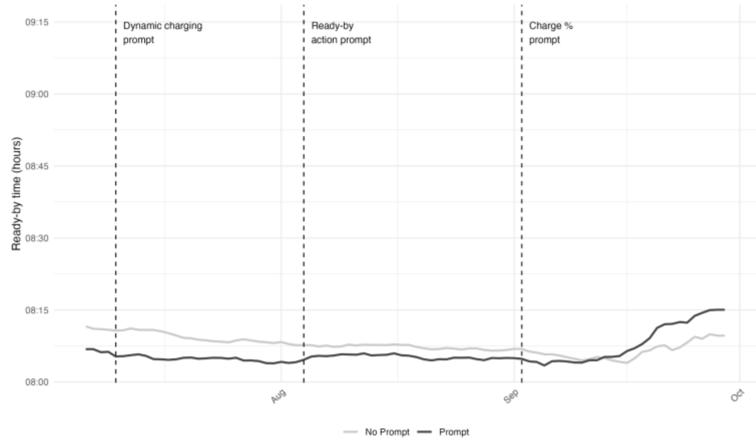


Figure 48: RBT time, by Prompt vs No Prompt (among treatment customers)



For the charging target, we see a small effect in both July – we hypothesise that this is a positive spillover from customers putting dynamic charging "on" – and again in September, when nudged specifically regarding the charge target.

Figure 49: Charging target, by trial group

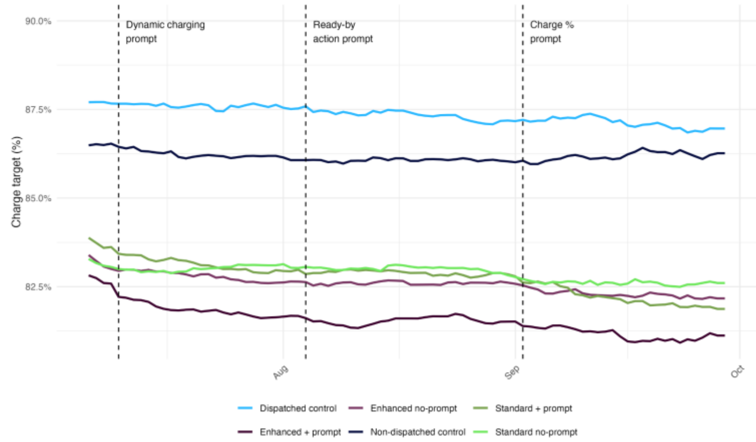
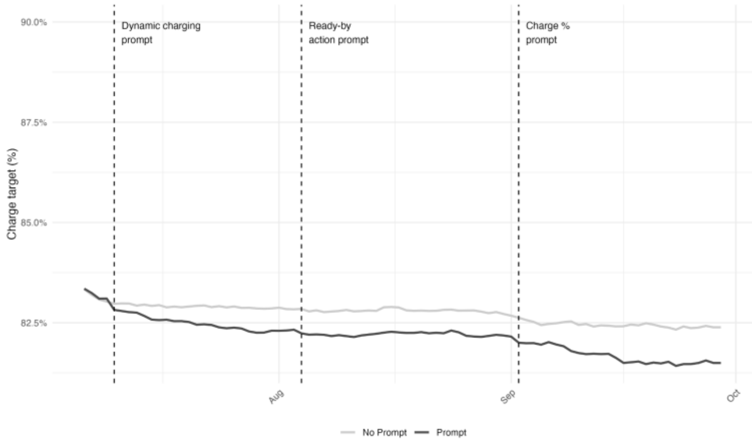


Figure 50: Charging target, by Prompt vs No Prompt (among treatment customers)



Finally, it is notable that these graphs suggest some concerning imbalances, both between Control and Treatment and between the treatment groups when compared to each other. First, comparing Control and Treatment:

- The Control groups have lower dynamic charging turned “on” at the beginning of the trial, before any prompts.
- The Control groups have later RBTs.
- The Control groups have higher charge targets.

These imbalances indicate that even after matching on observable variables such as consumption, there are imbalances on observable variables that were not part of the matching criteria, suggesting further unobservable differences between the groups. These differences mean that Treatment-Control comparisons should be treated with caution.

Among treatment groups, we also see that the Prompted treatment groups have higher rates of Dynamic Charging “on” and may have earlier RBTs. These imbalances may come from unintentional oversampling of customers with these characteristics in the randomisation. These differences mean that even treatment versus treatment comparisons may be affected by underlying differences between the groups.

9. Discussion

The summer 2025 availability trials provided further robust evidence that EVs deliver significant flexibility for turn-up and turn-down events, aligned with system balancing needs. In doing so, it provided robust evidence for summer flexibility potential from EVs, an important evidence gap. Across both OVO and Ohme trials, automated dispatch of EVs during events produced large, statistically significant shifts in consumption. Compared to the Dispatched Control, the Non-Dispatched Control had 9% (Ohme) and 27% lower consumption during turn-up events; and 92% (Ohme) and 31% (OVO) higher consumption during turn-down events.

In addition to automated dispatch, the trials suggested that availability payments can meaningfully shape EV plug-in behaviour to provide greater opportunity for demand flexibility. All incentive structures substantially increased plug-in sessions and hours, confirming that these can reliably encourage customers to make their vehicles available for flexibility services. However, how those incentives were structured appeared to make comparatively little difference, and more generous offers made the incentives less cost-effective. Put another way, we found very little additional benefit from higher availability payments or remunerated nudges/prompts to relax customer preferences that could theoretically enhance the impact of automation of EV charging. The standard availability payments trial arm had similar turn-up and turn-down to the other arms, in both Ohme and OVO's trials.

The increased plug-in behaviour from availability payments translated into greater turn-up flexibility during events over and above automated dispatch, but not into equivalent gains for turn-down events. For OVO, availability payments increased turn-up by roughly 20% relative to the Dispatched Control group, while these did little to increase turn-down effect sizes, which were 0-8% compared to the Dispatched Control. For Ohme, turn-up rose by 29-36% relative to the Dispatched Control group, while again turn-down was generally unaffected, with demand 3-14% lower than the Dispatched Control. It is worth noting that OVO and Ohme employ different dispatch mechanisms; in addition, some of this variation may be driven by the differing extent to which each provider accesses implicit flexibility.

We hypothesise this is because extra plug-in more directly aided automated turn-up than turn-down. For turn-up, having more vehicles plugged in during the 0.5- to 2-hour window when the DSRSP sought to increase charging demand directly increases the potential kWh delivered. In contrast, the way that extra plug-in could have aided turn-down – by providing the DSRSP with more alternative charging hours – is more indirect, and therefore perhaps weaker. It is also notable that turn-down events tended to occur at times of day when plug-in rates were lower, compared to turn-up windows, limiting the available capacity for further reduction. The result is that while availability payments enhanced *potential* for turn-down beyond the impact of dispatch, we did not detect that they realised that value. Finally, a good deal of turn-down flexibility may already be captured by standard optimisation, which generally avoids high-cost peak periods. This leaves little remaining load to turn-down during an event. In contrast, signals to turn-up during the day are distinct from standard charging patterns, allowing the event to deliver a clear, additive impact.

Availability payments may offer benefits that extend beyond the narrow event windows measured here, given higher plug-in rates increase the total connected hours of chargers, expanding the pool of assets that flexibility providers can draw on for both upward and downward services. These benefits may become more visible in longer or more frequent event regimes, where extended availability can be fully utilised. Future trials could therefore explore longer or more frequent events to test whether sustained availability translates into greater flexibility delivered.

Exploratory analysis also revealed some notable variation in how different customers responded to the incentives. Incentives appeared more effective for vehicles with larger batteries, which delivered proportionally higher increases in kW demand during turn-up events – reflecting both greater storage capacity and the ability to sustain higher charge rates. However, lower-consuming households (with annual electricity consumption below 7,300 kWh) showed larger percentage increases in plug-in hours, suggesting a stronger behavioural response from these customers. While this analysis is non-causal, it may imply different responses based on vehicle type or household profile, for providers to consider in optimising participation. While this analysis was non-causal, the ability to observe and describe these patterns was enabled by the randomised design and large scale of participation, which together provide unusually rich evidence on how responses may vary across customer and vehicle characteristics.

In summary, automated dispatch of EVs delivered flexibility, and incentives meaningfully shaped when customers plug in, but additional flexibility delivered depended on times that events are called and was more substantial for turn-up than turn-down events. Integrating availability payments into mainstream commercial flexibility products – aligned with event timing, vehicle mix, and consumer characteristics – could unlock substantial additional flexibility at low cost. However, more intricate or generous incentive structures beyond straightforward, predictable payments to plug in more often may bring little additional value. Overall, availability payments can provide significant demand flexibility to support a renewables-based grid; this may increase over time as automation improves and bidirectional charging enables exporting to the grid in addition to load-shifting.

10. Appendix

10.1 OVO results tables

Please note that SE refers to the Standard Error, and that the % difference is the difference between the 'Coefficient' and the Dispatched Control average (Avg.).

10.1.1 Primary analysis supporting table

Supporting tables for section 5.1.1: Interventions' effects on turn-up events.

Table 22: Average treatment effects of interventions during turn-up events

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	0.569				
Non-dispatched Control	0.412	-0.154	0.012	<0.001***	-27.11%
Escalating payments	0.699	0.120	0.015	<0.001***	21.07%
Escalating + RBT nudge	0.709	0.120	0.015	<0.001***	21.14%
Escalating + RBT payment	0.695	0.114	0.015	<0.001***	20.00%

Supporting tables for section 5.1.2: Interventions' effects on turn-down events.

Table 23: Average treatment effects of interventions during turn-down events

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	0.121				
Non-dispatched Control	0.158	0.038	0.007	<0.001***	31.46%
Escalating payments	0.123	0.000	0.006	0.9997	0.00%
Escalating + RBT nudge	0.119	-0.007	0.006	0.2745	-5.59%
Escalating + RBT payment	0.114	-0.010	0.006	0.0991	-8.37%

10.1.2 Secondary analysis supporting tables

Supporting tables for section 5.2.1: Interventions effects on plug-in sessions.

Table 24: Average treatment effects of interventions on plug-in sessions

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	2.25				
Non-dispatched Control	2.33	0.084	0.044	0.0582	3.74%
Escalating payments	2.81	0.522	0.047	<0.001***	23.22%
Escalating + RBT nudge	2.88	0.567	0.047	<0.001***	25.21%
Escalating + RBT payment	2.86	0.567	0.047	<0.001***	25.22%

Supporting tables for section 5.2.2: Interventions' effects on plug-in hours.

Table 25: Average treatment effects of interventions on plug-in hours

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	36.67				
Non-dispatched Control	37.80	1.195	0.793	0.1318	3.26%
Escalating payments	47.01	9.686	0.874	<0.001***	26.42%
Escalating + RBT nudge	48.35	10.647	0.879	<0.001***	29.04%
Escalating + RBT payment	47.97	10.555	0.874	<0.001***	28.79%

10.1.3 Exploratory analysis supporting tables

Supporting tables for section 5.3.1: Interventions' effects on weekly consumption.

Table 26: Average treatment effects of interventions on weekly consumption

Trial arm	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	50.7				
Non-dispatched Control	52.4	1.929	0.946	0.0415*	3.80%
Escalating payments	56.3	4.571	0.945	<0.001***	9.01%
Escalating + RBT nudge	56.5	3.753	0.961	<0.001***	7.40%
Escalating + RBT payment	55.5	3.531	0.967	<0.001***	6.96%

10.1.4 Sub-group analyses supporting tables: consumption during events

Supporting tables for Section 5.4.1.1: Time of day subgroup analysis of average kW demand

Table 27: Average kW demand during turn-up events by time block (EFA)

Trial arm	Split value	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	EFA 1: 23:00 - 03:00	1.083				
Treatment		1.333	0.22	0.025	<0.001***	20.35%
Dispatched Control	EFA 2: 03:00 - 07:00	0.731				
Treatment		0.891	0.143	0.023	<0.001***	19.54%
Dispatched Control	EFA 4: 11:00 - 15:00	0.134				
Treatment		0.169	0.031	0.006	<0.001***	23.48%

Table 28: Average kW demand during turn-down events by time block (EFA)

Trial arm	Split value	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	EFA 2: 03:00 - 07:00	0.161				
Treatment		0.163	-0.004	0.014	0.7566	-2.74%
Dispatched Control	EFA 3: 07:00 - 11:00	0.08				
Treatment		0.077	-0.005	0.006	0.4305	-6.04%
Dispatched Control	EFA 5: 15:00 - 19:00	0.106				
Treatment		0.108	0	0.008	0.973	-0.26%
Dispatched Control	EFA 6: 19:00 - 23:00	0.172				
Treatment		0.164	-0.012	0.011	0.2928	-6.71%

Supporting tables for Section 5.4.1.2: Day-type subgroup analysis of average kW demand

Table 29: Regression results for average kW demand (turn-up) by day type

Trial arm	Day type	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	Weekday	0.579				
Treatment	Weekday	0.711	0.116	0.013	<0.001***	20.10%
Dispatched Control	Weekend	0.537				
Treatment	Weekend	0.667	0.118	0.015	<0.001***	22.01%

Table 30: Regression results for average kW demand (turn-down) by day type

Trial arm	Day type	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	Weekday	0.121				
Treatment	Weekday	0.118	-0.006	0.005	0.2505	-5.14%
Dispatched Control	Weekend	0.121				
Treatment	Weekend	0.123	-0.002	0.011	0.8496	-1.66%

Supporting tables for Section 5.4.1.3: Battery size subgroup analysis of average kW demand

Table 31: Regression results for average kW demand (turn-up), split by EV battery size

Trial arm	Battery size	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	<50 kWh	0.522				
Treatment	<50 kWh	0.558	0.047	0.024	0.0454*	9.08%
Dispatched Control	>=50 kWh	0.58				

Trial arm	Battery size	Avg.	Coefficient	SE	P-value	% difference
Treatment	>=50 kWh	0.729	0.129	0.013	<0.001***	22.15%

Table 32: Regression results for average kW demand (turn-down), split by EV battery size

Trial arm	Battery size	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	<50 kWh	0.074				
Treatment	<50 kWh	0.074	0.001	0.008	0.9208	1.02%
Dispatched Control	>=50 kWh	0.132				
Treatment	>=50 kWh	0.127	-0.009	0.006	0.1372	-6.84%

Supporting tables for Section 5.4.1.4: EAC subgroup analysis of average kW demand

Table 33: Regression results for average kW response during turn-up events, split by EAC

Trial arm	EAC bracket	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	<7300 kWh	0.450				
Treatment	<7300 kWh	0.543	0.082	0.013	<0.001***	18.30%
Dispatched Control	>=7300 kWh	0.721				
Treatment	>=7300 kWh	0.879	0.151	0.02	<0.001***	20.90%

Table 34: Regression results for average kW response during turn-down events, split by EAC

Trial arm	EAC bracket	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	<7300 kWh	0.08				
Treatment	<7300 kWh	0.088	0.006	0.006	0.3094	7.07%
Dispatched Control	>=7300 kWh	0.174				
Treatment	>=7300 kWh	0.154	-0.021	0.009	0.0270*	-11.83%

10.1.5 Sub-group analyses supporting tables: plug-in hours

Supporting tables for Section 5.4.2.1: Time of day subgroup analysis of weekly plug-in hours

Table 35: Average treatment effects of weekly plug-in hours on day type

Trial arm	EFA comparison	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	EFA 1: 23:00 - 03:00	8.72				
Treatment		11.12	2.206	0.148	<0.001***	25.30%
Dispatched Control	EFA 2: 03:00 - 07:00	8.32				

Trial arm	EFA comparison	Avg.	Coefficient	SE	P-value	% difference
Treatment		10.55	2.043	0.143	<0.001***	24.56%
Dispatched Control	EFA 3: 07:00 - 11:00	4.64				
Treatment		6.09	1.344	0.106	<0.001***	28.99%
Dispatched Control	EFA 4: 11:00 - 15:00	3.18				
Treatment		4.4	1.14	0.093	<0.001***	35.79%
Dispatched Control	EFA 5: 15:00 - 19:00	4.41				
Treatment		5.98	1.478	0.109	<0.001***	33.52%
Dispatched Control	EFA 6: 19:00 - 23:00	7.43				
Treatment		9.67	2.081	0.137	<0.001***	28.02%

Supporting tables for Section 5.4.2.2: Day type subgroup analysis of weekly plug-in hours

Table 36: Average treatment effects of interventions on plug-in hours by day type

Trial arm	Day type	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	Weekday	26.25				
Treatment	Weekday	34.18	7.331	0.497	<0.001***	27.92%
Dispatched Control	Weekend	10.44				
Treatment	Weekend	13.62	2.961	0.205	<0.001***	28.36%

Supporting tables for Section 5.4.2.3: Battery size subgroup analysis of weekly plug-in hours

Table 37: Average treatment effects of interventions on plug-in hours by battery size

Trial arm	EV battery size	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	<50 kWh	56.36				
Treatment	<50 kWh	66.91	11.159	1.824	<0.001***	19.80%
Dispatched Control	>=50 kWh	32.18				
Treatment	>=50 kWh	44.01	10.705	0.684	<0.001***	33.26%

Supporting tables for Section 5.4.2.4: EAC subgroup analysis of weekly plug-in hours

Table 38: Average treatment effects of interventions on plug-in hours by household consumption

Trial arm	EAC	Avg.	Coefficient	SE	P-value	% difference
Dispatched Control	<7300 kWh	31.62				
Treatment	<7300 kWh	43.75	11.392	0.922	<0.001***	36.03%
Dispatched Control	>=7300 kWh	43.05				
Treatment	>=7300 kWh	52.33	8.655	1.03	<0.001***	20.10%

10.2 Further details on OVO methodology: Imbalance in matched control group inactivity and methodological adjustments

Ensuring balance between treatment and control groups is essential when evaluating the impact of interventions. During the analysis, an imbalance was identified in the activity levels of the matched control group pool compared to the treatment groups. This required methodological adjustments to protect the integrity and reliability of the findings.

10.2.1 The observed imbalance

The matched control group contained a higher proportion of inactive chargers than the treatment groups. Weekly data showed that 23% of matched control accounts were inactive, compared to 17% of treatment accounts. A pre-defined “stasis” flag (indicating no charging activity over a specified period) reinforced this pattern: 5.3% of the control group were in stasis compared with 3% of the treatment group.

10.2.2 Impact on analysis

This imbalance arose from how missing values (NULLs) were handled during the matching process. Initially, NULL values were replaced with average charging characteristics. While sensible for new customers without prior data, this approach unintentionally pulled in more permanently inactive customers from outside the trial into the matched control group.

The issue was most significant for consumption analysis. Here, NULL values were interpreted as 0 kWh consumption. Because the control group contained more inactive customers, this created the risk of bias:

- Simply removing inactive customers would have been inappropriate, since some inactivity could have been caused by treatment itself.
- Leaving the imbalance unaddressed risked overestimating treatment effects, as treatment groups showed slightly higher engagement.

10.2.3 Solution for robust analysis

To address this, a three-part solution was adopted:

1. Pre-trial stasis flag as a covariate

- A “stasis” flag was created for each customer based only on pre-trial data.
- This provided a baseline measure of inactivity (e.g. 5.3% in control vs. 3% in treatment).
- It was included as a covariate in analysis to control for pre-existing inactivity, while avoiding contamination from post-trial behaviour.

2. Refined imputation rules

- Weekly NULL values were imputed conservatively as 0.
- For consumption analysis, the pre-trial stasis flag ensured that customers already inactive before the trial were not automatically imputed as 0 kWh in ways that could bias results.

3. Empirical validation

- Different imputation strategies were tested using a base regression model.
- The chosen approach (NULLs → 0 + pre-trial stasis covariate) produced a treatment effect of 0.55 sessions, compared with 0.49 sessions under the base model and 0.59 sessions if all NULLs were imputed as 0.
- Although still producing a slightly larger treatment effect (due to higher NULL rates in the control group), this strategy was judged the most robust and transparent, providing the best balance between accuracy and methodological soundness.

10.3 Further details on OVO methodology: Introduction of SMS behavioural prompts

From 2 August 2025, an additional communication channel was introduced for customers in the two behavioural prompt treatment groups: Escalating + RBT nudge and Escalating + RBT payment. Alongside the existing email communication, eligible customers also received an SMS text message encouraging them to extend their ready-by-time for a specified period.

Eligibility for SMS was based on whether customers had previously provided consent to receive text messages from OVO Energy. The proportion of customers eligible to receive SMS was:

- Escalating + RBT nudge: 6%
- Escalating + RBT payment: 7%

A statistical test confirmed that this one-percentage-point difference in eligibility was not statistically significant ($z = 1.51$, $P=0.13$). In other words, the proportion of ineligible customers was very similar across groups.

As a result, the introduction of SMS messaging was not expected to affect the validity of the planned primary, secondary, or exploratory analyses.

10.4 OVO availability trial technical issue: Customer opt-out volume

10.4.1 Issue summary and identification

A higher trial "opt out" volume was observed in the Availability trial starting in late August 2025, specifically noted in data on 27 August, when compared to the volumes seen in July and September. Note that this only impacted OVO customers.

Upon investigation, a technical instance within the Charge Anytime system was identified. This instance temporarily removed customers from receiving the associated tariff add-on, thereby making them ineligible for the remainder of the CrowdFlex Summer 2025 trial.

10.4.2 Impact on customers and trial volumes

The technical issue resulted in 725 customers in the treatment groups being erroneously dropped from the trial.

- These customers missed out on the opportunity to participate and earn incentives during the final month of the trial (September 2025).
- Impacted customers were not dispatched or optimised in line with the trial event schedule.
- They did not receive incentives or notifications requesting changes to their RBT, if they belonged to those specific treatment groups.
- They were unable to earn trial incentives for their plug-in sessions/behaviours.

10.4.3 Impact on analysis

Due to the drop-out, plug-in and RBT data from 27 August to the end of the trial on 28 September is not available for these impacted customers.
However, checks were completed to ensure the non-inclusion did not adversely impact the remaining data and analysis. The dropouts were confirmed to be balanced by trial arm and appeared to be random.

10.4.4 Specific checks performed include:

1. Checking for an even volume spread across treatment groups (and misfit or control).
2. Checking weekly plug-in sessions, plug-in hours, and Half-Hourly (HH) consumption during events, where no statistical difference was found between the dropped customers and the remaining population.
3. Running weekly and kW regressions adding a binary flag for the dropped customers as a covariate; the results showed that none of the coefficients for the flag were statistically significant, suggesting the drop was not associated with charging behaviour or consumption.
4. It was noted that post-23 August, the dropped customers were no longer dispatched nor invited to change their RBT, thus no consumption/RBT data is available for that period.

10.4.5 Solution agreed

OVO agreed to run relevant analysis, in consultation with CNZ, covering periods pre and post 27 August. This analysis detailing the impact of the error will be included in the appendix of the end of Summer '25 trial analysis pack.

10.5 Ohme results tables

Please note that SE refers to the Standard Error, and that the % difference is the difference between the 'Coefficient' and the dispatched control average (Avg.).

10.5.1 Primary analysis

Supporting tables for section 8.1.1: Interventions' effects on turn-up and turn-down events.

Table 39: Regression results for average kW demand during turn-up events by trial arm

Trial Arm	Avg.	Coefficient	SE	P-value	% difference
Non-dispatched Control	0.432	-0.044	0.021	0.0392*	-9.23%
Dispatched control	0.476				
Standard no-prompt	0.614	0.138	0.017	<0.001***	28.97%
Standard + prompt	0.645	0.169	0.018	<0.001***	35.45%
Enhanced no-prompt	0.613	0.137	0.017	<0.001***	28.84%
Enhanced + prompt	0.649	0.173	0.018	<0.001***	36.41%

Table 40: Regression results for average kW demand during turn-down events by trial arm

Trial Arm	Avg.	Coefficient	SE	P-value	% difference
Non-dispatched Control	0.164	0.079	0.012	<0.001***	92.01%
Dispatched control	0.085				
Standard no-prompt	0.097	0.012	0.007	0.1050	14.14%
Standard + prompt	0.082	-0.004	0.007	0.6247	-4.21%
Enhanced no-prompt	0.087	0.002	0.007	0.7652	2.56%
Enhanced + prompt	0.075	-0.010	0.007	0.1729	-11.61%

10.5.2 Secondary analysis tables

Supporting tables for section 8.2.1: Interventions' effects on plug-in sessions.

Table 41: Regression results for interventions' effect on weekly plug-in sessions by trial arm

Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Non-dispatched Control	2.526	-0.021	0.098	0.8322	-0.82%
Dispatched control	2.547				
Standard no-prompt	3.380	0.834	0.080	<0.001***	32.74%
Standard + prompt	3.431	0.884	0.080	<0.001***	34.73%
Enhanced no-prompt	3.357	0.810	0.079	<0.001***	31.82%
Enhanced + prompt	3.479	0.932	0.080	<0.001***	36.60%

Supporting tables for section 8.2.2: Interventions' effects on plug-in hours.

Table 42: Regression results for interventions' effect on weekly plug-in hours by trial arm

Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Non-dispatched Control	28.161	-0.323	1.125	0.7742	-1.13%
Dispatched control	28.483				
Standard no-prompt	37.846	9.363	0.917	<0.001***	32.87%
Standard + prompt	39.112	10.629	0.920	<0.001***	37.32%
Enhanced no-prompt	37.881	9.398	0.915	<0.001***	32.99%
Enhanced + prompt	40.158	11.674	0.923	<0.001***	40.99%

10.5.3 Exploratory analysis

Supporting tables for section 8.3.1: Interventions' effects on arrival state of charge.

Table 43: Regression results for interventions' effect on SoC upon arrival

Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Non-dispatched Control	52.540	-4.397	2.667	0.0992	-7.72%
Dispatched control	56.933				
Standard no-prompt	52.330	-4.612	2.059	0.0251*	-8.10%
Standard + prompt	53.255	-3.677	2.050	0.0729	-6.46%
Enhanced no-prompt	53.117	-3.811	2.062	0.0646	-6.69%
Enhanced + prompt	52.838	-4.095	2.051	0.0459*	-7.19%

Supporting tables for section 8.3.2: Interventions' effects on weekly consumption

Table 44: Regression results for interventions' impact on average weekly kWh consumption

Trial Arm	Weekly avg.	Coefficient	SE	P-value	% difference
Non-dispatched Control	44.807	0.407	1.568	0.7952	0.92%
Dispatched control	44.401				
Standard no-prompt	51.102	6.702	1.223	<0.001***	15.09%
Standard + prompt	50.838	6.438	1.223	<0.001***	14.50%
Enhanced no-prompt	50.856	6.456	1.222	<0.001***	14.54%
Enhanced + prompt	50.466	6.066	1.218	<0.001***	13.66%

10.6 OVO balance checks and matching results

This section summarises balance checks conducted to assess whether both the matched control groups and the treatment group are statistically comparable across key variables. These checks ensure that differences observed in later analyses can be attributed to the treatments rather than pre-existing differences between groups. Each table reports mean estimates for both groups, sample sizes, test statistics, and p-values from two-sample tests.

Table 45: Comparison of key variables across treatment groups

Variable type	Treatment group (comparison 1)	Treatment group (comparison 2)	Estimate 1	Estimate 2	Sample size (comparison 1)	Sample size (comparison 2)	p
OVO tenure	1	2	5.14	5.13	2697	2695	0.93
	1	3	5.14	5.08	2697	2699	0.602
	2	3	5.13	5.08	2695	2699	0.665
Mean weekly consumption (last 3 months)	1	2	69.5	69.5	2644	2637	0.999
	1	3	69.5	70.4	2644	2651	0.48
	2	3	69.5	70.4	2637	2651	0.498
Mean weekly plug-in hours (last 3 months)	1	2	47.8	48.5	2644	2637	0.359
	1	3	47.8	48	2644	2651	0.785
	2	3	48.5	48	2637	2651	0.519
Mean weekly plug-in sessions (last 3 months)	1	2	2.92	2.97	2644	2637	0.322
	1	3	2.92	2.95	2644	2651	0.523
	2	3	2.97	2.95	2637	2651	0.722

Variable type	Treatment group (comparison 1)	Treatment group (comparison 2)	Estimate 1	Estimate 2	Sample size (comparison 1)	Sample size (comparison 2)	p
June 2025 total consumption	1	2	260	262	2682	2543	0.824
	1	3	260	264	2682	2568	0.509
	2	3	262	264	2543	2568	0.68

Table 46: Comparison of key variables between treatment and control groups, pre and post matching

Variable type	Arm type 1	Arm type 2	estimate1	estimate2	Sample size (arm type 1)	Sample size (arm type 2)	p
Mean weekly plug-in hours (last 3 months)	Control: Prior to matching	Pooled treatment groups	38.4	48.1	11364	7932	<0.001
Mean weekly plug-in hours (last 3 months)	Control: Dispatched	Control: Non-dispatched	47.1	48.1	2592	2581	0.230
Mean weekly plug-in hours (last 3 months)	Control: Dispatched	Pooled treatment groups	47.1	48.1	2592	7932	0.144
Mean weekly plug-in hours (last 3 months)	Control: Non-dispatched	Pooled treatment groups	48.1	48.1	2581	7932	0.986
Median ready by hour	Control: Prior to matching	Pooled treatment groups	6.88	6.95	12353	8091	<0.001
Median ready by hour	Control: Dispatched	Control: Non-dispatched	6.98	6.97	2743	2597	0.834
Median ready by hour	Control: Dispatched	Pooled treatment groups	6.98	6.95	2743	8091	0.243
Median ready by hour	Control: Non-dispatched	Pooled treatment groups	6.97	6.95	2597	8091	0.376
Mean weekly consumption	Control: Prior to matching	Pooled treatment groups	69.6	69.8	11364	7932	0.797

Variable type	Arm type 1	Arm type 2	estimate1	estimate2	Sample size (arm type 1)	Sample size (arm type 2)	p
Mean weekly consumption	Control: Dispatched	Control: Non-dispatched	71.5	70.6	2592	2581	0.503
Mean weekly consumption	Control: Dispatched	Pooled treatment groups	71.5	69.8	2592	7932	0.11
Mean weekly consumption	Control: Non-dispatched	Pooled treatment groups	70.6	69.8	2581	7932	0.43