Public

CrowdFlex: Full Programme Executive Summary

Utilisation Payments Trials 2024-2025

Submitted: December 2025

NESO
National Energy System Operator

Contents

1. Introduction

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.

This report summarises three Randomised Controlled Trials (RCTs) delivered by OVO Energy, in which customers were contacted and paid for their "**utilisation**" response – how much they increased (or decreased) their demand during turn-up (or turn-down) flexibility events. Conducted in Summer 2024, Winter 2024, and Summer 2025, these trials evaluated how different incentive levels and structures influenced consumer behaviour, as well as the impact of notice period ahead of events. Together, the trials involved approximately 107,000 unique OVO Energy customers across hundreds of flexibility events, providing one of the largest evidence bases to date on domestic flexibility in Great Britain.

This is a high-level summary of results from the three trials. Full trial reports are available on the CrowdFlex Beta document portal.

Table 1: The three CrowdFlex utilisation trials

Trial period	Core treatment mechanisms
Summer 2024 *May - July 2024*	Tested varying per-kWh payment levels, including mid-trial changes (High, Medium, Low, High-Low, Low-High).
Winter 2024 *October 2024 - March 2025*	Tested the effect of a Consistency Bonus (CB) alongside varying per-event payment levels (Low, Medium, High, Low + CB, Medium + CB).
Summer 2025 *July - October 2025*	This trial focused primarily on turn-up events and explored different incentive delivery models, including direct £/kWh, £/kWh + community pot, Free electricity (volumetric), and Free electricity (consistency). We also introduced an opt-in stage and tested varied notice periods (2 hours, 5 hours, 18:00 the evening before).

The project followed an iterative approach, adapting later trial designs based on insights from earlier ones.

Table 2: Incentive structures across the three trials

Trial	Treatment group	£/kWh	
		Turn-down	Turn-up
Summer 2024	High	£1.25	£0.29
	High to Low	£1.25 to £0.05	£0.29 to £0.05
	Low to High	£0.05 to £1.25	£0.05 to £0.29
	Med	£0.50	£0.20
Winter 2024	Low	£0.20	£0.10
	Medium	£0.50	£0.15
	High	£1.25	£0.25
	Low + Consistency Bonus	£0.20 + CB	£0.10 + CB
	Medium + Consistency Bonus	£0.50 + CB	£0.15 + CB
Summer 2025	£0.25/kWh	N/A	£0.25
	£0.25/kWh + down events	£0.50	£0.25
	£0.20/kWh + community pot	N/A	£0.20
	Free electricity hours (volumetric)		*
	Free electricity hours (consistency)		*

** These free-electricity incentives were designed to offer comparable headline value in £/kwh terms, but did not pay consumers on this basis; their realised cost to OVO was lower due to lower-than-expected electricity use during the Sunday redemption window, combined with lower overall consistency in event participation than was assumed in the original reward cost modelling.*

Table 3: The number of trial events and trial event hours

Trial	# of events		# total hours	
	Turn-up	Turn-down	Turn-up	Turn-down
Summer 2024	13	20	19	32
Winter 2024	36	53	60	56.5
Summer 2025	48	12	48	12
Total	97	85	127	101

2. Results from CrowdFlex's Utilisation trials

The tables below present the main findings from the three utilisation trials. As these were RCTs, each treatment group can be directly compared with the control group to estimate the effect of different incentive structures on electricity demand during turn-down and turn-up events. The results are based on regression analyses, which estimate how much average hourly electricity demand (in kW) differed in each treatment group versus the control group during event periods.

- The Control Demand (kW) column shows the average demand (kW) during trial events for the Control group.
- The % difference column shows the percentage difference in demand in each treatment group (relative to the Control group's average demand).
- A negative percentage in the *Turn-down* column indicates participants reduced demand more than the control group.
- A positive percentage in the *Turn-up* column indicates participants increased demand more than the control group.
- Stars reflect the statistical significance of the difference, based on the *p-value* from the regression. *** = $p < 0.001$. ** = $p < 0.01$. *= $p < 0.05$.

Larger percentage differences indicate stronger behavioural responses to the incentives, while the number of stars indicates a smaller p-value, where the p-value is the chance that an effect as large or larger would have occurred if the true effect was zero. Note that these effects compare trial groups to each other – including all customers in the group. In other words, we did *not* excluded disengaged customers; and even in the Summer 2025 trial, where customers had to opt in, we included all customers, not just opted in customers (partly, importantly, because there was no "opt-in" in the control group, so excluding non-opted-in customers from treatment groups would have made groups less comparable to each other).

Table 4: Estimated treatment effects on electricity demand during trial events

Trial	Treatment group	Turn-down		Turn-up	
		Control Demand (kW)	% difference	Control Demand (kW)	% difference
Summer 2024	High	0.368	-4.13%***	0.294	19.52%***
	High to Low		-2.55%*		19.46%***
	Low to High		-4.29%***		15.92%***
	Med		-5.05%***		19.80%***
Winter 2024	Low	0.533	-3.18%***	0.418	8.15%***
	Medium		-3.78%***		10.08%***
	High		-4.29%***		12.83%***
	Low + Consistency Bonus		-5.05%***		11.63%***
	Medium + Consistency Bonus		-4.34%***		12.65%***
Summer 2025	£0.25/kWh			0.304	19.55%***
	£0.25/kWh + down events	0.400	-1.42%*		16.29%***
	£0.20/kWh + community pot				18.08%***
	Free hours (volumetric)				16.39%***
	Free hours (consistency)				17.98%***

The utilisation trials demonstrated that financial incentives shifted household electricity demand in the desired direction, with all tested incentive structures yielding statistically significant changes in consumption compared to the relevant control group. Across all observed trials, turn-up events (increasing consumption) generated stronger overall effect sizes compared to turn-down events (reducing consumption). We hypothesise that this may reflect the fact that the flexibility potential for demand turn-up is greater than for demand turn-down: reductions are limited by baseline consumption, whereas increases are constrained only by the household's import capacity.

2.1 Price responsiveness and cost effectiveness

In general, the trials proved a core hypothesis; that customers were price-sensitive – delivering more response when presented with more generous rewards. However, even lower-remunerated treatments delivered meaningful response. In other words, there seemed to be a trade-off between maximizing response volume and minimizing cost per kWh of flexibility response.

This was particularly clear in the Winter 2024 trial results. (Note that the £/kWh costs in these results are higher than the face-value incentive levels because 1) flexibility was measured relative to the control group, while remuneration was calculated against household baselines, and 2) payments had a floor of 0 p, so negative responses were not penalised.)

- The Low Incentive group resulted in the least turn-down (3.18%), but was the most cost-effective option at £1.69/kWh.
- The High Incentive group resulted in slightly higher response, and the greatest turn-down (4.29%), but was significantly less cost-effective at £2.89/kWh.
- Groups featuring the Consistency Bonus delivered very similar response to the High Incentive group, but they were the least cost-effective, with costs up to £4.43/kWh.

The Summer 2025 trial presented a similar picture, wherein treatment effects were similar between incentive groups – thus leading to lower-remunerated groups being much more cost-effective than higher-remunerated ones. In this trial's case, it was the treatments featuring free electricity rewards that were lowest-total-cost and thus lowest cost-per-kWh.

- Free electricity rewards achieved similar demand response as the direct cash payment group but cost only roughly one-third of the expense per unit of flexibility compared to the direct payment mechanisms.
- The Free Hours (Volumetric) group was the most cost-effective at £0.13/kWh.
- The direct payment of £0.25/kWh delivered the highest response, 0.059 kW, but only by a small amount. It cost £0.32/kWh.

Note that care should be taken when comparing cost-per-kWh figures across trials. The estimates are sensitive to event frequency, event type, and season of the year, all of which varied systematically between trials. Within-trial comparisons are therefore more reliable.

2.2 Effect of notice period on demand shift

In Summer 2024, notice period was not randomised and only a small number of short-notice events occurred. As a result, the results from the analysis were non-causal. Within this constrained dataset, shorter-notice events appeared to show slightly larger turn-down responses and longer-notice events slightly larger turn-up responses, but these patterns should be taken as purely suggestive – they could have been driven by event timing and the very small number of observations, not *caused* by notice period.

In Winter 2024–25, the subgroup analysis was repeated with a substantially larger set of events across both shorter and longer notice periods. In contrast to the exploratory findings from Summer 2024, the Winter results showed no consistent or statistically meaningful relationship between notice period length and customer response. Both short and long notice events demonstrated similar treatment effects relative to control, for both turn-up and turn-down events.

Given the relevance of notice period to the design of demand response programmes – and the contrasting, non-causal findings from the previous two trials – the Summer 2025 trial was designed to randomise notice period explicitly, providing the first opportunity to causally identify its effect. The results showed a clear and intuitive pattern: longer notice periods (5-hour and day-ahead notice) produced modestly stronger turn-up responses, while short notice (2-hour) still generated a substantial response. For turn-down events, only the day-ahead notice group showed a statistically significant reduction (though note that this result is in tension with the winter utilisation payments trial results mentioned immediately above, where customers had meaningful turn-down in <4-hour notice events roughly equivalent to turn-down in longer events, perhaps due to turn-down being somewhat easier, even at short notice, in the winter than the summer). These findings provided causal evidence that additional notice can enhance flexibility response, while confirming that for turn-up events, customers remain responsive even with only a few hours' advance warning.

Table 5: Estimated treatment effects of notice period on electricity demand during trial events

Trial	Notice period	Turn-up		Turn-down	
		Control Demand (kW)	% difference	Control Demand (kW)	% difference
Summer 2025	2 hours	0.304	14.94%***	0.400	-0.23%
	5 hours		19.49%***		-1.08%
	Day ahead		18.55%***		-2.95%**

2.3 Demand displacement and spillover (Winter 2024-25 and Summer 2025)

Across both the Winter 2024–25 and Summer 2025 trials, we observed small displacements in consumption around events (in the order of 1-5% per settlement period), but these effects were modest relative to the in-event response and often not statistically significantly different from zero. In addition, for both turn-down and turn-up events, both trials showed spillover into the periods just after an event, as if customers sometimes continued increasing (decreasing) their demand for turn-up (turn-down) events for a few hours even after the event finished. This "crowding-in" of the event direction shift after the event was often of a similar magnitude to the suggestive evidence of displacement. This seeming lack of demand destruction and creation was at odds with our expectations *ex ante* that we would see mostly "shifting" of demand.

However, despite these results suggesting limited displacement – i.e., mostly demand creation or destruction – we caution against drawing a firm conclusion on this question. Displacement may have been too diffuse to detect reliably, even with our large sample sizes and high-frequency (half-hourly) data. This is particularly the case because displacement could theoretically occur many days after an event.

2.4 Evidence concerning fatigue and participation dynamics

Behavioural fatigue was difficult to identify causally, but each trial contributed useful suggestive evidence. In the Summer 2024 trial, treatment effects appeared to decline over the event season, but this trend may be confounded by timing – i.e., events in July may have had lower effects than May events regardless of fatigue. By contrast, the Winter 2024–25 trial did not show this pattern – treatment effects remained relatively consistent over the six-month period of the winter trial, with no evidence of systematic decline in response. Finally, in Summer 2025, turn-up appeared to *increase* across the event season, while turn-down showed the opposite pattern. Again, these apparent month-to-month differences should be interpreted cautiously, as events occurring in different months are not strictly comparable due to underlying seasonal and contextual factors.

To test fatigue more causally, the Summer 2025 trial introduced a randomised comparison between customers facing four turn-up events per week and those facing the same plus one additional turn-down event. This provided the first causal test of whether higher event frequency reduced engagement. The additional weekly event led to a small (2–3%) decline in turn-up performance, suggesting mild behavioural fatigue from more frequent calls.

The Summer 2025 trial also introduced an opt-in stage to each event, allowing direct observation of participation. Longer notice periods increased participation rates: day-ahead notifications produced the highest opt-in rates at 36% for turn-up and 41% for turn-down, compared with 33% and 36% under five-hour notice and 25 and 29% under two-hour notice. However, this did not translate into proportionally greater demand response. The day-ahead notice group had similar turn-up to the five-hour notice group, and only 31% more turn-up than the two-hour notice group despite having 44% more opt-in. One hypothesis is that the extra customers encouraged to opt-in by longer notice may tend to have lower response per household; another is that these customers were more likely to forget the event after opting in. In addition, midway through the trial, SMS event reminders and notifications were randomly assigned to participants and proved effective: they increased participation by around 3% and boosted consumption during turn-up events by a further 3–4 percentage points (a 15–20% uplift on top of the 16–17% baseline). Their impact was most pronounced for shorter-notice events, suggesting that timely reminders can compensate for reduced notice.

2.5 Baseline accuracy

Our trials provided an opportunity to examine the accuracy of “baselines” used in demand response programmes. A baseline establishes the expected level of demand in the absence of a flexibility event, providing a reference against which actual consumption can be compared to calculate the flexibility delivered, particularly when that flexibility is remunerated directly. In CrowdFlex, OVO used a baseline based on recent historical consumption in the same settlement periods to estimate delivered flexibility. This was a variant on P376, a commonly used baseline to compensate customers in Great Britain.

A standard practice when using baselines in this way is to “clip” customers’ demand response at 0. The purpose of clipping is to avoid penalising customers for individual-level shifts that move in the opposite direction of the event’s call – for example, demand increases during turn-down events or decreases during turn-up events. Providers treat these opposite-direction shifts as 0 kWh rather than negative values. In some cases, the system operator also settles with providers using the same clipped flexibility. OVO adopted this standard approach for both customer remuneration and settlement with the flexibility procurer (NESO, in the case of these trials).

We found that clipping reduced the accuracy of baseline-based estimates of flexibility. This inaccuracy was smaller when actual demand response represented a larger share of typical consumption; in other words, a higher signal-to-noise ratio reduced the inflation effect. For example, in the Winter trial, the inflation factor dropped from 8.2 for turn-down events to 3.6 for turn-up events, driven by the effect size being larger for turn-up events. Inaccuracy was also smaller when the demand response provider, in this case OVO, used an opt-in mechanism – thereby excluding non-participating customers from the flexibility calculation.

In the Summer 2025 trial, OVO remunerated only customers who had opted in. We calculated that if there had been no opt-in filter, the remunerated flexibility would have been 169% above what we calculated the true flexibility to be (based on the control group baseline). Filtering on opted-in customers reduced this inaccuracy substantially; the remunerated flexibility was 48% higher than the true flexibility.

Baselining issues are complex, and flexibility market design needs to balance a range of factors, including accuracy, fairness, simplicity, and interoperability. CrowdFlex was able to quantify the accuracy of different baselining approaches with high evidentiary quality, thanks to its RCT structure. These findings showed that some commonly used practices – such as clipping and the inclusion of customers not opted in – can introduce substantial inaccuracies. As flexibility services expand, operators should continue to explore ways to develop robust and common baselining methods to establish a fair playing field for market participants and consumers.

2.6 Subgroup differences

Responses were broadly consistent across customer types. However, there were some interesting subgroups that showed differences. Notably, households with low-carbon technologies (LCTs) displayed the largest differences, with turn-up responses nearly 50% stronger than households without LCTs. Across both Winter 2024–25 and Summer 2025, customers on the Economy 7 (E7) tariff showed slightly higher turn-up performance than non-E7 customers; despite already “turning up” overnight, these customers seemed to still have additional response capability than non E7 customers. In Summer 2025, customers on the Priority Services Register (PSR), a list of electricity customers with specific vulnerabilities or support needs, also exhibited higher turn-up responses relative to non-PSR customers. Note that these subgroup patterns should be interpreted cautiously, as they are not causal – i.e., here, differences may reflect underlying characteristics *correlated* with tariff type, technology ownership, or customer need, rather than being *caused* by those differences.

3. Conclusion

The results from the three CrowdFlex trials offer practical lessons for the design and delivery of future domestic flexibility programmes:

- **Consumer response:** households consistently responded to incentives to provide flexibility. All treatment effects were statistically significant, showing that incentives effectively change electricity use during events. Overall effect sizes for turn-up events were far higher than for turn-down events.
- **Incentive design:** Lower incentives delivered more flexibility per pound spent, while higher incentives delivered greater volumes, pointing to a trade-off between maximising response and minimising cost. At the incentive levels tested, free electricity incentives achieved comparable demand shifts at lower cost than direct cash payments, highlighting the potential for low-cost, scalable models.
- **Event scheduling:** Notice period had a clear influence on response, including day-ahead notification for turn-down events increasing response. When designing frequent or mixed event schedules (i.e., combining turn-up and turn-down events), system operators should consider the potential for mild behavioural fatigue which was observed to slightly dampen turn-up performance.
- **Communication strategy:** Clear, timely communications - such as text message (SMS) notifications and reminders – helped convert customer intent into active participation and performance.
- **Role of technology:** Adoption of low-carbon technologies such as EVs, heat pumps, and batteries may increase the scope for response to flexibility calls from system operators, with significant potential for linking flexibility with wider electrification and automation to deliver scalable demand response.

Across three seasons of experimentation, the trials provided robust causal evidence that households consistently shifted demand in response to incentive. While customers were sensitive to payment levels, even modest rewards produced measurable changes in consumption, underscoring the potential for cost-effective programme design. These findings show that well-designed domestic flexibility can be a valuable component of a clean, reliable electricity system.