Supporting Information for

Embedding Usage Sensors in Point-of-Use Water Treatment Devices: Sensor Design and Application in Limpopo, South Africa

David D. Meyer,[*,†,§] Courtney Hill,[‡] Kelly McCain,[‡] James A. Smith,[‡] Pascal O. Bessong,[‡,¶] Elizabeth T. Rogawski McQuade,[‡] and Natasha C. Wright[*,†,‖]

†*Massachusetts Institute of Technology, Cambridge, MA 02139 USA*

‡*University of Virginia, Charlottesville, VA 22904 USA*

¶*University of Venda, Limpopo Province, Thohoyandou, 0950 South Africa*

§*Current Address: University of Toronto, Toronto, ON M5S 0C9 Canada*

‖*Current Address: University of Minnesota, Minneapolis, MN 55419 USA*

E-mail: david.meyer@utoronto.ca; wrigh677@umn.edu

This file has 14 pages and contains the following supplementary information:

- Extended Methods.
- Figure S1: Data processing and noise classification method and example.
- Figure S2: Data compression method.
- Figure S3: Noise sensitivity of group usage trends.
- Figure S4: Noise sensitivity of sensed and surveyed prevalence of use.
- Figure S5: Noise sensitivity of silver versus use prevalence model.
- Figure S6: Noise and long-entry sensitivity of time of day and day of week analysis.

- Figure S7: Households observed via surveys and SmartSpouts.
- Figure S8: Use prevalence versus sensor functionality.
- Figure S9: Distribution of event durations by group.
- Figure S10: Use prevalence before and after fieldworker visits.

Extended Methods

Data Interpretation

Analysis included data from all sensors reporting at least two weeks of data during the 8.9-week reporting period, resulting in 232 sensors reporting over 2000 sensor-weeks of data. Included sensors reported data during an average of 98% of the observation window (range: 82-100%).

A firmware error obscured the exact timing of one entry per 32 hours of device disuse. These entries were categorized as occurring at an uncertain time and the first and last time they could have occurred was computed. To determine the prevalence of single-day and consistent use, sequential entries of uncertain timing were clustered into events meeting the usage threshold (5 s/day) and these events were randomly distributed within the window of uncertainty. Entries occurring at uncertain times were excluded from the time-of-day analysis.

Data Analysis

Unless otherwise noted, reported p-values were computed using (non-parametric) randomization tests with >10,000 replications.

POU Device Design and Use Prevalence

To examine the prevalence of single-day use within groups, we considered each group's daily-averaged use prevalence over time and its distribution. To study why these trends might have occurred, we examined the duration and prevalence of all events, and events excluding noise and Long entries (entries >120 s). Differences in means between groups were obscured by the varied prevalence of Long entries. Instead we report (in the Main Text) the shift in the median duration of events excluding noise and truncating long entries. A similar median shift occurs when noise is included and long entries are left untruncated (Figure S9c).

Distinguishing Single-Day and Consistent Use

We considered households to have 'used' their device if their total duration use was ≥ 5 seconds/day on a given day (*single-day use*), excluding events categorized as Noise. We considered households to have 'consistently used' their device if they had met the single-day use criteria (≥ 5 s/day) for at least 7 continuous days (*consistent use*).

Daily-averaged prevalence of use was reported for any day with at least 33 reporting sensors per day per group. The 33-sensor threshold was selected to bifurcate the 2-week period during which tablets were replaced with filters (see Figure S7).

Surveyed Versus Sensed Single-Day and Consistent Usage

To compare sensor and survey data, we analyzed all monthly surveys (488 survey results from 259 households) and all households with reporting SmartSpouts (232 households observed for an average of 60.6 days each) during the reporting window.

Relating Use Prevalence with Contact Times and Silver Concentrations

We constructed a probabilistic model using the geometric distribution and its continuous equivalent (exponential) to model the probability that water would sit stagnant in a POU device for a duration of $d > 0$ days, as a function of the probability that a household uses all of their stored water on a given day ($0 \leq q < 1$). Non-integer values of d are enabled by the model's continuous equivalent, which is parameterized based on the average rate of device uses per day (λ), where each device use is assumed to use and then replace all stored water:

$$\lambda = -\ln(1-q) \tag{1}$$

We denote the probability that water's stagnation duration D is more than d days at a given household on a given day as the exceedence probability p_e:

$$p_e = P(D > d|\lambda) = e^{-(\lambda d)} = (1-q)^d \tag{2}$$

$$\therefore q = 1 - p_e^{1/d} \tag{3}$$

If p_e is to be limited to a given likelihood (e.g. 5%), Equation 3 suggests the minimum necessary adherence probability. For example, Equation 2 could be used to understand the probability that silver levels exceed 100 µg/L (when $d = 100/35$) as a function of q, and Equation 3 would then suggest $q \geq 65\%$ was necessary to enable a given house in the Tablet group to stay below the WHO limit with 95% probability on a given day.

Assuming each household in a cohort of size N has an independent and identical use probability q, the probability that at least one of the N households has a stagnation duration $\geq d$ is

$$p_N = 1 - (1-p_e)^N \tag{4}$$

$$\therefore p_e = 1 - (1-p_N)^{1/N} \tag{5}$$

Equations 5 and 3 combine to suggest the required probability of single-day use in each of the N households to limit the worst-case stagnation to $< d$. For example, for 100 households to each comply with the WHO silver limit on a given day with 95% probability ($p_{N=100} = 0.05$), Equation 5 suggests each households' $p_e \leq 0.0005$, corresponding to $q \geq 92.9\%$ (Equation 3).

We model the probability that a given house uses their device $k \geq 1$ times over a period of $T > d$ days with the Poisson distribution, using the rate of device use λ as before:[1]

$$P(k|T,\lambda) = \frac{e^{-\lambda T}(\lambda T)^k}{k!} \tag{6}$$

If the device is used k times over the interval T, the probability of any one of those uses occurring after a stagnation period $> d$ is $(1 - d/T)^k$.[1] Therefore we take the probability of having at least one stagnation period $\geq d$ as:

$$p_T = 1 - P(D < d|T,\lambda,k) = 1 - \left[1 - (1 - d/T)^k\right]^k \tag{7}$$

Combining Equations 6 and 7, yields:

$$p_T = 1 - P(D < d|T,\lambda) = 1 - \sum_{k=1}^{\infty} \frac{e^{-\lambda T}(\lambda T)^k}{k!} \left[1 - (1 - d/T)^k\right]^k \tag{8}$$

Which can be solved numerically for a critical value of λ and thereby q.

In the case of silver-releasing tablets, for a single household to have a 95% chance of staying below the WHO limit continuously for two years, Equation 8 suggests $\lambda \geq 3.8$ uses per day, corresponding to a single-day use probability $q \geq 97.8\%$.

To ensure that a cohort of N households all have stagnations times $\leq d$ over a period T, we substitute $p_e = p_T$ in Equation 5 and solve for the minimum required use. For the silver example, to ensure a 95% chance of having a maximum stagnation time of $d = 100/35$, over the whole cohort of $N = 100$ households over a period of $T = 760$ days, requires $p_T \leq 0.0005$ for each house (Equation 5), corresponding to a single-day use probability of $q > 99.6\%$ (Equation 8).

We demonstrate the explanatory power of this probabilistic model using measured silver concentrations and SmartSpout data. Silver concentrations were measured 6-weeks before SmartSpout installation. Sensed adherence data was available from 73 (84%) of households in the Tablet Group; silver data was available for 72 of these 73 households.

Fieldworker Efficacy and Insights from Time of Device (Non)Use

To examine the effect of fieldworkers administering monthly surveys, we compared the prevalence of use for 5-7 days before and after each survey, excluding the date of the survey. The full week before and after was considered when data was available; before and after windows were always of equal size.

Time-of-use analysis shows the distribution of use throughout an average week by day of week, hour of day and both. The distribution of use duration was computed for each household by day of week and hour of day. This distribution was then normalized by the household's total use. The overall use pattern was then averaged across households, weighting each household equally.

Supplementary Figures

Entries → Truncated (if >120 s) → Classified Events

Start: 9:00 AM July 1

Inter-Entry Interval (min)	Duration (s)	Truncated Duration (s)
30	1600	max{5, μ}
0	0.1	0.1
0	2	2
12	6	6
240	30	30
6	0.4	0.4
0	0.6	0.6
0	1	1
24	12	12
0	6	6

Date & Time	Duration (s)	Average Duration per Entry (s)	Classifi-cation
9:30 AM July 1	8.56	2.85	Water
9:42 AM July 1	6	6	Water
1:42 PM July 1	30.4	15.2	Water
1:48 PM July 1	2	0.67	Noise
2:12 PM July 1	18	9	Water

μ : Household's mean duration of entries $\leq$120 s
here, μ = 6.46 s

Figure S1: Data processing and noise classification method and example. *Entries* recorded the elapsed time since the last entry (inter-entry interval) and duration of device use. *Long* entries (>120 s) were truncated to the greater of either 5 s or the mean (non-long) entry duration. Entries with zero inter-entry spacing were combined into a single *Event*. Average duration per entry (s) was used to classify events as *Noise* or *Water* (see Figure 2).

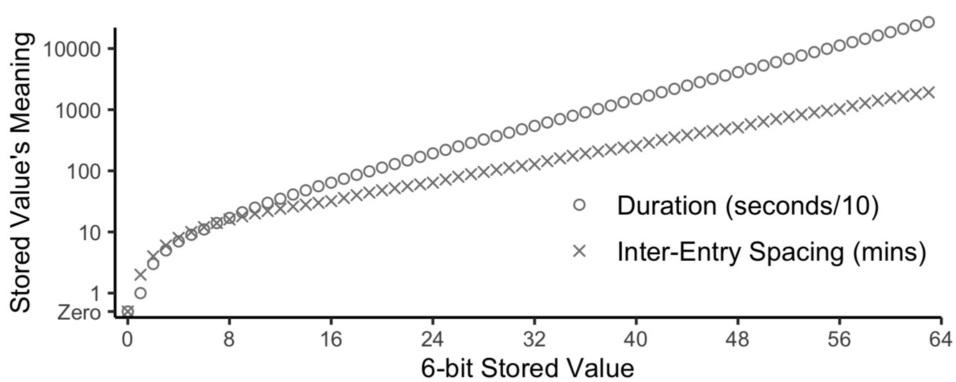

Figure S2: Data compression method. 6-bit dictionaries were used to compress durations (open circles) and inter-entry spacing (x's). Stored, 6-bit values correspond to the duration (in tenths of seconds) or inter-entry spacing (in mins) shown on the vertical axis (log spaced, with zero superimposed).

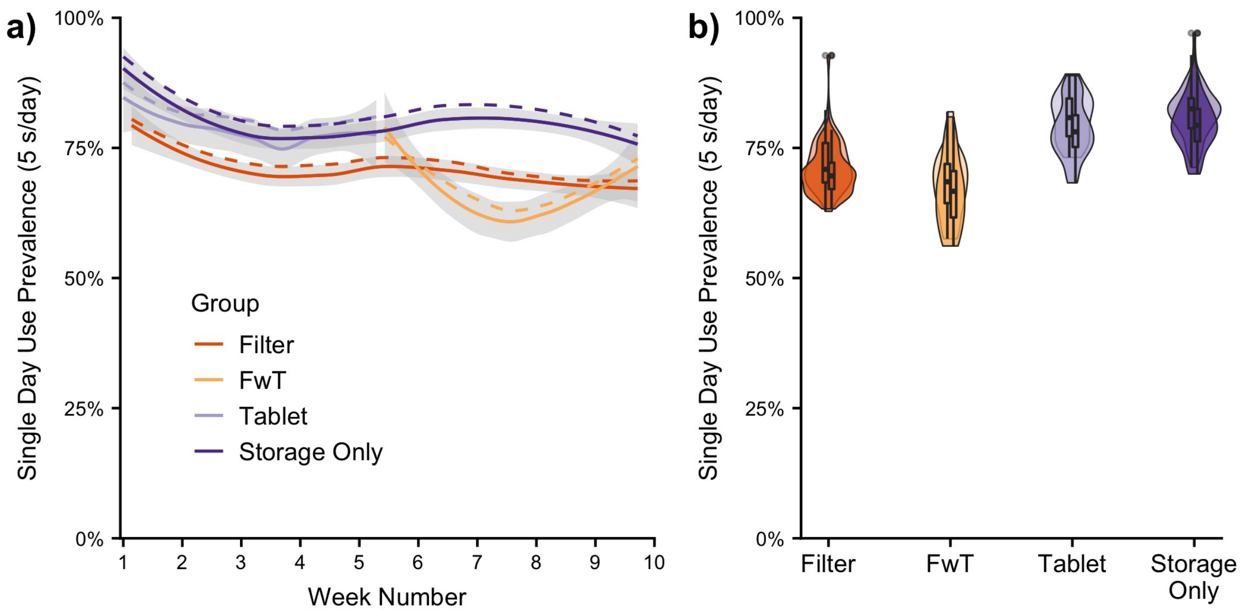

Figure S3: Noise sensitivity of group usage trends. Plot mirrors Figure 3: a) Group-averaged prevalence of single-day use, summarized by locally-averaged scatter plot smoothing (lines) with shaded 95% confidence intervals; lines show the trend with (dashed) and without (solid) noise included in the analysis. b) Variations (by day) in group-averaged single-day use are summarized by violin plots; results with noise included are shown as silhouettes behind the original violin plots. Inset boxplots summarize the data with (left) and without (right) noise. FwT: Filter, was Tablet.

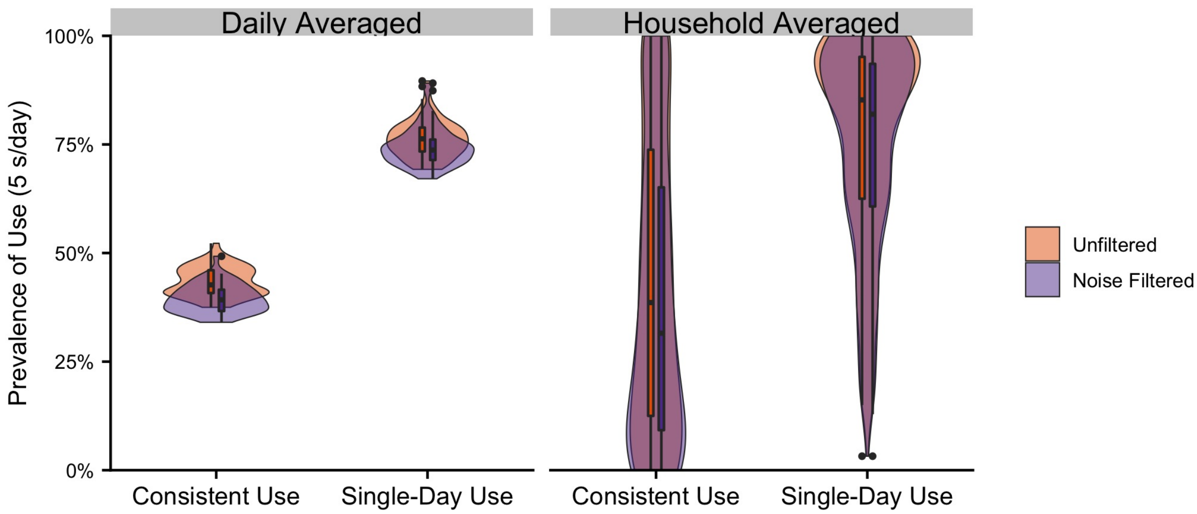

Figure S4: Noise sensitivity of sensed and surveyed prevalence of use. Corresponds to Figure 4. Variations in average (sensed) prevalence of use ($\geq$ 5 s/day) between days (left panel) and between households (right panel) was summarized by violin plots (kernel densities) with inset boxplots. Results with (orange) and without (purple) noise shown. Daily-averaged use was more sensitive to noise filtering than household-averaged. Reported trends of low adherence are not sensitive to noise filtering.

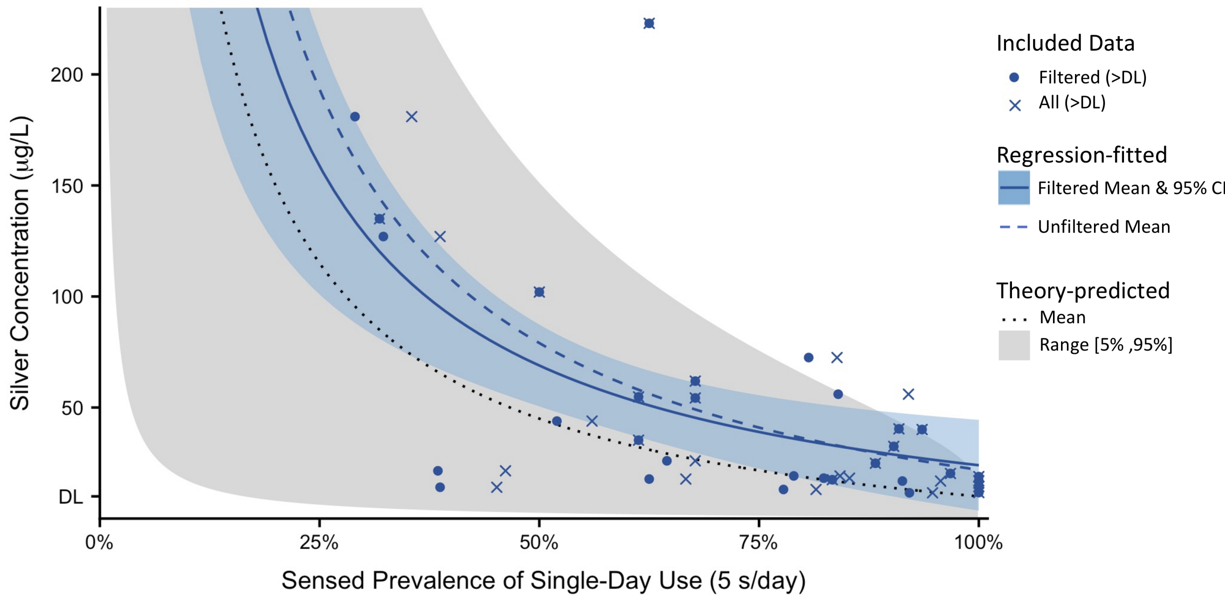

Figure S5: Noise sensitivity of silver versus use prevalence model. Corresponds to Figure 5. Household-averaged prevalence of single-day use in the Tablet Group compared to measured silver concentrations above the detection limit (DL=10 µg/L), with (x's) and without (dots) Noise included. Based on random use, our theory-predicted the range of expected silver levels, spanning 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles is shaded. The theory-predicted mean concentration of samples above the DL is shown (dotted line). Least squares (regression) best fit (with $y \sim 1/x$; R^2=0.36) is shown as a solid blue line with shaded 95% confidence interval based on data without noise; best fit regression line based on all data is shown as dashed (R^2=0.36). Most silver measurements were within the theory-predicted range, given daily usage (with and without noise). Regression-fitted relationships between single-day use and silver concentrations were not significantly affected by noise inclusion/exclusion.

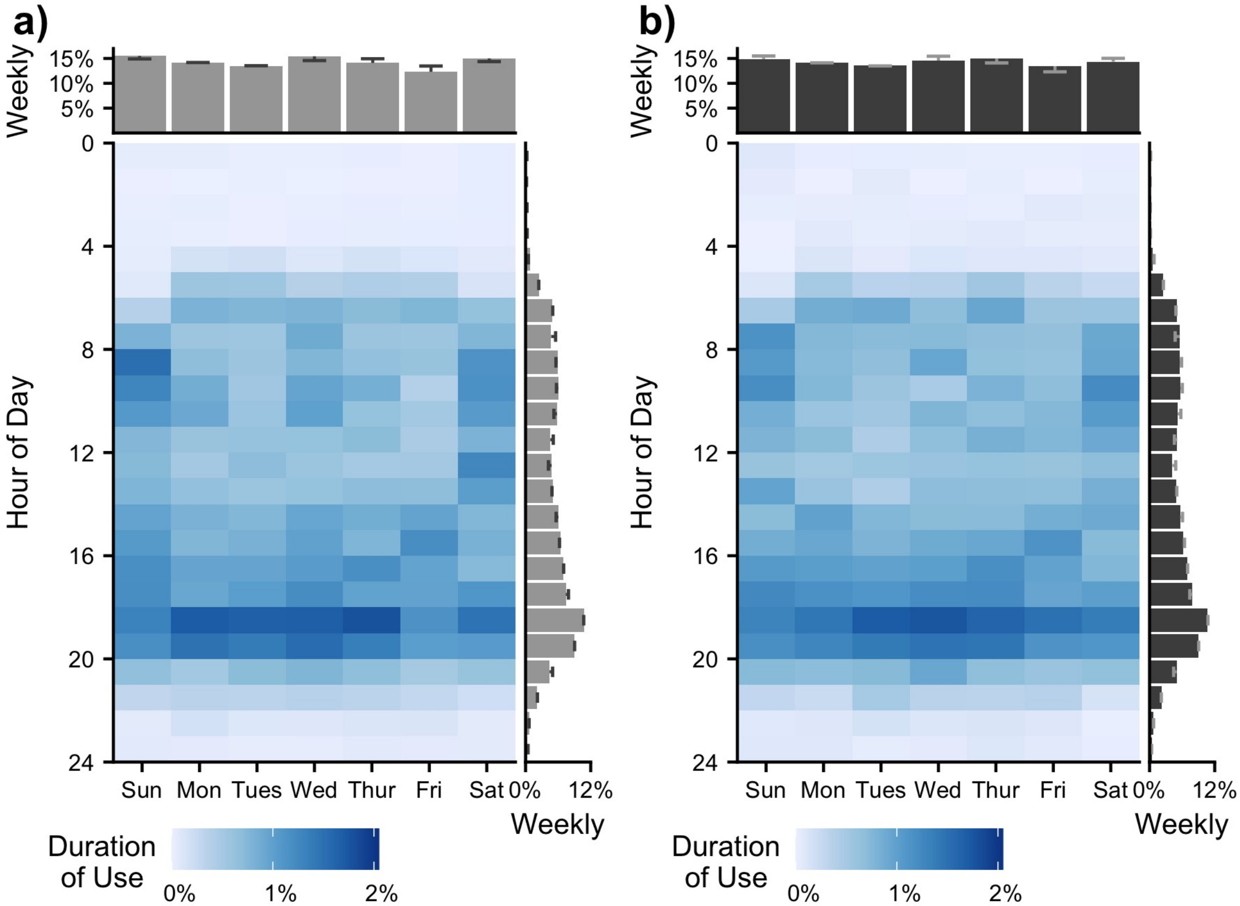

Figure S6: Noise and long-entry sensitivity of time of day and day of week analysis. Relative duration of use by hour and weekday with (a) and without (b) noise filtering and long-entry trimming. The percentage of weekly use duration is colored (bottom left) and aggregated by weekday (top) and hour of day (right). a) repeats with Figure 6. b) Equivalent to a) except that it includes the duration of all noise events and the untruncated duration of long entries. Error bars on the weekly (top) and hourly (right) marginals indicate the equivalent value in the other subplot; error bars in a) indicate value shown in b) and vise versa. Overall trends in time of use during the week are relatively comparable; the largest divergence is on Fridays, where trimming entries >120 seconds and removing events classified as noise substantially reduced the relative duration of use.

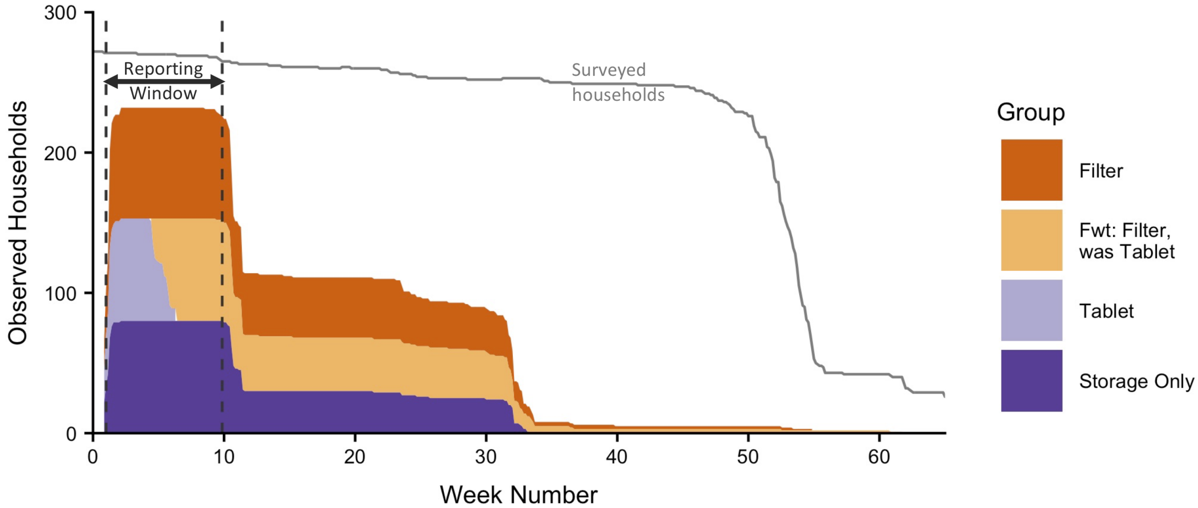

Figure S7: Households observed via surveys and SmartSpouts. The total number of households observed via surveys (grey line) and SmartSpouts (shaded by group): Filter Group (orange); Filter, was Tablet Group (beige/light orange); Tablet Group (light purple); and Storage Only Group (dark purple). Vertical dashed lines demark the two-month period in which we report SmartSpout data.

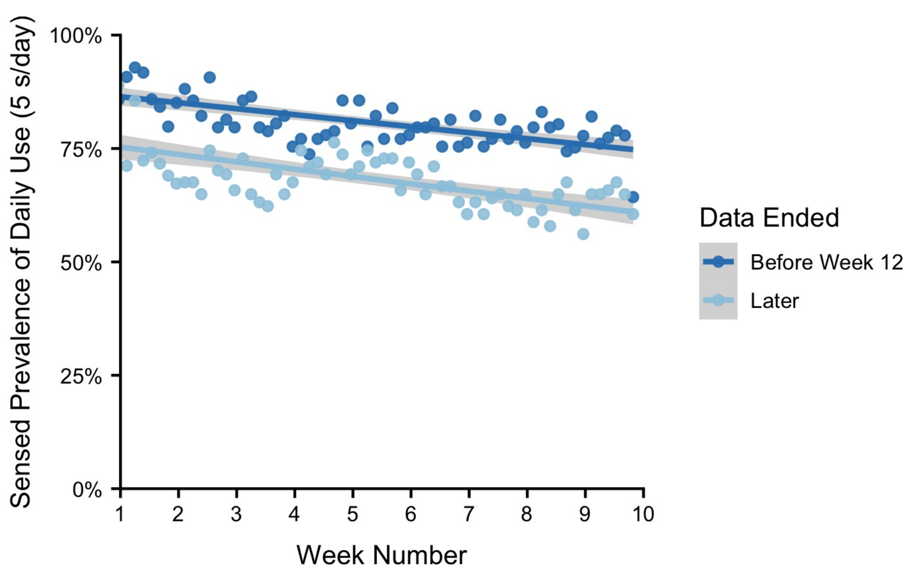

Figure S8: Use prevalence versus sensor functionality. Sensors that report data on or after Week 12 (light blue) report significantly lower sensed prevalence of single day use than sensors that stopped reporting data before Week 12 (dark blue). Straight lines are the least squares linear best fits with shaded 95% confidence intervals.

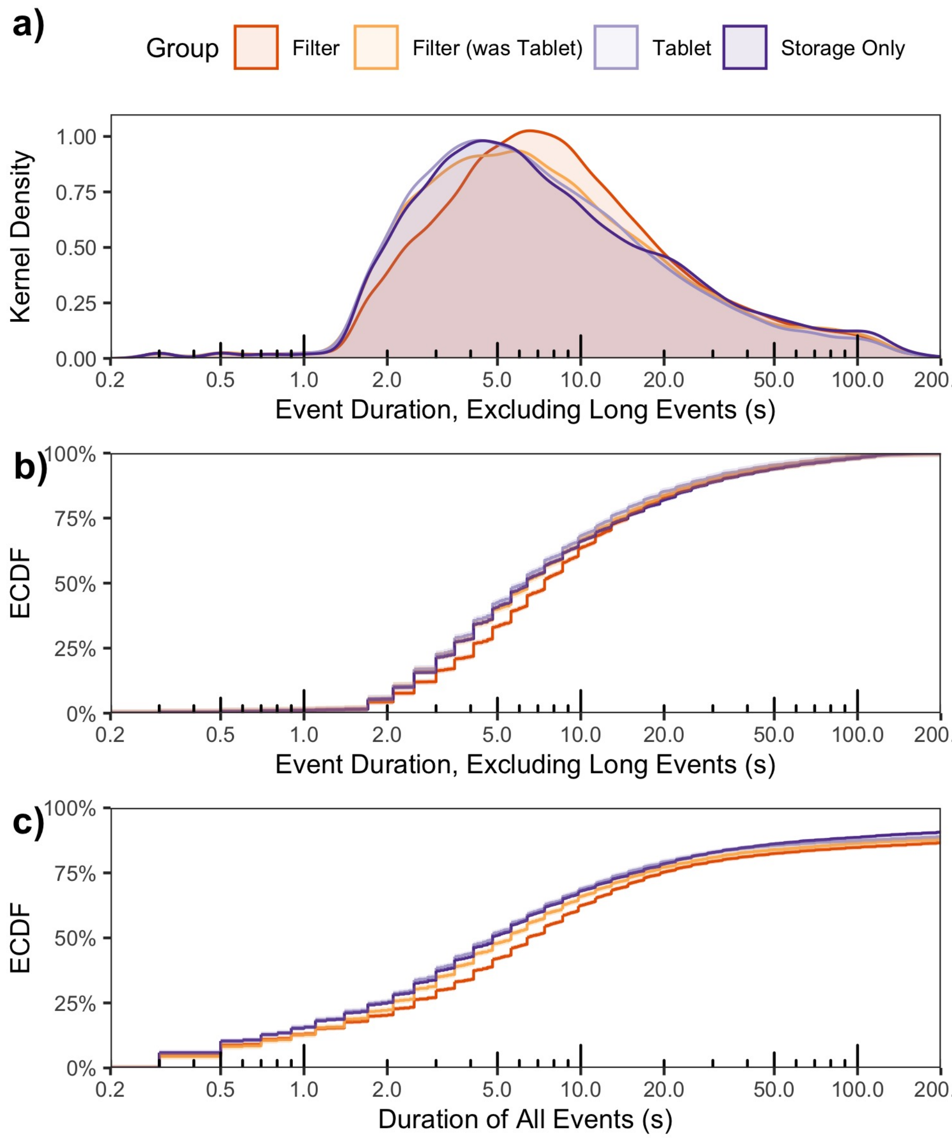

Figure S9: Distribution of event durations by group. a) The kernel density (shows the distribution of values) of events excluding noise and entries >120 seconds, b) empirical cumulative distribution function (ECDF) of a), and c) ECDF of every event's duration by group. In b) and c), semi-transparent shading shows the 95% confidence bands, generated using the Dvoretzky–Kiefer–Wolfowitz inequality. Filter (orange), Filter was Tablet (beige), Tablet (light purple), and Storage Only (dark purple). Median events in the Filter Group were statistically significantly longer than in other groups, perhaps because of slower flow rates. This difference was significant with and without noise filtering (b & c).

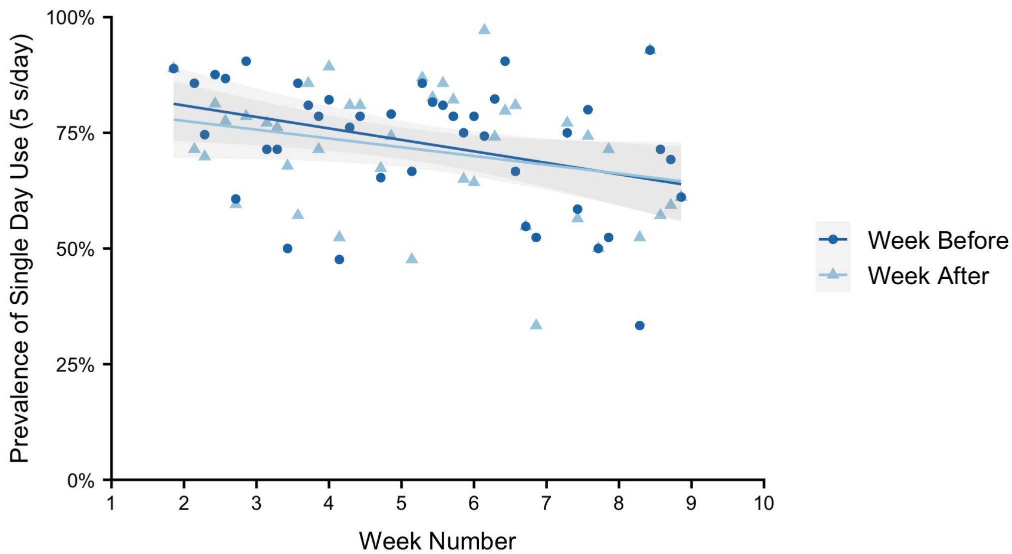

Figure S10: Use prevalence before and after fieldworker visits. The average single-day use prevalence during the week before (dark blue dots) and after (light blue triangles) fieldworker contact. Single-day use prevalence was slightly reduced following fieldworker contact, but this reduction was not significantly different than overall usage declines (lines of best fit with 95% confidence intervals shaded).

References

(1) Gallager, R. G. *Stochastic Processes: Theory for Applications*; Cambridge University Press, 2013.