



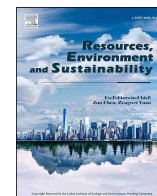
Event-level monitoring of water-saving interventions in a living lab: device effectiveness and user experience

Downloaded from: <https://research.chalmers.se>, 2026-10-10 21:32 UTC

Citation for the original published paper (version of record):

Knutsson, J., Hiller, C., Klingberg, J. et al (2026). Event-level monitoring of water-saving interventions in a living lab: device effectiveness and user experience. *Resources, Environment and Sustainability*, 27. <http://dx.doi.org/10.1016/j.resenv.2026.100381>

N.B. When citing this work, cite the original published paper.



Research article

Event-level monitoring of water-saving interventions in a living lab: device effectiveness and user experience

Jesper Knutsson^{a,*}, Carolina Hiller^b, Josefine Klingberg^b, Sara Renström^b, Tobias Svanberg^c^a Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, SE-412 96, Sweden^b RISE Research Institutes of Sweden AB, Box 857, Borås, 501 15, Sweden^c IKEA of Sweden AB, Älmhult, SE-343 81, Sweden

ARTICLE INFO

Keywords:

Event-level monitoring
Living lab
Difference-in-differences
Water-saving devices
User experience
Smart metering

ABSTRACT

Most field studies of residential water-saving devices rely on aggregate meter data that cannot isolate device-level effects. We present a tap-level monitoring framework applied at HSB Living Lab, a 29-apartment facility in Gothenburg, Sweden. Over 32 months, 205,097 water use events were recorded across 80 taps (31 treated, 49 control) for nine water-saving products installed in three rounds. Event-level non-parametric tests, a difference-in-differences (DiD) framework, and three user experience surveys ($n = 26$) were combined.

Event-level tests found that 65–81% of treated taps showed statistically significant shifts in the distribution of individual use events (median volume change -7.1%); because an event here reflects a ~ 10 -min reporting window rather than a fully independent discrete use, we treat this as exploratory description rather than confirmatory evidence. The aggregate, tap-clustered difference-in-differences estimate—our confirmatory analysis—was not significant. Product-specific effects were heterogeneous; in this small, exploratory survey ($n = 26$ respondents, 3–10 per product), the least-liked product showed the largest measured reduction and the most-liked showed none, a suggestive pattern that would need confirmation in larger samples. Event-level monitoring can surface intervention effects that aggregate data may not resolve; we interpret this as a consequence of finer measurement resolution rather than evidence of savings hidden within aggregate figures. We emphasise, however, that statistically significant event-level shifts indicate changes in the distribution of individual water-use events, not necessarily reductions in total water use at the building scale.

1. Introduction

Residential water consumption constitutes a significant share of urban water demand, and reducing it is a key component of sustainable water management strategies (Addo et al., 2018). A wide range of water-saving devices—including aerators, flow restrictors, sensor-activated taps, and shower monitors—are commercially available and frequently promoted by utilities and housing providers. However, robust evidence on the real-world effectiveness of these devices in occupied buildings remains limited (Inman and Jeffrey, 2006).

Most existing studies rely on aggregate billing data or monthly meter reads, which obscure the mechanisms through which savings occur (or fail to occur). The temporal resolution of such data makes it difficult to distinguish intervention effects from seasonal patterns, occupancy changes, and behavioral adaptation. Recent advances in smart metering have enabled higher-resolution monitoring (Cominola et al., 2015), and

end-use analysis methods have been developed to disaggregate consumption into individual events (Nguyen et al., 2014; Beal and Stewart, 2014). Yet few studies combine high-resolution quantitative data with systematic user experience assessments, despite growing recognition that user experience is critical for sustained adoption of water-saving technologies (Ehret et al., 2021). We emphasise that this study does not claim that user experience explains measured water volumes; rather, effectiveness (measured consumption change) and user acceptance (satisfaction) are treated as two distinct, parallel outcomes, both of which matter for the real-world performance and sustained adoption of water-saving technologies.

Living labs—real-world research environments where technologies are tested under authentic conditions—offer a promising setting for water intervention studies (Bergvall-Kåreborn et al., 2015). They allow controlled comparisons while maintaining ecological validity. However, methodological challenges remain: living labs often feature pre-existing

* Corresponding author.

E-mail address: jesper.knutsson@chalmers.se (J. Knutsson).<https://doi.org/10.1016/j.resenv.2026.100381>

Received 19 May 2026; Received in revised form 3 September 2026; Accepted 3 September 2026

Available online 7 September 2026

2666-9161/© 2026 The Authors. Published by Elsevier B.V. on behalf of Lishui Institute of Ecology and Environment, Nanjing University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

efficient installations, occupancy may not be randomly assigned, and sample sizes are inherently constrained by building capacity.

This study addresses these challenges by presenting a monitoring and evaluation framework that tracks water consumption at individual tap points, identifies discrete use events, and assesses intervention effects at multiple levels of analysis. We apply this framework to data from HSB Living Lab, a 29-apartment residential research facility in Gothenburg, Sweden, where nine commercially available water-saving products were installed in three rounds during 2024–2025. The installations were accompanied by informative communication materials designed to raise residents' awareness of the intervention.

While event-level monitoring is established in the water research literature (Beal et al., 2011; Nguyen et al., 2014; Cominola et al., 2019), this study contributes by: (1) applying direct tap-level monitoring rather than algorithmic disaggregation of whole-house flow, eliminating classification errors; (2) integrating quantitative consumption data with systematic user experience surveys; and (3) evaluating multiple product categories within a single controlled setting under consistent conditions.

Taken together, these considerations frame the objective of this study: to rigorously evaluate the effectiveness of water-saving intervention strategies using commercially available efficiency devices in a living lab setting, combining the assessment of device effectiveness and user experience as two parallel outcomes. The study context—an already efficient building with environmentally aware tenants—provides conservative estimates of achievable savings in modern Swedish residential buildings.

2. Background and related work

2.1. Water-saving devices in residential buildings

The literature on residential water-saving devices can broadly be divided into studies of technology performance (laboratory tests of flow rates and efficiency) and studies of real-world effectiveness (field trials measuring actual consumption changes). Laboratory studies consistently demonstrate that aerators, flow restrictors, and sensor taps reduce flow rates by 20–60% compared to standard fixtures. However, field studies show more modest and variable results, with reported savings ranging from near-zero to 30%, depending on installation context, user behavior, and measurement methodology (Inman and Jeffrey, 2006; Addo et al., 2018).

A key limitation of many field studies is their reliance on aggregate consumption data. Monthly or quarterly billing data cannot distinguish changes at individual taps, making it impossible to attribute savings to specific devices. Some studies have used sub-metering at the fixture level (Beal et al., 2011), but typically at daily or hourly resolution, which still aggregates multiple use events.

2.2. Event-level water consumption analysis

Event-level analysis—also termed end-use analysis—disaggregates total consumption into individual use events based on flow patterns. Pioneering work by Nguyen et al. (2014) developed autonomous classification systems for residential water end uses, while Beal and Stewart (2014) demonstrated the value of end-use data for understanding peak demand. More recently, Cominola et al. (2019) applied data mining techniques to smart meter data to uncover heterogeneous consumption behaviors.

Our approach differs in that we monitor individual taps directly via dedicated meters rather than disaggregating whole-house flow traces, enabling before–after comparisons at the tap level with thousands of events per tap (see Section 4.1 for details).

2.3. Difference-in-differences in resource studies

The difference-in-differences (DiD) method is a quasi-experimental

approach widely used in economics and increasingly adopted in environmental and resource studies. It compares changes in outcomes between treated and control units over time, controlling for time-invariant differences and common temporal trends. Ferraro and Price (2013) applied DiD to evaluate the impact of social norm messages on household water use in a large-scale field experiment, demonstrating its value for causal inference in water research. When combined with time fixed effects and standard errors clustered at the unit level, DiD provides more credible estimates than simple before–after comparisons, which are confounded by secular trends (Allcott, 2011). A substantial and growing body of work has applied this design to high-frequency residential water data. Daminato et al. (2021) exploited a staggered smart-meter roll-out as a natural experiment and used difference-in-differences to isolate a behavioral reduction of roughly 2% from the mechanical effect of meter replacement. Closely related is the literature on online “Home Water Use Reports,” the water analogue of the social-comparison Home Energy Reports evaluated by Allcott (2011): randomised trials by Ferraro and Price (2013) and Brent et al. (2015) and the social-information campaign of Jaime Torres and Carlsson (2018) report reductions on the order of 4–7%, while feedback experiments such as Fielding et al. (2013) and the high-frequency smart-meter feedback synthesised by Sönderlund et al. (2016) span a wider 2.5–28.6% range. Datta et al. (2021) similarly document persistent reductions in household water use following a low-cost behavioural nudge intervention (Datta et al., 2021). Jessoe and Rapson (2014) show, in the energy sector, that high-frequency information sharply increases responsiveness, underscoring why the temporal resolution of smart meters matters. Our study complements this literature: whereas these studies operate at the household or aggregate-meter level, the present tap-level event data resolve which fixtures and use-points drive—or fail to drive—the aggregate response, a distinction that whole-household smart-meter data cannot make.

3. Study site and data

3.1. HSB living lab

HSB Living Lab is a purpose-built four-story modular research building with a footprint of approximately 400 m². It is located on the Chalmers University of Technology campus in Gothenburg, Sweden. The building was initiated in 2013 as a joint venture between Chalmers and HSB (a major Swedish housing cooperative) and has been operational with tenants since June 2016 (Knutsson and Marx, 2016). The first and third floors contain studio apartments with shared kitchens and shower facilities, while the second floor contains five standalone one-bedroom apartments. The ground floor houses a social laundry studio, a co-creation lab, and data/communication infrastructure. In total the building contains 29 apartments and is equipped with over 2000 sensors for real-time monitoring of utility consumption, indoor climate, and building performance. The apartments are let to students and guest researchers who live in the building as ordinary residential tenants, selected through HSB's standard allocation process. The rental contracts are short-term (typically 1–3 years), and most residents are students or young professionals affiliated with Chalmers. Because the kitchens and showers on floors 1 and 3 are shared between multiple residents, tap-level data at those locations reflects usage by more than one person. Water consumption costs are included in the monthly rent; residents do not receive individual water bills and thus have no direct financial incentive or awareness of their personal water use. Tenant turnover is higher than in typical Swedish housing, particularly during summer months when occupancy may temporarily decrease.

Each apartment is equipped with ZigBee-enabled rotary vane water meters (Lund Science AB, Sweden) on all individual water use points: kitchen tap, bathroom sink tap(s), and shower. The meters measure water flow continuously and record cumulative volumes with a resolution of 1 L; the aggregated volume data are transmitted wirelessly at

approximately 10-min intervals via the building's IoT platform (Knutsson and Knutsson, 2020). Although the volumetric resolution is coarse for individual events, the continuous long-term monitoring across all taps provides a large dataset from which event-level patterns can be extracted.

Demographic information was collected through the first survey round ($n = 15$ respondents out of 29 households, 52% response rate). The majority of respondents were aged 20–35 and lived alone.

The building was already equipped with water-efficient fixtures (cold-start kitchen taps, low-flow bathroom aerators), creating a high-efficiency baseline that limits marginal impact of additional interventions.

3.2. Intervention design

Nine water-saving products were installed across three rounds in 2024–2025 as part of the DROPIT project (Drop-in ICT solutions for water savings). Products were allocated to apartments based on practical considerations (compatibility, availability) rather than random assignment; each treated tap is paired with untreated control taps of the same use-point type in non-treated apartments. Each round was accompanied by informational posters and leaflets. We therefore define “treatment” as the combined package of the installed device and the accompanying information, rather than the device hardware in isolation; the reported effects should be read accordingly. Because the informational materials were displayed in shared kitchens, bathrooms and common areas of buildings that contain both treated and control taps, general awareness raised by the campaign reached treated and control residents alike and is therefore largely absorbed by the difference-in-differences comparison against contemporaneous controls. We nonetheless cannot fully separate a localized information or novelty (Hawthorne) response among residents of treated taps from the effect of the device itself, and we flag this explicitly as a limitation (Section 6.5). Table 1 provides an overview of the products. Brief mechanisms are summarized in Table 1; fuller technical specifications (e.g. rated flow rates and operating principles for the eco-booster, mist nozzle and prototype kitchen taps) are provided in Supplementary Section S1. The cold start tap warrants specific comment: it defaults the single-lever mixer to the cold position so that users do not draw hot water unintentionally, and is marketed primarily for energy saving (avoiding needless heating) rather than for reducing water volume. We included it because drop-in mixer replacements are a common retrofit and its volumetric effect had not, to our knowledge, been quantified at the tap level; consistent with its mechanism, our data show it does not reduce—and may slightly increase—water volume, which we report as an informative null/negative result rather than an expected saving.

3.3. Data collection period

The analysis uses flow data from April 18, 2023, to December 31,

2025, spanning 32 months. The pre-intervention baseline period varies by installation round, ensuring at least 6 months of pre-data for each treated tap. In total, the dataset comprises 205,097 water use events across 31 treated taps and 49 control taps, after excluding one product (Orbital Shower, $n = 3069$ events) that was removed from the analysis due to technical issues during the trial.

3.4. Survey data

A baseline survey was distributed to all 29 households before the first installation round. Three follow-up surveys were distributed after each installation round to the 15 households that received products. Survey 1 (October–December 2024, $n = 10$), covered rounds 1 and 1b; Survey 2 (April 2025, $n = 11$), covered round 2; and Survey 3 (October 2025, $n = 5$), covered round 3. Each survey included product-specific questions on satisfaction and functionality. Responses were collected on a 5-point Likert scale ranging from ‘strongly disagree’ (1) to ‘strongly agree’ (5). Surveys 2 and 3 additionally included a perceived water-saving question. Some respondents participated in multiple survey rounds. Total unique respondents across all rounds: 26.

4. Methods

4.1. Event detection

Consumption volumes were derived by differencing consecutive meter readings (Section 3.1). Water use events were identified using a gap-based segmentation algorithm: consecutive non-zero flow records separated by fewer than 620 s were grouped into a single event. The 620-s threshold was determined empirically from the long-tailed distribution of inter-record time gaps across all taps (see Supplementary Fig. S2), separating within-event pauses from between-event idle periods. Because the threshold falls within a wide, sparsely populated trough of the gap distribution, event segmentation is highly insensitive to its exact value. We confirmed this with a sensitivity analysis: re-segmenting the full record stream with thresholds ranging from 300 to 1000 s changed the total number of detected events by less than 0.2%, so the event partition—and therefore every downstream event-level and daily-level result—is effectively unchanged across this range.

Each event is characterized by its total volume (liters), duration (seconds), and mean flow rate (L/min). Given the 1-liter volumetric resolution, events involving very small volumes (<1 L) may be under-represented, and duration estimates for brief events are limited by the approximately 10-min reporting interval. This primarily affects very short sink events such as brief hand-rinsing. Consequently, the dataset is biased toward larger events; derived metrics such as duration and flow rate are approximate, and any measured intervention effects are conditional on this detection threshold. To quantify this, 35% of detected events sit at the 1-liter measurement floor and 61% are at or below 3 L, while the reporting interval is approximately 600 s; an “event” here is

Table 1

Overview of water-saving products installed in HSB Living Lab. Round indicates installation wave; n indicates number of treated taps.

Product	Location	Round	n	Mechanism
Sensor tap	Kitchen/Sink	1	10	Automatic shut-off
Spray tap	Kitchen	1	2	Spray mode for rinsing
Cold start tap	Sink	1	9	Cold water as default
Shower monitor	Shower	1	7	Real-time feedback display
Mist nozzle	Sink	1b, 3	6	Mist spray atomization
Eco-push tap	Sink	2	2	Timed push mechanism
Eco-booster	Shower, Kitchen	2	3 + 2	Flow limitation device
Auto water diverter	Shower	3	6	Cold start-up water capture
Kitchen tap (proto.)	Kitchen	3	4	Low-flow prototype

Note: Round 1b indicates a delayed installation within the first round. Some taps received combinations of products across rounds; for example, sensor taps in Round 1 were complemented by eco-boosters in Round 2 at certain locations. Round 1 products were installed in May 2024, Round 1b (mist nozzles) in August 2024, Round 2 in September 2024, and Round 3 in November 2024.

therefore better understood as a 10-min active reporting window than as a single physical water-use action. We accordingly treat the event-level tests as exploratory descriptions of distributional shifts (Section 4.2) and rely on the daily-level difference-in-differences as the confirmatory analysis, for which the 1-liter and 10-min resolution is not a limitation. This contrasts with high-frequency end-use studies that resolve individual fixture events from sub-minute trace data; our tap-specific metering trades that temporal resolution for direct, classification-error-free attribution to a known fixture, and the two approaches are complementary. The direct tap-level monitoring approach eliminates the classification errors inherent in algorithmic end-use disaggregation from whole-house flow traces (Nguyen et al., 2014), at the cost of requiring dedicated metering infrastructure at each tap. The procedure produced 205,097 events across 80 monitored taps after exclusion of one product (Orbital Shower, $n = 3069$ events) removed due to technical issues.

4.2. Event-level statistical tests

For each treated tap, we compared the distribution of event volumes in the pre-intervention period against the post-intervention period using the Mann–Whitney U test, a non-parametric test suitable for the right-skewed distributions typical of water use data (Fig. 2). Effect sizes were quantified using the rank-biserial correlation (r) and Cohen's d . Significance was assessed at $\alpha = 0.05$. Because 31 simultaneous tests were conducted, we applied both Bonferroni correction and the Benjamini–Hochberg false discovery rate (FDR) procedure to control for multiple comparisons. We report uncorrected, FDR-corrected, and Bonferroni-corrected results to allow assessment of the robustness of findings across correction stringencies. A limitation is that events within the same tap are not independent (temporal autocorrelation, nesting within households). Accordingly, event-level results are treated as exploratory evidence of distributional shifts; we rely on the daily-level DiD (Section 4.4) for more conservative estimates.

4.3. Daily-level aggregation

Event volumes were aggregated to daily totals per tap. Days with zero recorded volume were retained, as excluding them could bias the comparison if the intervention changed usage frequency. Daily volumes for each treated tap were compared pre vs. post using the same Mann–Whitney U test.

4.4. Difference-in-differences estimation

To move beyond simple before–after comparisons, we employ a difference-in-differences (DiD) design at the daily level. Because interventions were not randomly assigned, the DiD estimates should be interpreted as suggestive of causal effects rather than definitive; we therefore refer to these results as quasi-causal throughout. The estimating equation is:

$$Y_{i,t} = \alpha + \beta \cdot \text{Treated}_i \cdot \text{Post}_t + \gamma_t + \varepsilon_{i,t} \quad (1)$$

where Y is the daily volume (L) for tap i on day t , Treated_i is a binary indicator for treated taps, Post_t is a binary indicator for the post-intervention period, γ_t represents time (month) fixed effects that absorb seasonal and other common temporal variation, and β is the DiD estimator. The specification also includes tap fixed effects, which absorb time-invariant level differences between taps, so that the treatment effect is identified from within-tap variation around each tap's intervention date, with never-treated taps of the same use-point type providing the counterfactual time path. Standard errors are clustered at the tap level to account for serial correlation (Ferraro and Price, 2013). We estimate this model both in aggregate (pooling all treated taps) and separately by product category, where controls are restricted to

untreated taps of the same use-point type. Because daily observations are repeated within taps and taps are nested within apartments, we additionally re-estimate the daily-level model as a hierarchical linear mixed-effects model with random intercepts for tap nested within apartment, as a robustness check on within-tap and within-household dependence (Section 5.6). This explicitly addresses concerns about pseudoreplication; we report it alongside, rather than in place of, the cluster-robust DiD, for reasons discussed in Section 5.6.

A key assumption of the DiD design is that treated and control groups would have followed parallel trends in the absence of treatment. To assess this, we conducted a pre-trend analysis by estimating the interaction between a linear time trend and the treatment indicator using only pre-intervention data (April 2023 to May 2024). A statistically significant interaction would indicate differential pre-trends, undermining the DiD identification strategy.

4.5. Survey analysis

Surveys were distributed by email to all residents in apartments that had received product installations in the corresponding round. Participation was voluntary, anonymous, and without incentive. Response rates were 34% (10/29) for Survey 1, which was distributed to all residents; Surveys 2 and 3 were distributed only to the 15 households that had received a product installation by that point, with response rates of 73% (11/15) and 33% (5/15) respectively. Responses on the 5-point Likert scale were summarized as means, medians, and standard deviations by product and survey round. Given the small sample sizes (3–10 responses per product-question combination), we report descriptive statistics without inferential testing. A general attitude question (“Has the product affected your attitude toward water saving?”) was included in all three surveys. Qualitative responses were thematically summarized. We note that non-response bias may be present; residents with strong opinions (positive or negative) may have been more likely to respond. Because survey responses could not be reliably linked to individual apartments, we are unable to compare metered consumption between responding and non-responding households to assess this directly.

5. Results

5.1. Descriptive overview

The final dataset comprises 205,097 events from 80 taps (31 treated, 49 control) over 32 months. Fig. 1 shows the weekly mean daily volume for treated and control taps by use-point type, with installation round dates indicated. The time series reveal substantial temporal variation in all use points, with distinct seasonal patterns and periods of anomalous consumption (likely reflecting tenant turnover or extended absences). The most pronounced excursions—for example the building-wide rise across most taps in spring 2025 and the isolated shower peak in late 2023—coincide with periods of changed occupancy (move-ins/outs and, in the shared facilities, visitors), rather than with any intervention round. Because monitoring was non-intrusive, we cannot attribute individual peaks to specific tenancy events with certainty; the month fixed effects and contemporaneous controls in the difference-in-differences design are intended to absorb such common excursions.

5.2. Event-level results

Fig. 2 shows the distribution of event volumes before and after intervention for treated taps, separated by use-point type. Kitchen and sink events are heavily right-skewed with most events below 5 L, while shower events show a wider distribution extending to 160 L. The visual shift between pre and post distributions is subtle, consistent with the modest effect sizes observed statistically. The pronounced mode of low-volume shower events—which increases in relative frequency after

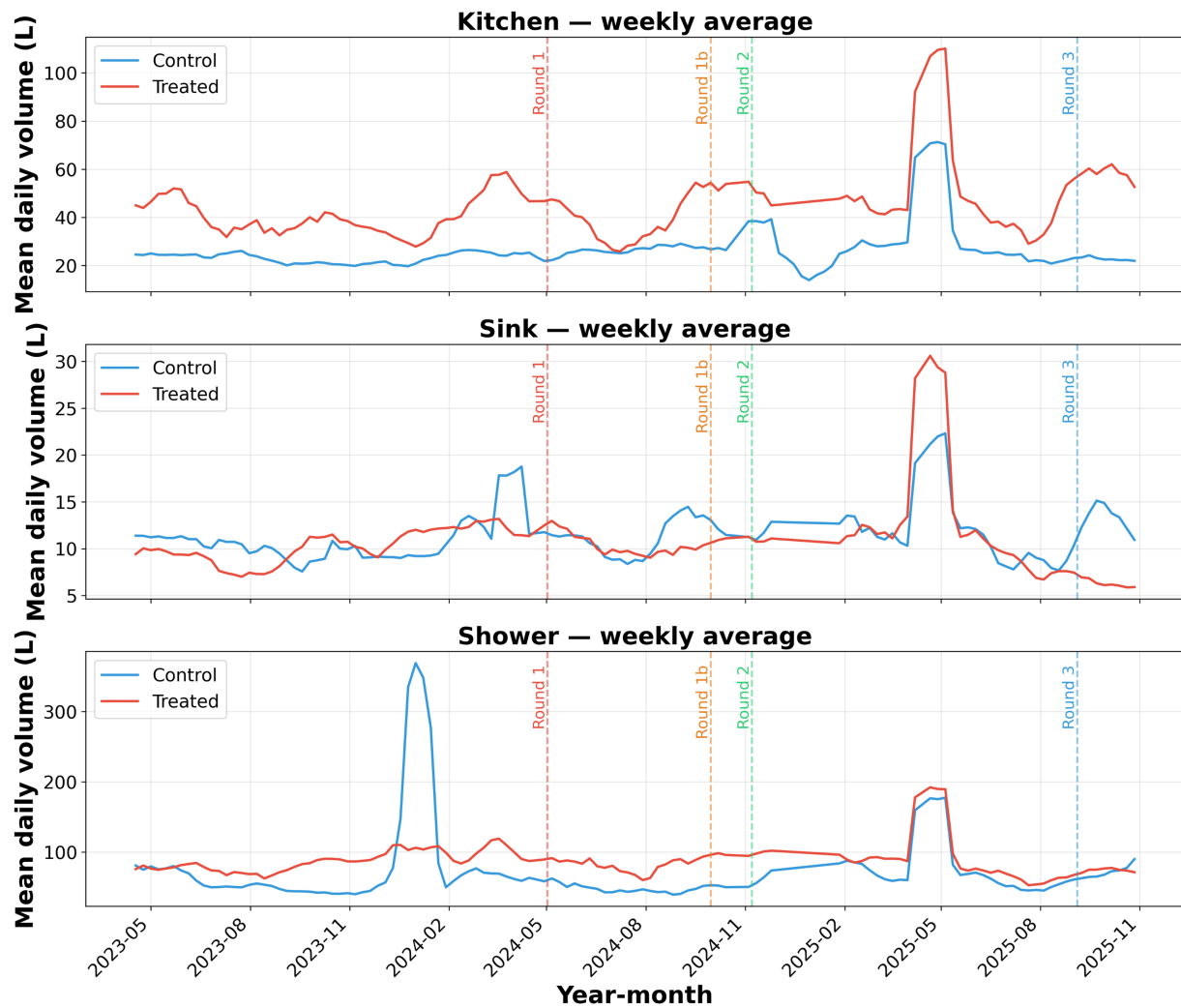


Fig. 1. Weekly mean daily water volume (L/day) for treated (orange) and control (blue) taps, by use-point type: kitchen (top), bathroom sink (middle), and shower (bottom). The x-axis shows year-month. Vertical dashed lines indicate installation round dates. To reduce x-axis crowding, the revised figure thins the year-month tick labels to quarterly intervals and enlarges all axis labels.

Event volume distributions before vs after (treated taps)

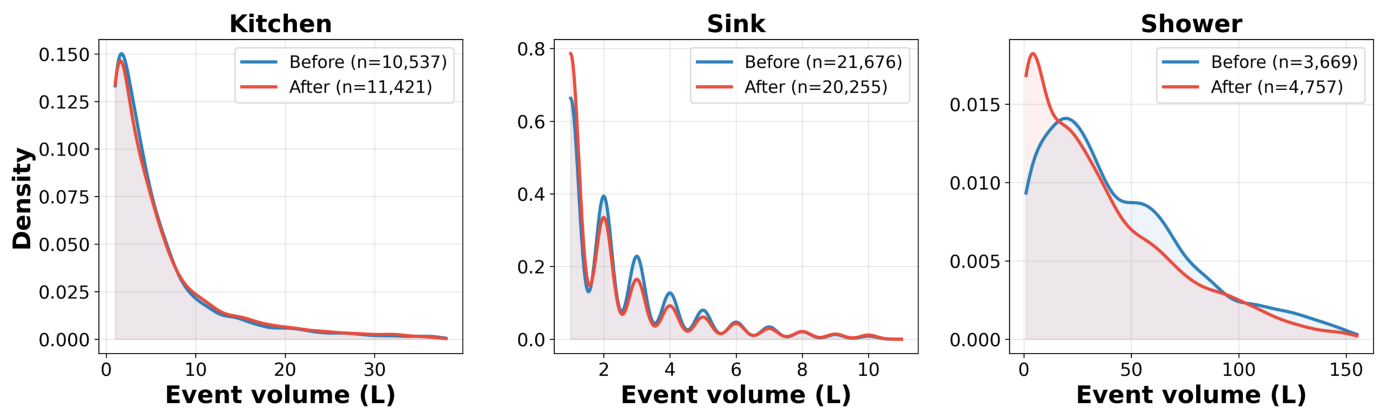


Fig. 2. Density distributions of event volumes (L) for treated taps before (blue) and after (red) intervention, by use-point type. Semi-transparent overlapping histograms show distribution overlap considerably. Sample sizes are indicated in the legends. Note the different x-axis scales. For clarity, the revised figure plots the pre- and post-intervention distributions as outlined kernel-density curves rather than filled overlapping histograms, so that the two distributions can be distinguished where they overlap.

intervention—most plausibly reflects a mix of very short showers and 10-min reporting windows that capture only the start or tail of a shower at low recorded volume, rather than a change in flow rate; with non-intrusive metering we cannot fully separate short-duration use from low instantaneous flow.

As noted in Section 4.1, an event under our ~10-min reporting protocol is best understood as an active-use window rather than a single discrete use; the figures below should be read with that caveat, and with the tap-clustered daily-level difference-in-differences (Section 5.6) as the confirmatory analysis. Of the 31 treated taps, 25 (81%) showed statistically significant differences in event volume distributions at the uncorrected $p < 0.05$ level. After applying multiple testing corrections, 23 taps (74%) remained significant under Benjamini–Hochberg FDR control, and 20 taps (65%) under the more conservative Bonferroni correction. The robustness of results across correction methods indicates that the majority of detected effects are unlikely to be false positives. Importantly, while many taps showed statistically significant changes, most effect sizes were small to moderate: the median percentage change in mean event volume was -7.1% , with a median rank-biserial correlation of 0.12, indicating modest practical differences. Individual taps ranged from -56% to $+174\%$, reflecting considerable heterogeneity. Fig. 3 summarizes the product-level event effects as a forest plot, showing that sensor taps and shower monitors tended toward reductions, while cold start taps and eco-push taps showed increases.

Effect sizes (rank-biserial r) were generally small to moderate, with $|r|$ ranging from 0.003 to 0.53. The largest effect sizes were observed for sensor taps in bathrooms ($|r|$ up to 0.53) and for one shower monitor installation where volumes decreased by 47% ($r = -0.49$). However, other shower monitor installations showed non-significant changes, highlighting substantial heterogeneity across apartments.

5.3. Daily-level results

At the daily aggregation level, 27 of 31 treated taps (87%) showed statistically significant changes (see Supplementary Fig. S3 for daily

volume distributions by product category). Aggregating events to daily totals removes the within-day dependence that complicates the event-level tests while retaining tap-level resolution. As Supplementary Fig. S3 shows, the before-versus-after daily distributions overlap substantially for most product categories, with the largest visible shifts for the shower monitor and sensor taps and little change for the cold start and eco-push taps—mirroring the heterogeneity seen in the event-level and difference-in-differences results and reinforcing that statistically detectable daily changes are generally modest in magnitude.

5.4. Pre-trend analysis

The pre-intervention parallel trends analysis (Fig. S1) shows that treated and control groups followed broadly similar trajectories across all three use-point types during the 13 months before the first intervention (April 2023 to May 2024). The mean pre-period duration was 395 days for Round 1 taps, 297 days for Round 2 taps, and 214 days for Round 3 taps. Formal testing of the time $\times$ treatment interaction in the pre-period yielded non-significant results for all use points: kitchen ($p = 0.64$), sink ($p = 0.73$), and shower ($p = 0.72$). These results support the parallel trends assumption underlying the DiD design, though we acknowledge that the absence of statistical significance does not definitively prove the assumption holds.

5.5. Difference-in-differences results

The aggregate DiD estimate, pooling all 31 treated taps against all 49 controls with time fixed effects, was not statistically significant. Fig. 4 shows the aggregate DiD effects by use point and outcome measure, with wide confidence intervals for daily volume and shower-level outcomes.

Product-specific DiD estimates (Table 2) reveal meaningful heterogeneity. Two results warrant particular discussion: the spray kitchen tap showed a DiD reduction of 11.8 L/day relative to controls, while the auto water diverter was associated with a DiD reduction of 29.6 L/day at the shower tap—the largest effect observed (see Section 5.5).

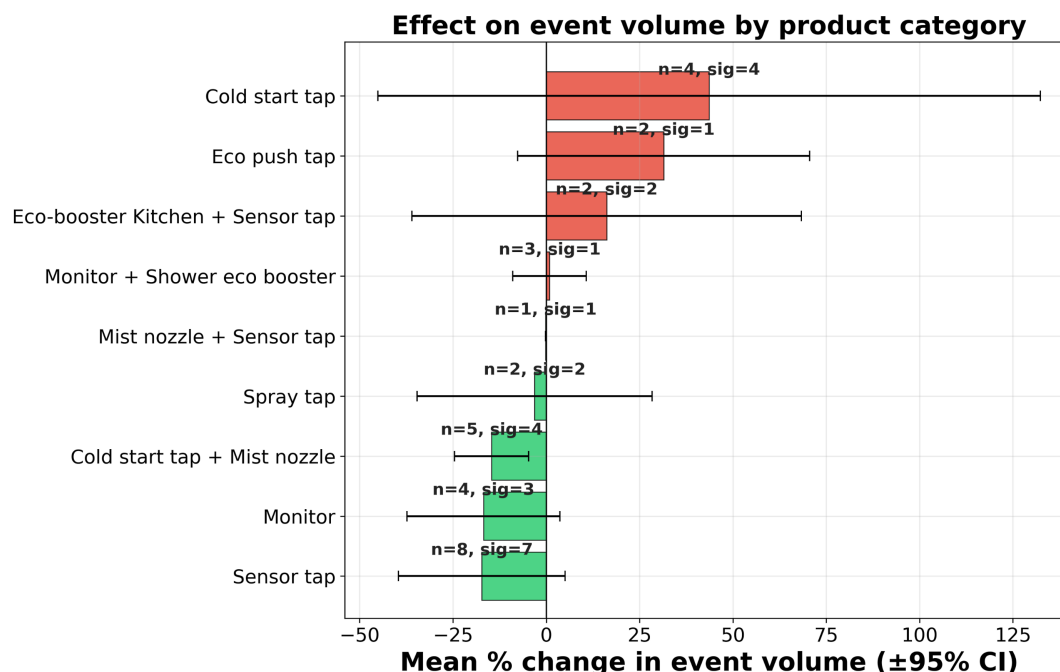


Fig. 3. Forest plot of mean percentage change in event volume by product category (treated taps). Bars indicate reductions (left of zero) or increases (right of zero). Error bars show 95% confidence intervals across taps within each product. Labels show number of taps (n) and number significant at $p < 0.05$ (sig). Here “sig” denotes the number of treated taps within the product category whose pre-versus post-intervention event-volume distribution differs significantly at $p < 0.05$ (Mann–Whitney U test); the n and sig labels are positioned above the corresponding error bars so as not to overlap them. Axis labels have been enlarged and set in bold for legibility. Note that these event-level tests are exploratory (Section 4.2).

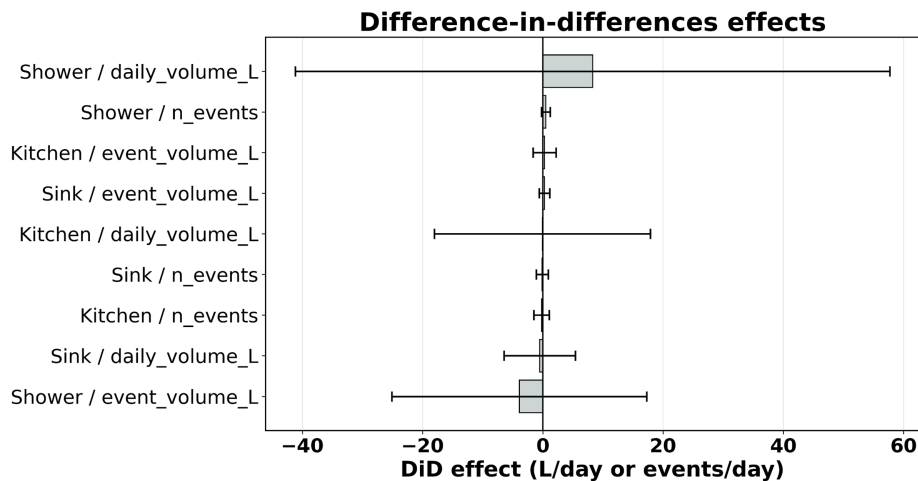


Fig. 4. Aggregate difference-in-differences effects by use point and outcome measure. Bar length indicates the DiD coefficient (L or count); error bars show 95% confidence intervals. No aggregate estimate is statistically significant at $p < 0.05$.

Table 2

Product-specific difference-in-differences estimates for daily water volume (L/day). Δ treated and Δ control show the percentage change from pre to post for treated and control groups respectively. DiD is the difference between treated and control changes in absolute terms. These are raw pre/post mean differences by product; they are independent of the covariate-adjusted, cluster-robust regression estimates reported in Table 4.

Product	Location	n	DiD (L/d)	Δ treated (%)	Δ ctrl (%)	n ctrl
Spray tap	Kitchen	2	-11.8	-24.8	+18.3	27
Auto water diverter	Shower	6	-29.6	-15.5	+23.0	5
Sensor tap	Sink	9	-3.9	-17.2	+15.0	17
Eco-push tap	Sink	2	-2.6	-7.9	+11.0	17
Cold start tap	Sink	9	-0.1	+16.9	+15.0	17
Shower monitor	Shower	7	+5.7	-6.9	-15.8	5
Eco-booster	Shower	3	-2.1	-0.3	+2.6	5
Kitchen tap (proto.)	Kitchen	4	+14.2	+16.3	-21.8	27
Spray/sensor tap	Kitchen	2	+9.3	+21.2	+18.3	27
Sensor tap + eco-booster	Kitchen	2	+11.5	+18.5	+4.7	27
Mist nozzle	Sink	6	-2.5	-14.3	+13.2	17

Table 3

Summary of user satisfaction scores by product (Likert scale 1–5). Values are median (mean $\pm$ SD). — indicates the question was not included in that survey round.

Product	Satisfied	Functions	Saves water	Surv.	n
Shower monitor	5 (4.5 $\pm$ 1.1)	5 (4.2 $\pm$ 1.1)	—	1	8
Spray tap	4 (3.6 $\pm$ 0.9)	4 (4.2 $\pm$ 0.8)	—	1	5
Sensor tap (kitchen)	4 (2.8 $\pm$ 1.6)	4 (3.3 $\pm$ 1.2)	—	1	5
Sensor tap (sink)	3 (2.8 $\pm$ 1.3)	3 (2.8 $\pm$ 1.1)	—	1	6
Mist nozzle	3 (2.4 $\pm$ 1.2)	3 (2.6 $\pm$ 0.9)	—	1	8
Eco-booster	4 (3.0 $\pm$ 1.4)	4 (3.6 $\pm$ 1.3)	3 (3.3 $\pm$ 1.3)	2	5–9
Eco-booster (kitchen)	3 (3.0 $\pm$ 1.4)	3 (3.2 $\pm$ 0.8)	3 (3.2 $\pm$ 1.0)	2	8–10
Eco-push tap	4 (3.0 $\pm$ 1.7)	3 (3.0 $\pm$ 0.8)	3 (3.0 $\pm$ 1.6)	2	3–4
Auto water div.	1 (1.3 $\pm$ 0.6)	1 (1.5 $\pm$ 1.0)	1 (1.5 $\pm$ 1.0)	3	3–4
Kitchen tap (proto.)	4–4.5	n/a	3 (3.0)	3	1–4

5.6. Hierarchical robustness check (mixed-effects model)

To verify that the daily-level findings are not artefacts of treating repeated daily observations from the same tap as independent, we re-estimated the difference-in-differences model as a hierarchical linear mixed-effects model with random intercepts for tap nested within apartment, fitted on the same daily panel and using the same year-month fixed effects (Table 4). The difference-in-differences point estimates are essentially unchanged: the pooled aggregate effect is -1.5 L/day under the cluster-robust DiD and -1.4 L/day under the mixed-effects model, and every per-use-point estimate retains its sign and order of magnitude. The substantive conclusions—modest, heterogeneous device effects and no robust aggregate reduction—are therefore preserved when within-tap and within-household dependence is

modelled explicitly.

Notably, the random-intercept standard errors are uniformly smaller, not larger, than those of the tap-clustered DiD—for the pooled daily-volume effect, 0.83 against 3.83 L/day. The point estimates are similar across specifications, but inference is sensitive to the assumed within-tap covariance structure: the shower events-per-day effect is far from significant under clustering ($p = 0.18$) yet would be reported as highly significant under the mixed model ($p < 0.001$), and the same reversal occurs at the kitchen taps ($p = 0.74$ against $p = 0.041$). This is expected, because a random-intercept structure absorbs differences in mean level between taps but does not account for residual day-to-day serial correlation within a tap, whereas the cluster-robust estimator makes no parametric assumption about the within-tap covariance and is robust to exactly this dependence. We therefore retain the tap-clustered DiD as

Table 4

Difference-in-differences estimates from the cluster-robust two-way fixed-effects specification (primary) and the hierarchical mixed-effects model with random intercepts for tap nested within apartment (robustness check). Point estimates are stable across specifications; the mixed-model standard errors are smaller because the random-intercept structure does not absorb within-tap day-to-day serial correlation, which the cluster-robust estimator accommodates. Because each apartment contributes at most one tap per use point, the apartment and tap levels coincide within each use-point model, and the nested random-effects structure is identified only in the pooled specification, where it is used. T = treated taps, C = control taps.

Use point/outcome	Taps (T/C)	DiD, cluster-robust (L/day)	p	DiD, mixed (L/day)	p
Kitchen — daily volume	5/27	−0.07	0.99	−0.07	0.96
Kitchen — mean event volume	5/27	+0.30	0.76	+0.41	0.35
Kitchen — events/day	5/27	−0.22	0.74	−0.21	0.041
Sink — daily volume	19/17	−0.54	0.86	−0.54	0.26
Sink — mean event volume	19/17	+0.27	0.55	+0.20	0.33
Sink — events/day	19/17	−0.10	0.84	−0.10	0.11
Shower — daily volume	7/5	+8.28	0.74	+8.74	0.051
Shower — mean event volume	7/5	−3.91	0.72	−3.91	0.18
Shower — events/day	7/5	+0.50	0.18	+0.50	<0.001
Pooled (all taps) — daily volume	31/49	−1.53	0.69	−1.43	0.086

our primary specification and report the mixed-effects model as a robustness check rather than as confirmation. Cluster-robust inference is asymptotic, however, and the number of treated clusters here is small—five treated kitchen taps and seven shower taps—so we subjected the daily-volume estimates to three further checks: jackknife (CRV3) standard errors, a wild cluster bootstrap imposing the null (Rademacher weights, 9999 replications), and re-clustering the pooled model at the apartment rather than the tap level. All three leave the estimates far from conventional significance. With as few as five treated clusters, bootstrap inference can itself be unreliable (MacKinnon and Webb, 2018); we therefore rely on it only to show that the null result is not marginal. The bootstrap p-values for the pooled, kitchen, sink and shower daily-volume effects are 0.69, 0.99, 0.87 and 0.82; CRV3 standard errors are 10–32% larger than their CRV1 counterparts without altering any conclusion; and apartment-level clustering changes the pooled standard error from 3.83 to 3.76 L/day. We do not claim that tap-level clustering is adequate in general, only that the aggregate null

reported here is not an artefact of the assumed dependence structure.

5.7. Auto water diverter (exploratory)

The auto water diverter yielded the largest nominal DiD estimate (−29.6 L/day), but this result should be interpreted with caution. The device captures cold start-up water for reuse; it does not reduce the volume passing through the shower meter. The observed effect may reflect behavioral awareness (Hawthorne), confounding temporal trends in the small control group ($n = 5$, which increased consumption by +23% in the post-period), or tenant turnover. A detailed discussion of candidate explanations is provided in [Supplementary Section S2](#).

5.8. Tap-level heterogeneity

[Supplementary Fig. S4](#) illustrates the heterogeneity at the individual tap level. Sink taps (baseline ≤ 5 L) show the widest range of responses

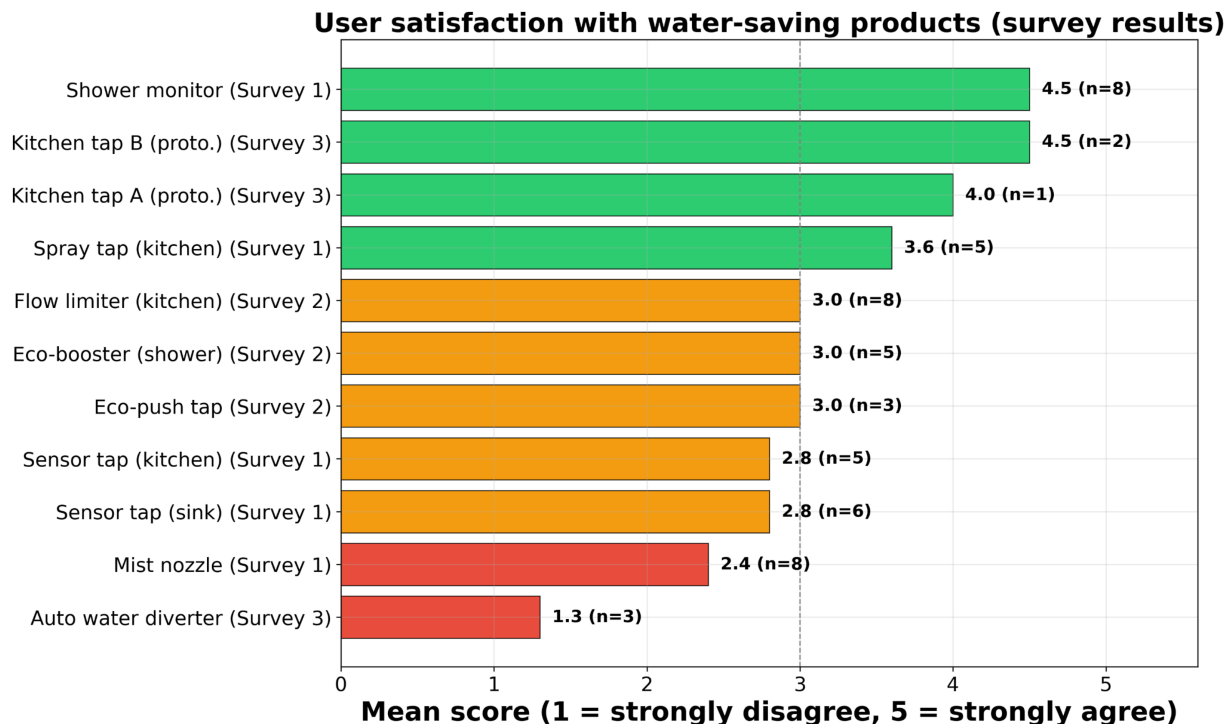


Fig. 5. Mean user satisfaction scores by product at different locations (kitchen/bathroom/shower) (Likert scale 1–5, where 1 = strongly disagree, 5 = strongly agree). Colors indicate satisfaction level (yellow ≥ 3.5 , orange 2.5–3.5, red < 2.5). Sample sizes shown in parentheses.

(−56% to +174%), shower taps cluster more tightly, and there is no clear relationship between baseline volume and percentage change, suggesting effects are not simply regression to the mean.

5.9. Survey results

Survey responses reveal a clear hierarchy of user satisfaction (Fig. 5). The shower monitor received the highest satisfaction scores (median 5/5 for satisfaction, 5/5 for functionality in Survey 1), with residents appreciating the real-time feedback without any change to their shower experience. In contrast, the auto water diverter received the lowest scores (median 1/5 for satisfaction, 1/5 for functionality in Survey 3), with residents reporting the bucket system as impractical and intrusive.

Table 3 summarizes the quantitative survey results. Across all three survey rounds, the general attitude question showed modest scores (medians of 3, 2, and 2 on the 1–5 scale), suggesting that the products had limited impact on residents' self-reported water-saving awareness.

A suggestive but exploratory pattern is the apparent disconnect between measured consumption effects and user satisfaction. The shower monitor, which received the highest user ratings, showed no consistent DiD effect on water volumes. Meanwhile, the auto water diverter, with the lowest user satisfaction scores, was associated with the largest (though potentially confounded) DiD reduction. This suggests that user experience and measured effectiveness may be driven by different factors. Given the small number of respondents per product (3–10) and response rates of 33–73%, this pattern is presented as hypothesis-generating rather than definitive, and should be confirmed in larger samples.

6. Discussion

6.1. Methodological contributions

This study suggests that event-level monitoring offers important advantages for evaluating water-saving interventions. Compared to daily or monthly data, event-level analysis revealed that (1) statistically significant changes were present at 65–81% of treated taps (depending on multiple testing correction), even when the aggregate DiD estimate was null; (2) the direction and magnitude of effects varied substantially across products and individual taps; and (3) most effect sizes were small to moderate (median change −7.1%, median rank-biserial correlation 0.12), consistent with the challenge of achieving marginal savings in an already-efficient building. These findings extend the growing body of work on smart water metering (Cominola et al., 2015) by suggesting that event-level granularity enables qualitatively different inferences than aggregate data. However, a trade-off exists between the high statistical sensitivity of event-level tests—which, due to large sample sizes per tap, can detect very small distributional shifts—and the practical (policy) relevance of those shifts. The DiD framework, which operates at the daily level with far fewer degrees of freedom, provides more conservative and arguably more policy-relevant effect estimates. The high rate of event-level significance (65–81%) should therefore be interpreted as exploratory evidence of within-tap distributional changes, not as confirmation of large or economically meaningful water savings.

The product-specific DiD—comparing treated taps only to untreated taps of the same type—provided more interpretable estimates than the fully pooled model, but for products with few treated or control taps, estimates should be treated as suggestive rather than definitive.

6.2. External validity and baseline efficiency

HSB Living Lab was already equipped with water-efficient fixtures (cold-start taps, low-flow aerators), which constrains interpretation: marginal savings from additional devices are inherently limited, the control group also had efficient fixtures (making the DiD comparison

conservative), and null results should not be taken as evidence of general product ineffectiveness. Our results therefore represent a conservative lower bound on achievable savings in buildings with conventional fixtures.

6.3. User experience versus measured effectiveness

The disconnect between user satisfaction and measured consumption effects is a suggestive, exploratory pattern that warrants confirmation in larger samples. Products that scored highest with users (shower monitor, spray tap) generally showed modest or inconsistent consumption effects, while the product with the lowest user satisfaction (auto water diverter) was associated with the largest DiD estimate. This pattern has practical implications for utilities and housing providers: optimizing for user satisfaction could, if this pattern proves robust, select different products than optimizing for consumption reduction.

Viewed through the lens of the Technology Acceptance Model (Davis, 1989), user satisfaction reflects perceived ease of use and perceived usefulness—dimensions that may be largely orthogonal to volumetric savings. Products that require minimal behavioral adjustment (aerators, monitors) score high on perceived ease of use but may achieve savings through mechanisms that users do not notice, whereas products that impose visible inconvenience (auto water diverter) face adoption barriers regardless of objective effectiveness (Rogers, 2003). This framing suggests that long-term adoption, and hence sustained savings, may depend more on user acceptance than on short-term volumetric performance.

We caution, however, against interpreting this disconnect as established evidence of a genuine trade-off between user preference and effectiveness, because several measurement-related explanations are equally consistent with the data. The shower monitor may reduce shower duration rather than flow rate, a channel that our volume- and event-based metrics capture only partially given the 1-liter resolution and ~10-min reporting interval; a real saving achieved through shorter showers could therefore go undetected. Conversely, the apparent effect of the auto water diverter is most plausibly an artefact of its small control group and confounding rather than a true reduction (Section 5.7). The observed independence between satisfaction and measured savings may thus reflect the limits of volumetric detection as much as any underlying orthogonality, and we present it as a hypothesis to be tested with flow-rate- and duration-resolved metering rather than as an established finding.

Interpreted more concretely through the Technology Acceptance Model and diffusion-of-innovations theory (Davis, 1989; Rogers, 2003), the product ranking maps onto distinct adoption constructs: the shower monitor scores high on perceived usefulness and observability because it makes consumption visible without altering the showering experience; aerator- and sensor-type devices score high on perceived ease of use and low on required behavioral change, easing adoption but limiting their behavioral leverage; and the auto water diverter, despite a high notional relative advantage, scores poorly on ease of use, compatibility and trialability because it imposes a visible bucket-handling routine, predicting weak sustained adoption irrespective of its measured effect. This mapping implies that durable building-scale savings are likely to come from devices that combine low behavioral cost with observable feedback, a hypothesis that the present cross-sectional satisfaction data can motivate but not confirm.

Well-liked products may produce sustained effects through long-term behavioral change, while disliked ones risk removal; longitudinal follow-up is needed to test this.

Beyond the pre-existing efficient fixtures discussed above, the representativeness of HSB Living Lab's tenant population warrants consideration. Tenants are self-selected through HSB's standard allocation process but must agree to participate in a research environment, which may attract individuals who are younger, more environmentally

aware, or more tolerant of experimental installations than the general population. Average tenure is 2–3 years, which is shorter than typical Swedish rental housing. Several additional structural features further distinguish this setting from mainstream housing. First, the kitchens and showers on two of the four floors are shared between multiple residents, creating a communal usage dynamic that may dilute individual accountability for water consumption. Second, water costs are included in the rent, so residents have no direct price signal incentivizing conservation—a factor that may dampen responsiveness to efficiency devices relative to settings with volumetric billing. Third, the tenant population is predominantly young (20–35 years), transient, and affiliated with Chalmers, which constitutes a demographically narrow sample. Taken together, these factors mean that HSB Living Lab represents a structurally distinct demand context—not only a conservative baseline—and caution is warranted when extrapolating to owner-occupied, family, or individually metered housing. (Over the 32-month monitoring window, the short lease terms imply that a substantial share of the monitored apartments changed occupants at least once; exact tenancy-change counts were not available to us, so turnover is characterized qualitatively and treated as a potential time-varying confounder that the difference-in-differences design (with month fixed effects and contemporaneous controls) can only partially absorb.)

For utilities and policy-makers, the key implication is that our results likely represent a conservative lower bound on achievable savings: buildings with conventional fixtures and less environmentally aware occupants may show larger absolute effects. However, the qualitative finding—that user satisfaction varies dramatically across product types—is likely generalizable and should inform product selection strategies regardless of building context.

6.4. Comparison with previous studies

Our event-level median reduction of 7.1% for treated taps is at the lower end of the 5–30% range reported in the literature (Inman and Jeffrey, 2006; Addo et al., 2018), consistent with the already-efficient baseline hypothesis. The null aggregate DiD result echoes large-scale trials where technical devices alone had modest effects (Ferraro and Price, 2013). The disconnect between user perception and measured effects parallels findings in energy conservation (Allcott, 2011).

6.5. Limitations

Three main limitations warrant discussion. First, sample sizes are small, particularly for some products ($n = 1\text{--}2$ treated taps), and the control group for shower products ($n = 5$) is especially limited, which may contribute to wide confidence intervals and reduces generalizability. Second, the quasi-experimental design—products were not randomly assigned to apartments—introduces potential selection bias, and while the DiD design mitigates time-invariant confounding, it cannot address time-varying confounders correlated with treatment assignment. Because products were allocated on the basis of compatibility and availability rather than at random, treated and control apartments may also differ systematically in unobserved occupant or fixture characteristics. The matched same-use-point controls and the pre-trend tests (Section 5.4; all $p > 0.6$) mitigate this concern but cannot eliminate it, and residual selection bias together with unobserved time-varying confounders correlated with treatment timing remain threats to internal validity. We therefore interpret all difference-in-differences estimates as quasi-causal rather than definitive. Third, the 1-liter volumetric resolution and ~10-min reporting intervals constrain the precision with which very small or very brief events can be captured.

Additional limitations include the following. The pre-existing water-efficient fixtures limit external validity to similar high-efficiency buildings. Tenant turnover during the monitoring period may have introduced uncontrolled variation in occupancy-related consumption—a time-varying confounder that the DiD framework cannot

fully address. We were unable to quantify it. The number and timing of tenancy changes cannot be recovered from meter data, and we did not have access to the lease records for the monitoring period, so we cannot report how many changes occurred, whether they were balanced between treated and control apartments, or whether any coincided with an installation round. The sensitivity analyses this would permit—excluding apartments in which a tenancy changed, conditioning on occupancy spell, or splitting each series into separate household episodes—therefore remain open, and we present this as an unresolved limitation rather than one the present design overcomes. A random-intercept structure does not address it either, since turnover is a time-varying change within an apartment rather than a fixed difference in level between apartments.

The event-level Mann–Whitney tests treat individual events as independent observations; however, events within the same tap are serially correlated and nested within households, which may inflate the effective sample size and the proportion of statistically significant results. The tap-level comparison framework and multiple testing corrections partly mitigate this concern, and we additionally implemented a hierarchical (mixed-effects) model with random intercepts for tap nested within apartment as a robustness check (Section 5.6). The difference-in-differences point estimates were essentially unchanged, confirming that the conclusions are not artefacts of within-tap dependence; we note further that the cluster-robust standard errors used throughout are, if anything, more conservative than the random-intercept alternative.

The 1-liter volumetric resolution means that the dataset is biased toward larger events: any measured intervention effects are conditional on this detection threshold, and products that primarily reduce very small events may appear less effective than they are. Furthermore, the monitoring system cannot capture water savings that are realized outside the metered flow path—notably the auto water diverter, whose intended benefit (reuse of captured cold start-up water) occurs downstream of the shower meter. Survey response rates, while adequate for qualitative insights, are too small for robust statistical inference.

In addition, because each installation was accompanied by informational materials (Section 3.2), the estimated effects reflect the combined device-plus-information intervention; although building-wide awareness is largely absorbed by the contemporaneous control comparison, a localized information or novelty (Hawthorne) response among treated residents cannot be fully separated from the effect of the device itself. Hot and cold water volumes were recorded separately at each tap; all volumetric analyses in this paper are based on total (hot + cold) volumes unless otherwise noted.

Despite these limitations, the living lab setting enabled continuous tap-level monitoring over 32 months—a level of resolution and duration rarely achieved in occupied residential buildings. This methodological contribution offsets some of the constraints imposed by the small sample and quasi-experimental design.

7. Conclusions

This study introduced an event-level monitoring framework for evaluating water-saving interventions and applied it to a 32-month dataset from HSB Living Lab, encompassing 205,097 events, 31 treated taps, 49 controls, and 26 survey respondents. The key conclusions are as follows.

First, event-level analysis at tap-level resolution substantially enriches the evaluation of water-saving devices. We found that 74–81% of treated taps (depending on multiple testing correction) showed statistically significant changes in event volume distributions, even when the aggregate DiD estimate was not significant. This suggests that event-level data can detect intervention effects that may be obscured at higher aggregation levels. Because the aggregate difference-in-differences estimate was not statistically significant, however, these event-level shifts should be interpreted as changes in the distribution of

individual water-use events rather than as demonstrated reductions in total water use at the building scale.

Second, product-specific effects are highly heterogeneous. Spray kitchen taps and sensor bathroom taps showed consistent reductions, while kitchen sensor taps and cold start taps produced mixed results. A nominally large DiD estimate was observed for the auto water diverter (-29.6 L/day), but this result should not be interpreted as causal: the device's intended mechanism (capturing cold start-up water for reuse) does not directly reduce measured shower flow, the control group was small ($n=5$), and multiple plausible confounders (Hawthorne effects, tenant turnover) cannot be ruled out. We therefore report this estimate only as an exploratory case (Section 5.7) and do not treat it as a headline finding.

Third, user experience does not correlate with measured consumption effects. The most-liked product (shower monitor) had no consistent DiD effect, while the least-liked product (auto water diverter) was associated with the largest reduction. This has implications for program design: both user-centered and data-driven perspectives are needed when selecting water-saving technologies for deployment.

Fourth, pre-existing building efficiency matters. Testing water-saving devices in an already-efficient building provides conservative estimates that should not be generalized without adjustment to buildings with conventional fixtures.

Fifth, a DiD framework with time fixed effects and clustered standard errors strengthens quasi-causal interpretation compared to simple before–after comparisons in building-scale water intervention studies, although non-random assignment and small samples mean that definitive causal claims remain beyond the reach of this design. Pre-trend analysis should be included as standard practice to verify the parallel trends assumption. We recommend this approach, combined with explicit acknowledgement of remaining threats to internal validity, for future evaluations.

Future work should extend the framework to buildings with conventional fixtures, incorporate longer follow-up periods to assess behavioral persistence, and use structured interviews to understand mechanisms behind unexpected findings such as the auto water diverter result.

CRediT authorship contribution statement

Jesper Knutsson: Writing – review & editing, Writing – original draft, Visualization, Investigation, Funding acquisition, Formal analysis, Data curation. **Carolina Hiller:** Writing – review & editing, Conceptualization. **Josefine Klingberg:** Writing – review & editing, Funding acquisition, Conceptualization. **Sara Renström:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization. **Tobias Svanberg:** Writing – review & editing, Resources, Conceptualization.

Data and code availability

Code implementing the corrected difference-in-differences and hierarchical mixed-effects analyses (Sections 4.4, 5.6) is available at [10.5281/zenodo.22278790](https://doi.org/10.5281/zenodo.22278790) under the MIT license. The underlying tap-level water-use data are not publicly available due to resident privacy.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work the author(s) used Claude (Anthropic) in order to assist with manuscript revision, statistical interpretation, and preparation of submission documents. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tobias Svanberg reports a relationship with IKEA of Sweden AB that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was carried out as part of the DROPIT project (Drop-in ICT solutions for water savings), funded by the Swedish Research Council for Sustainable Development (Formas), grant number 2023-01988. The project is a collaboration between Chalmers University of Technology, RISE Research Institutes of Sweden, and IKEA of Sweden AB. We thank the residents of HSB Living Lab for their participation and patience with the numerous product installations. We gratefully acknowledge the HSB Living Lab infrastructure and management team for facilitating data collection and product installations. The water-saving products were provided by IKEA of Sweden.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.resenv.2026.100381>.

References

- Addo, I.B., Thoms, M.C., Parsons, M., 2018. Household water use and conservation behavior: a meta-analysis. *Water Resour. Res.* 54 (10). <https://doi.org/10.1029/2018WR023306>.
- Allcott, H., 2011. Social norms and energy conservation. *J. Publ. Econ.* 95 (11–12), 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>.
- Beal, C.D., Stewart, R.A., 2014. Identifying residential water end uses underpinning peak day and peak hour demand. *J. Water Resour. Plann. Manag.* 140 (7). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000357](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000357).
- Beal, C.D., Stewart, R.A., Sharma, A.K., 2011. South East Queensland residential end-use study: final report. Urban Water Security Research Alliance Technical Report No. 47. Queensland Government, Brisbane.
- Bergvall-Kärebörn, B., Ihlström Eriksson, C., Ståhlbröst, A., 2015. Places and spaces within living labs. *Technol. Innov. Manag. Rev.* <https://doi.org/10.22215/timreview/951>.
- Brent, D.A., Cook, J.H., Olsen, S., 2015. Social comparisons, household water use, and participation in utility conservation programs: evidence from three randomized trials. *J. Assoc. Environ. Resour. Econ.* 2 (4), 597–627. <https://doi.org/10.1086/683427>.
- Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E., 2015. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: a review. *Environ. Model. Software* 72, 198–214. <https://doi.org/10.1016/j.envsoft.2015.07.012>.
- Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E., 2019. Data mining to uncover heterogeneous water use behaviors from smart meter data. *Water Resour. Res.* 55 (9). <https://doi.org/10.1029/2019WR024897>.
- Daminato, C., Diaz-Farina, E., Filippini, M., Padrón-Fumero, N., 2021. The impact of smart meters on residential water consumption: evidence from a natural experiment in the Canary Islands. *Resour. Energy Econ.* 64, 101231. <https://doi.org/10.1016/j.reseneeco.2021.101231>.
- Datta, S., Darling, M., Lorenzana, K., 2021. Toward achieving persistent behavior change in household water conservation. *Proc. Natl. Acad. Sci.* 118 (25). <https://doi.org/10.1073/pnas.2023014118>.
- Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13 (3), 319–340. <https://doi.org/10.2307/249008>. <https://misq.umn.edu/misq/article/13/3/319/191/Perceived-Usefulness-Perceived-Ease-of-Use-and>. (Accessed 1 September 1989).
- Ehret, P.J., Hodges, H.E., Kuehl, C., Brick, C., Mueller, S., Anderson, S.E., 2021. Systematic review of household water conservation interventions using the information–motivation–behavioral skills model. *Environ. Behav.* <https://doi.org/10.1177/0013916519896868>.
- Ferraro, P.J., Price, M.K., 2013. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Rev. Econ. Stat.* 95 (1), 64–73. https://doi.org/10.1162/REST_a_00283.
- Fielding, K.S., Spinks, A., Russell, S., McCrea, R., Stewart, R., Gardner, J., 2013. An experimental test of voluntary strategies to promote urban water demand management. *J. Environ. Manag.* 114, 343–351. <https://doi.org/10.1016/j.jenvman.2012.10.027>.

- Inman, D., Jeffrey, P., 2006. A review of residential water conservation tool performance and influences on implementation effectiveness. *Urban Water J.* 3 (3), 127–143.
- Jaime Torres, M.M., Carlsson, F., 2018. Direct and spillover effects of a social information campaign on residential water-savings. *J. Environ. Econ. Manag.* 92, 222–243. <https://doi.org/10.1016/j.jeem.2018.08.005>.
- Jessoe, K., Rapson, D., 2014. Knowledge is (less) power: experimental evidence from residential energy use. *Am. Econ. Rev.* 104 (4), 1417–1438. <https://doi.org/10.1257/aer.104.4.1417>.
- Knutsson, J., Knutsson, P., 2020. Water and energy savings from greywater reuse: a modeling scheme using disaggregated consumption data. *International Journal of Energy and Water Resources* 4, 17–30. <https://doi.org/10.1007/s42108-020-00096-z>.
- Knutsson, J., Marx, C., 2016. The habitation lab as a platform for sustainable innovations in the built environment: a case study. In: *Green Buildings Technologies and Materials*. Vilnius Gediminas Technical University.
- MacKinnon, J.G., Webb, M.D., 2018. The wild bootstrap for few (treated) clusters. *Econom. J.* 21 (2), 114–135. <https://doi.org/10.1111/ectj.12107>.
- Nguyen, K.A., Stewart, R.A., Zhang, H., 2014. An autonomous and intelligent expert system for residential water end-use classification. *Expert Syst. Appl.* 41 (2), 342–356. <https://doi.org/10.1016/j.eswa.2013.07.049>.
- Rogers, E.M., 2003. *Diffusion of Innovations*, 5th ed. Free Press.
- Sönderlund, A.L., Smith, J.R., Hutton, C.J., Kapelan, Z., 2016. Effectiveness of smart meter-based consumption feedback in curbing household water use: knowns and unknowns. *J. Water Resour. Plann. Manag.* 142 (12), 04016060. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000703](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000703).