Rational Habit Formation: Experimental Evidence from Handwashing in India

Reshmaan Hussam, Atonu Rabbani, Giovanni Reggiani, Natalia Rigol

Online Appendix

A. Learning and child health

Do households internalize these substantial child health returns and increase their valuation of handwashing (and thereby their handwashing rates in the long run) accordingly? To test the extent to which learning about the health returns to handwashing generates persistence, we run the following regression separately for dispenser-only, monitoring, and incentive households:

$$(1)\quad Persistence_{cv} = \alpha + \beta_1 Health_{cv} + \beta_2 HandwashStock_{cv} + \beta_3 BaselineHealth_{cv} + \delta_c + \gamma_v + \epsilon_{cv}$$

in which $Persistence_{cv}$ is the average handwashing performance during the month following the withdrawal of incentives or monitoring for child c in village v, *Health* is a health index constructed using Anderson (2008) separately for self-reported disease incidence and anthropometric outcomes[1], *HandwashStock* is the average likelihood of washing during dinnertime over the course of the intervention, *BaselineHealth* is the identical incidence or anthropometric index constructed using baseline health variables, δ is a vector of child and household-level characteristics (sex and age of child, whether child was breastfed exclusively, household occupation, number of rooms, mother's age at marriage, and mother's education) and γ is village fixed effects. Standard errors are clustered at the household level. A significant and positive β_1 coefficient implies that, conditional on having accumulated the same amount of consumption stock of handwashing, households that experience larger improvements in health are more likely to persist in their handwashing behavior.[2]

Appendix Table 13a presents the results separately for each treatment arm and health index type. All estimates of the coefficient on the health index are statistically insignificant and close to zero. It does not appear that households are internalizing health gains and updating their handwashing performance accordingly.

Despite the host of controls for child health and household characteristics, it is possible that learning effects are washed out by endogeneity in handwashing behavior to household type: households who experience larger health returns may also be the types of households who handwash little (for example, the sick children who experience the largest health improvements may reside in poor households - who are on average less likely to wash than their affluent counterparts - in a manner that is not sufficiently controlled for in our vector of child and household characteristics). Therefore, we also exploit our panel data on illness collected during months three through five of the experiment and consider the following exercise: conditional on households having built the same amount of handwashing stock and experiencing equal levels of sickness, does a household that experiences an illness the week before a handwashing observation behave differently from a household that experiences an illness in the week after the observation? Any difference can plausibly be attributed to the reaction to the health event rather than changes in consumption stock, since the latter is equivalent across comparison households. To evaluate this, we run the following regression for households who report an ARI episode in either the week before or after the week of handwashing observation, run separately for each week of child health panel data[3]:

$$(2)\quad Handwashing^t_{cv} = \alpha + \beta_1 Sick^{t-1}_{cv} + \beta_2 Sick^t_{cv} + \beta_3 SickStock^{t-1}_{cv} + \beta_4 HandwashStock^{t-1}_{cv} + \gamma_v + \epsilon_{cv}$$

[1] We include anthropometric outcomes for completeness, although given the magnitude of effect size, these are likely much more difficult for a mother to internalize and learn from than changes in diarrhea and ARI incidence.

[2] This translates into a learning effect of health returns given two assumptions: first, that the relationship between handwashing and health is not one-to-one, but rather there is a random component to the health improvements that a child experiences from a unit of handwashing; and second, that households are unable to separate the random from the direct components of health improvements in their learning process: a household that observes a large child health improvement will attribute the full gain to handwashing, even if their neighbor accumulates the same amount of handwashing stock and sees only a small improvement in child health.

[3] We examine only ARI outcomes for the panel data given the complications in collecting child diarrhea outcomes prior to the revised question formatting in the midline survey.

In which $Handwashing_{cv}$ is the total number of days the dispenser was used at dinnertime in week t for child c in village v, $Sick^{t-1}$ is a binary variable that equals one if the child is sick in the previous week and zero of the child is sick in the following week, $Sick^{t}$ is a binary variable that equals one if the child is sick in the current week, $SickStock^{t-1}$ is the total number of episodes the child experiences from the first day of observation to the start of the previous week, $HandwashStock^{t-1}$ is the total number of days the dispenser was used from the first day of observation to the start of the previous week, and γ is village fixed effects. Standard errors are clustered at the household level. Our coefficient of interest is β_1: a negative and significant coefficient would suggest that, holding total sickness and handwashing stock constant, children (households) who experience a sickness in period $t-1$ devalue handwashing and wash less in period t relative to those children (households) who experience a sickness in period $t+1$. Conversely, households that remain healthy in period $t-1$ learn that handwashing is good for health and therefore wash more in period t relative to those that remain healthy in period $t+1$.

Appendix Table 17 presents the results. Panel A presents results for households in either the dispenser-only or the monitoring arms, and Panel B presents results for households in either the dispenser-only or the incentivized arms. These samples thus correspond to those of the persistence analysis in Table **??** (Panel A). Over the course of the weeks in which we can observe before, during, and after ARI incidences, no consistent pattern emerges. Estimates are noisy, with an equal distribution of negative and positive coefficients. It does not appear that households are - at least coherently or consistently - internalizing the health returns of their children and updating their valuation and performance of handwashing accordingly.[4]

Finally and most decisively, note that the rational habit formation effect can only be driven by intertemporal complementarities in the stock of consumption, *not* by learning about child health effects. This is because the experiment exogenously increased only the value of handwashing in the future, not that of the health returns to handwashing in the future.[5] Evidence of rational habit formation by households anticipating the monitoring of handwashing behavior therefore offers further evidence that learning about the health returns cannot be the primary driver of intertemporal complementarities in handwashing. Rather, the persistence we observe is most likely driven by the accumulation of consumption stock, or the building of a habit.

B. *Household beliefs around handwashing behavior*

To examine the degree of sophistication households possess around planned handwashing behavior, we elicited biweekly forecasts of handwashing: we asked each respondent (mother) to forecast how many days in the upcoming week she and her children expected to wash their hands with soap before dinnertime. Appendix Table 11 reports the results. Forecasts offer further suggestive evidence that households for whom the future service matters (those anticipating being monitored but not those anticipating a tripling of tickets) internalize the anticipated change in behavior: households anticipating monitoring forecast that they will wash 0.23 more days (4% more) than their unanticipating counterparts (Columns 1 and 2). We see no such forecasting effect among anticipating incentivized households. However, note that neither set of households forecasts greater

[4]Our test of learning about health benefits is not perfect: perhaps mothers notice more subtle health improvements in their children that are not captured in our loose stool and ARI metrics. Though possible, we find this unlikely; in our field experience, whenever mothers were probed on their comment that the dispenser has made their children healthier, they consistently described "healthier" as "fewer coughs, runny noses, fevers, or loose stool." Given that the health measures were self-reports, we should have been able to capture perceived changes along precisely these dimensions of health.

[5]Upon being randomized into receiving the future price change or monitoring service, treatment households face an increased future return to the behavior but, in a world without rational habit formation, identical current returns to the behavior. In the typical risky technology and learning experiment, one subsidizes current behavior and examines effects on future returns. In this study, we subsidize future behavior and examine effects on current behavior, which yields clear evidence of intertemporal complementarities, the hallmark of habit formation. It is in this way that the learning and habit formation stories can be distinguished, and our experimental design identifies only the latter mechanism.

handwashing rates when the future change arrives relative to those who experience no change (Columns 3 and 4): even monitoring households, for whom we observe a real increase in handwashing relative to their dispenser control counterparts, do not articulate this change insofar as the forecasting question elicits. It is possible that these near-zero effects are due to a ceiling effect: the average respondent in the respective control groups already forecasts washing more than six days in the week; treated respondents cannot forecast much higher. Finally, Columns 5 and 6 suggest that, at least on the extensive margin of receiving versus not receiving a dispenser, forecasts appear to be strongly predictive of the truth: those who receive a dispenser (whether in the incentives, monitoring, or dispenser control treatment arm) forecast washing more than twice as much as their pure control counterparts. We interpret this forecast data with considerable caution and minimal weight: forecasts are exceedingly difficult to elicit well given the unusual nature of the question and the potentially strong experimenter demand effects.

C. Alternative measures of household hygiene and sanitation

While the sensor data of dinnertime dispenser use is our primary source of hand hygiene data, we collected a series of additional observational and self-reported hygiene outcomes that are commonly employed in the literature. Surveyors observed the cleanliness of respondent hands and nails at the time of survey and graded each on a three point Likert scale: 0 indicating no visible dirt, 1 indicating some visible dirt, and 2 indicating extensive visible dirt. This direct observational measure is a popular primary outcome in the handwashing literature (Bennett, Naqvi and Schmidt, 2018; Ruel and Arimond, 2002; Luby et al., 2011; Halder et al., 2010). However, given the subjective nature of the rating and the fact that surveyors are not blinded to treatment assignment in this (and most) hygiene experiments, this measure is vulnerable to surveyor bias. If subjects realize they are being observed (which is not uncommon in practice despite efforts to remain discreet) it is also subject to observation bias. We also collected respondent ratings on handwashing habit formation. Respondents were asked "Has handwashing with soap before eating become habitual for you?" and were rated on a five point scale using the following metric: 0 = "How? You did not give us soap"; 1 = "No, not at all"; 2 = "No, not yet, but it is growing"; 3 = "Yes, mostly, but still needs time"; 4 = "Yes, definitely, the habit has been established." Third, surveyors asked the respondent whether they had any liquid soap in the household; for treated households, the question specified that we were interested in non-project liquid soap. If households mentally assign barsoap to purposes like bathing and laundry, the presence of liquid soap may be a signal that handwashing is a household priority. These three hygiene measures were collected at midline, seven to eight months after rollout. Finally, we proxy for the amount of soap consumed by a household using the total number of dispenser presses per day.

Results are presented in Appendix Table 12 for pooled and disaggregated treatment arms. Treatment assignment in the pooled sample is predictive of all alternative hygiene measures. The disaggregated samples broadly follow the pattern established by our primary hygiene outcome measure of dinnertime dispenser use, with the incentive arm reflecting larger treatment effects within most measures.[6] However, the disaggregated treatment effects are statistically indistinguishable from one another. These results suggest that alternative, inexpensive measures of hand hygiene are informative for high-intensity interventions; however, more precise measurement techniques are essential for identifying the underlying mechanisms behind behavioral change in handwashing.

We also explore the impact of the interventions on the household's sanitation behavior. A change in hand hygiene may be complemented by changes in other sanitation practices, if for example the

[6] In particular, the incentive effect is half the size of the monitoring effect in the observed hand cleanliness measure; this may be reflective of the measure's vulnerability to Hawthorne effects and/or surveyor bias, as monitored households may have been more conscious of keeping their hands clean when the surveyor visited, or surveyors may have felt a greater (subconscious) obligation to report cleaner hands among households they monitored

act of having handwashing top of mind makes remembering to maintain other preventive health practices easier. It is also important to examine effects of the interventions on other sanitation outcomes as they affect our interpretation of the results on child health: improvements in sanitation may be the real cause of improvements in child health and handwashing merely a correlate. Appendix Table 13 presents the two household level sanitation outcomes collected during the midline survey: whether the household practices open defecation and whether they treat their drinking water. Treatment assignment is not predictive of either of these outcomes: coefficients on treatment are small in magnitude and imprecise, suggesting that the interventions had no complementary effect on other dimensions of household sanitation.

D. Household willingness to pay for soap

Despite the evidence that the intervention lowered the cost of handwashing by making it habitual and significantly improved child health outcomes, it is *ex ante* unclear whether households internalize these impacts of handwashing when making their hygiene and sanitation-related purchasing decisions. One way to explore this question is through the elicitation of a household's willingness to pay (WTP) for soap. We play a WTP game using the Becker-DeGroot-Marschak methodology with households at the eight month mark after all interventions have been phased out. Respondents (mothers, often with their children accompanying them) were presented with a series of prizes of increasing value.[7] At each level, the respondent was asked whether she would prefer to take the prize or take a month's worth of soap.[8] To ensure incentive compatibility, each choice was made in the form of a token and dropped into a bag; after the completion of all choices, the respondent chose one token at random and received the drawn prize.

Results are presented in Appendix Table **??**. Contrary to expectations, treated households value an additional one month of soap significantly *less* than control households. A disaggregation by treatment arm (Column 2) reveals that this difference arises entirely from formerly incentivized households, who express a willingness to pay that is 11% lower than that of control households. Valuations among monitoring and dispenser-only arms are statistically indistinguishable from those of pure control. One interpretation of this result is that the prizes from the incentives intervention gave the mothers (and/or children) a taste for such rewards which crowded out, rather than complementing, the value of soap. Households may have anchored their valuation of soap to a negative price as they became accustomed to being paid to use it.

However, formerly incentivized households are also significantly more likely than their pure control counterparts to have non-project liquid soap in the household (Appendix Table 12, Column 8), so their lower valuation may be due to having already established a source for liquid soap once project soap provision ends. Column 3 therefore excludes all households that report having non-project liquid soap in the household. Coefficients change only marginally; incentive households still have a 14% lower valuation of soap than control households. Appendix Figure 9 plots the average WTP across each treatment arm for this restricted sample.

Echoing the results on child health and the absence of learning, this valuation exercise underscores a problem at the heart of behavioral change in preventive health: health benefits of preventive behaviors are often too small, too delayed, or too difficult to observe relative to what is required for households to internalize the causal relationship between behavior and health. Even in a setting where behavioral change generates health effect sizes that are twenty percent at the lower bound,

[7] Because of logistical and contextual concerns, we were not permitted to offer respondents cash. We therefore generated a list of prizes of increasing market value, ranging from Rs. 5 to Rs. 150, which were distinct from the prizes formerly offered to incentive households, and which households, in extensive piloting, could accurately estimate the market value of.

[8] Respondents were informed that their prize or soap would be delivered to them in six months time. This was a necessary caveat because treatment households had been promised free soap for one year from rollout; if the soap from the game were to come during this period, its marginal value would be lower by construction, preventing a valid comparison with pure control households.

the household's decision-makers on child health do not appear to draw the link between liquid soap provision, the likelihood of handwashing, and child health outcomes.[9] Importantly, the same argument applies to habit formation: despite the considerable handwashing stock accumulated over eight months and evidence of persistence in handwashing, households do not increase their willingness to pay for soap. At the point of playing the willingness-to-pay game, neither the return from habit nor the return to health was sufficiently internalized (or sufficiently high) to shift households' monetary valuations of soap.[10]

E. *Behavioral spillovers*

Despite no obvious changes imposed on dispenser-only households throughout the experiment, these households demonstrate a rise and fall in handwashing rates that closely mimics the pattern of monitored households (Appendix Figure 8). This pattern could be due to parallel time trends, the dispenser control households undergoing their own process of habit formation, or to spillovers in behavior from neighboring monitored households.

Because treatment assignment between dispenser-only and monitoring was randomized at the household level, we capitalize on the random variation in the concentration of monitoring households nearby dispenser-only households to estimate the size of spillovers in handwashing behavior.[11] We choose a radius of one kilometer around each dispenser-only household, as this is a typical distance within which children play with one another and attend the same government nursery school, mothers walk to the local pond or road-side shop, and most conversations are likely to occur. We examine spillovers at three points in time: Day -40 to -30, when there is little that dispenser households can learn from monitoring households; Day 40 to 50, ten days after monitoring households have received their first calendar (which gives them time to share their experiences with neighbors), and Day 120 to 130, after monitoring is officially over. If spillovers drive the rise in rates among dispenser-only households, we should only observe the effects of spillovers in the middle specification, and potentially remnants in the third specification.[12] Results are presented in Appendix Table 15. Consistent with the prediction, there are zero spillovers in the early part of the experiment, some evidence of positive spillovers during the peak of discovery in the monitoring regime (unadjusted for multiple hypothesis testing, the coefficient is significant at the ten percent level), and a dropoff after monitoring ends. However, the magnitude of these spillovers is modest relative to the upward trend in handwashing observed among dispenser households over the same time period: at the peak of the monitoring regime, having one more monitoring household within one kilometer of a dispenser household is associated with a 1.3 percentage point (4%) increase in dispenser household handwashing rates. Thus while spillovers from monitored neighbors may have played some role in the handwashing behavior of dispenser households, they can only explain a fraction of the observed rise (nearly a doubling) in handwashing among dispenser households in the first three months of the experiment.

The pattern we observe may alternatively be due to parallel time trends or the natural process

[9] This WTP exercise was in fact biased towards finding a higher WTP among treated households: the liquid soap was presented in a refill pouch, which is more valuable if one has a liquid soap dispenser in the home.

[10] Note that our rational habit formation result provides evidence that the effects of habit formation are sufficiently large to affect *behavior*; this, however, appears not to translate into changes in monetary valuation for soap. This could be due to a variety of reasons, such as mental accounting (households allocate a fixed budget to soap/hygiene that is difficult to shift) or price anchoring (formerly incentivized households anchor their perceived price of soap at a negative value given that they were effectively paid to use soap for four months).

[11] We define concentration of treated households in levels (number of households) rather than percentages because our sample is far from a complete census of all households in a village, so our denominator would be an ineffective proxy for total number of neighboring households.

[12] These time bins were not specified in the pre-analysis plan, but were specified prior to running this analysis; given the large set of choices one could make in this analysis, alternative time bins were not explored. Alternative distances were explored: 0.5 km radius and 2 km radius both yield estimates nearly identical in magnitude, with the former the least precise (results available upon request).

of habit formation. While we cannot rule out the former, habit formation is not unlikely. Consider a habit formation model in which there exists some fixed amount of consumption stock which must be accumulated before σ kicks in. This permutation of the model is consistent with the initial shallow decay of handwashing rates in dispenser control households (Appendix Figure 8, Day -70 to 0) followed by their steady rise (Day 0 to 90). Given that surveyors switched from twice-monthly visits to collect health data to monthly visits to collect data (across all sample households) around Day 110, which can be regarded as a positive shock to x_t, the subsequent decay in handwashing rates is likewise consistent with the habit formation model. Therefore the pattern of a secular rise in handwashing rates amongst dispenser households suggests the role of habit formation in handwashing over time even absent monitoring or incentive interventions.

F. Health spillovers

Despite the lack of significant behavioral spillovers, we may expect to see spillovers in health given that viral and bacterial contamination are the primary sources of diarrhea and ARI morbidity. To measure these spillovers, we exploit the random variation in the concentration of treated households (pooled) within a one kilometer radius of pure control households. We run this exercise separately in monitoring villages (MV) and incentive villages (IV) as households were randomized into pure control and treatment only within these village categorizations. Appendix Table 16 presents these results. While most coefficients are negative, as one would expect with positive health spillovers, nearly all are small and imprecise. We find some evidence that having one additional treated neighbor reduces a pure control child's days of ARI by 0.03 days and reduces her likelihood of having ARI symptoms by 0.2 percentage points in monitoring villages (coefficients significant at the ten percent level, unadjusted for multiple hypothesis testing). Therefore despite substantial positive health benefits, the habit of handwashing at dinnertime produces modest health externalities for neighboring children. This is not especially surprising given the timing of the behavioral change we focus on: while children are most prone to spreading germs during the daytime at school and as they play, our intervention improves hand hygiene only at night. To maximize positive spillovers, we may want to focus on hand hygiene interventions linked to schools or a child's midday meal. This is an important direction for future research.

Appendix Figures

Figure 1. : Soap dispenser anatomy

Note: The dispenser is a standard wall mounted handsoap dispenser with a foaming pump. It is opened with a special key available only to the surveyors. The sensor module is secured inside between the pump and the liter container.

Figure 2. : Typical dispenser location

Note: An infant sleeps on the verandah of a home. The dispenser is nailed to a wall of the verandah at a height accessible by young children. The verandah is the common space for dining.

Figure 3. : Child using dispenser

Note: A child uses the dispenser by pushing the black button once or twice. The foaming soap can be rubbed on the hands without water. He then goes to the nearby water pail or tubewell in the courtyard and rinses the soap off with the help of the mother, who pours the water

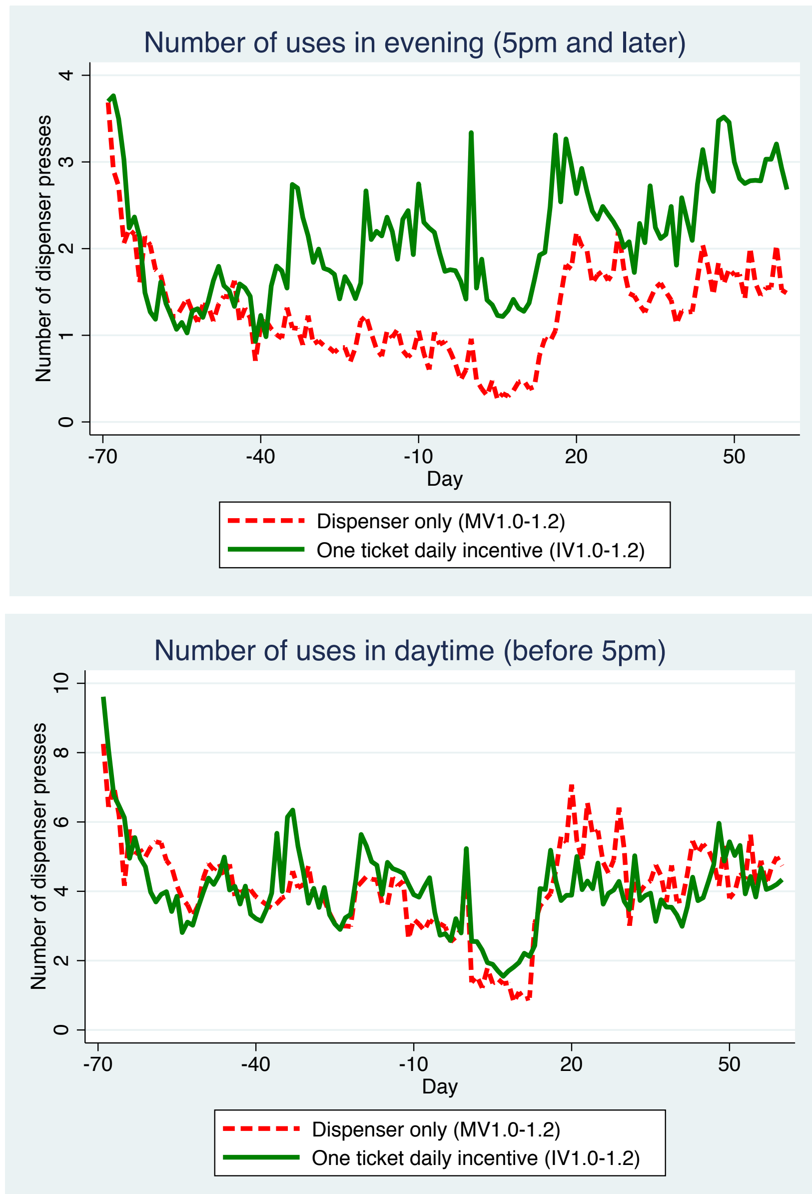

Figure 4. : Dispenser use over 24 hours

Note: Figures show the average number of individual presses per day after 5pm and before 5pm, respectively. Dashed red line represents households who received only the dispenser; green line represents households who received the dispenser, feedback, and one ticket for every night the dispenser was active around their self-reported dinnertime. Day -70 is the day of rollout.

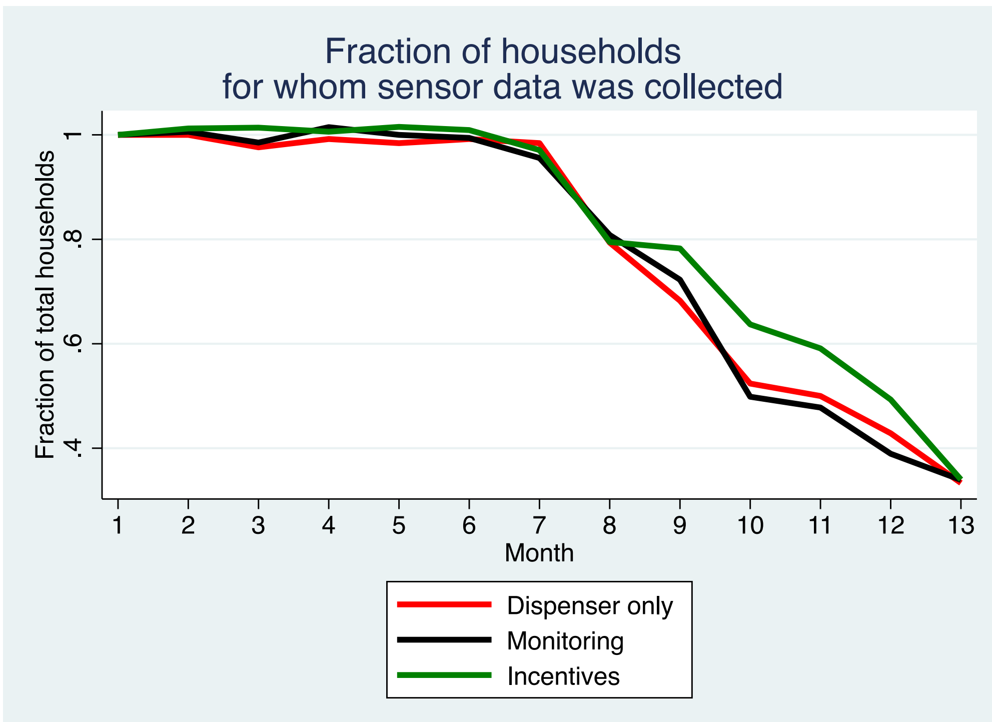

Figure 5. : Time trend in attrition of sensor data

Note: Figure plots fraction of households in each treatment arm that have sensor data collected over the course of the experiment.

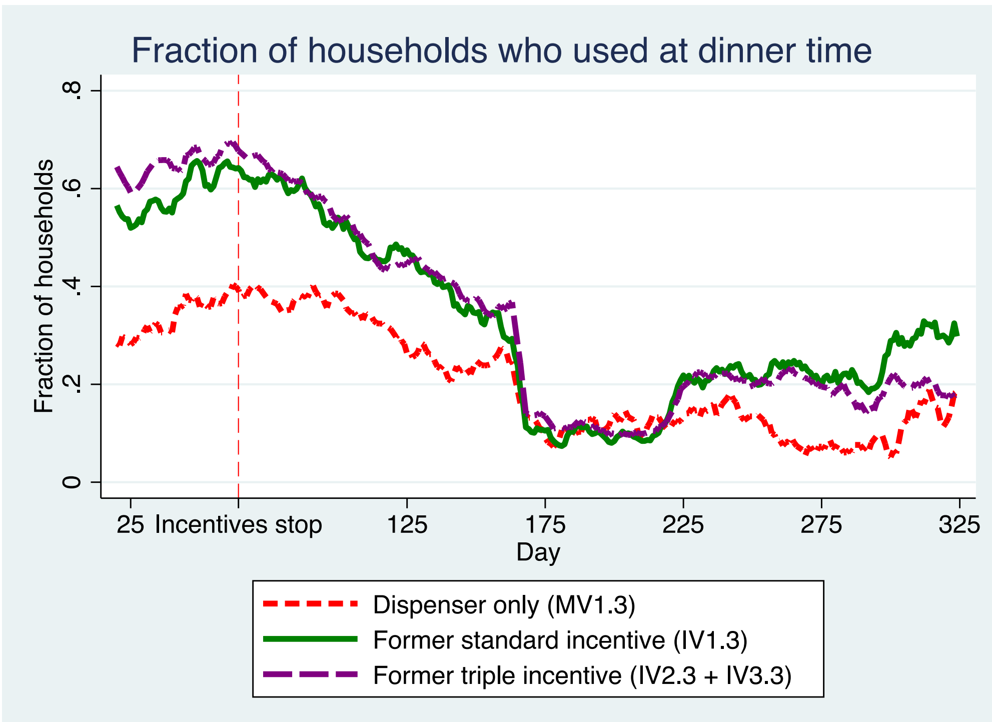

Figure 6. : Persistence of incentive effect

Note: Figure shows the five day moving average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Red dashed line represents households who received the dispenser only; green line represents households who received the dispenser, feedback, and one ticket until the point of the "Incentives stop" (Day 60), after which they stopped receiving tickets or feedback and therefore became identical to dispenser-only households; purple line represents households who received three tickets until Day 60 and none thereafter.

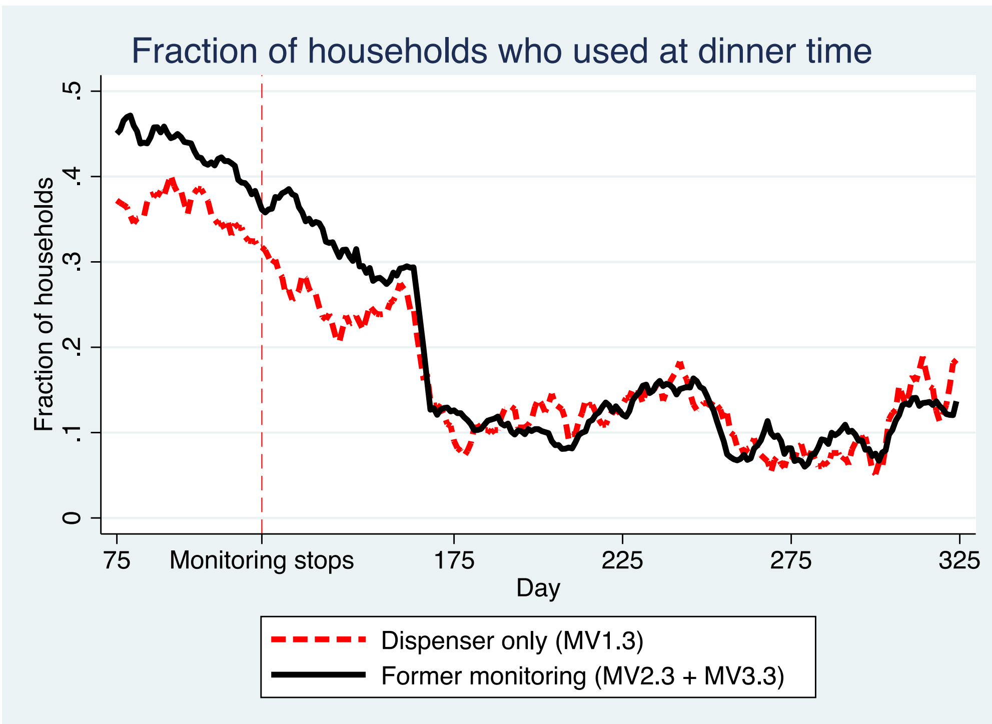

Figure 7. : Persistence of monitoring effect

Note: Figure shows the five day moving average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Red dashed line represents households who received the dispenser only; black line represents households who received the dispenser and feedback until the point of the "Monitoring stops" (Day 117), after which they stopped receiving feedback and therefore became identical to dispenser only households.

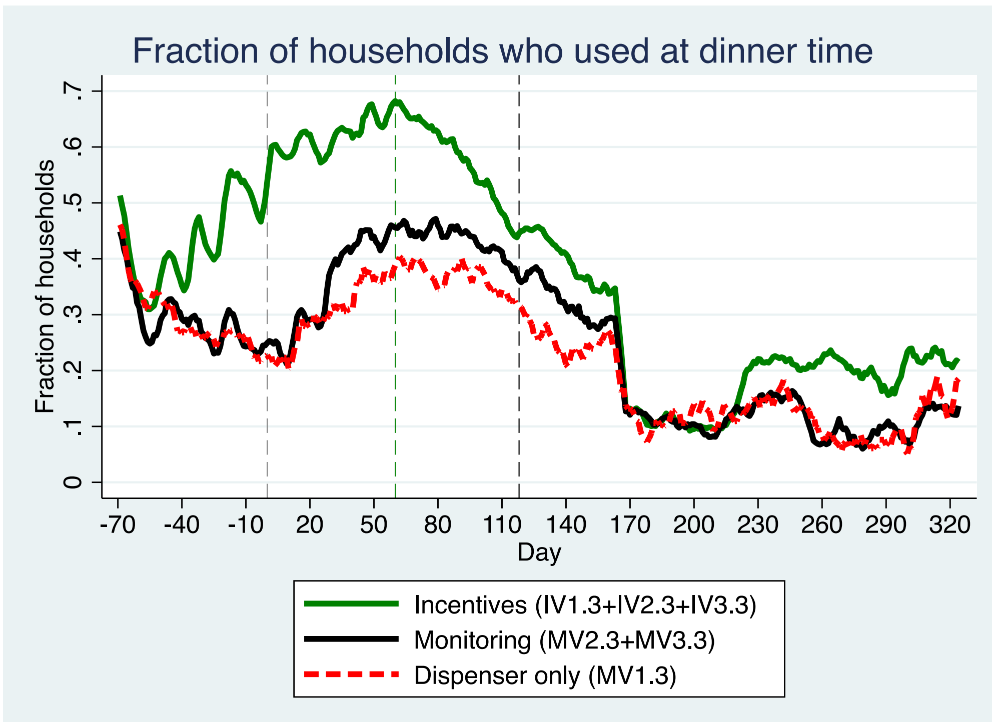

Figure 8. : Time trends across treatment arms

Note: Figure shows the five day moving average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Red dashed line represents households who received the dispenser only; black line represents households who received the dispenser only until Day 0 (gray vertical dashed line) after which they additionally received feedback/monitoring; green line represents households who received the dispenser, feedback, and one ticket for every evening the dispenser was active during the evening mealtime. Tickets and feedback were stopped for this group on Day 60 (green vertical dashed line) and feedback was stopped for the black group on Day 117 (black vertical dashed line).

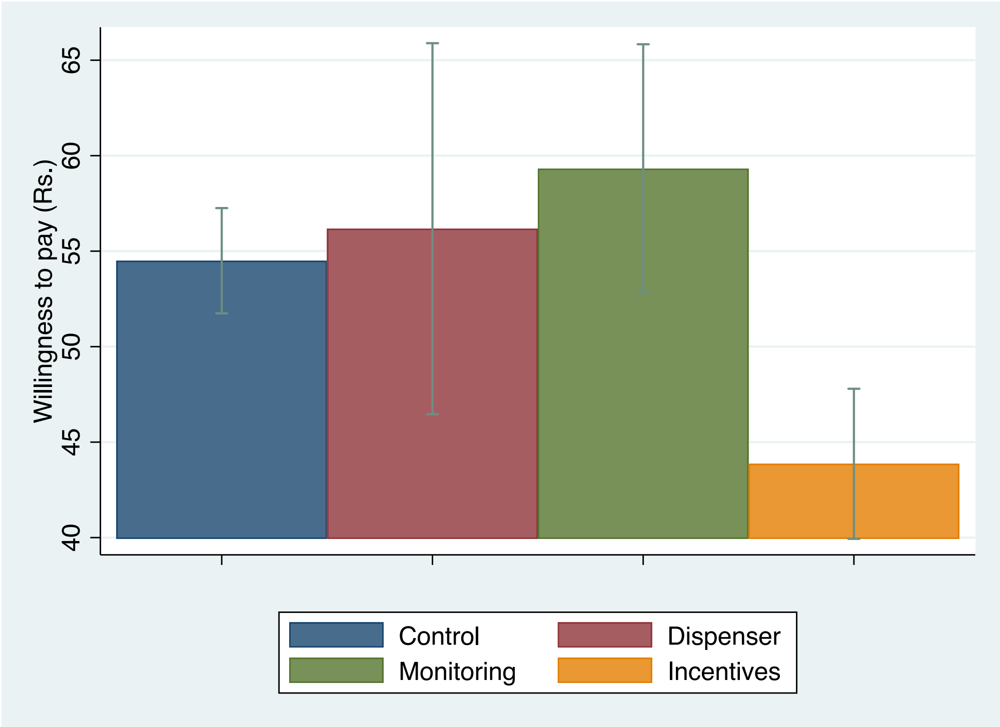

Figure 9. : Willingness to pay for soap

Note: Figure plots the average willingness to pay (WTP) for soap by treatment arm with standard errors in gray. Rupee to USD exchange rate is approximately 65:1. WTP was collected eight months after rollout in using a BDM mechanism in which households chose between a one month soap supply and various household items of increasing (and commonly known) market value.

Appendix Tables

Table 1—: Balance table for disaggregated treatments

	(1) Control mean	(2) Incentives mean	(3) p-value	(4) N	(5) Control mean	(6) Incentives mean	(7) p-value	(8) N	(9) Control mean	(10) Incentives mean	(11) p-value	(12) N
Panel A: Household												
Access to electricity	0.953	0.955	0.977	1918	0.955	0.957	0.837	840	0.955	0.969	0.534	620
Daily labor occupation	0.534	0.544	0.618	1920	0.566	0.576	0.947	840	0.566	0.488	0.188	620
Agriculture occupation	0.224	0.211	0.626	1920	0.201	0.213	0.870	840	0.201	0.173	0.761	620
Number of rooms	2.150	2.190	0.456	1918	1.843	1.899	0.454	839	1.843	1.953	0.457	619
Deep tubewell drinking source	0.548	0.534	0.487	1920	0.584	0.608	0.112	840	0.584	0.583	0.575	620
Distance to drinking source (min)	8.991	9.406	0.149	1919	9.957	9.916	0.997	838	9.957	10.528	0.582	619
Latrine	0.361	0.320	0.046	1919	0.424	0.435	0.430	840	0.424	0.488	0.282	620
Mobile	0.773	0.745	0.188	1919	0.761	0.769	0.699	841	0.761	0.795	0.340	621
Breakfast start hour	8.047	8.064	0.508	1914	7.979	8.072	0.062	840	7.979	8.118	0.044	620
Lunch start hour	12.899	12.945	0.287	1914	12.967	12.960	0.826	840	12.967	13.008	0.922	620
Dinner start hour	20.377	20.372	0.845	1920	20.340	20.416	0.206	841	20.340	20.327	0.955	621
Panel B: Hygiene and sanitation												
Cold can spread	0.629	0.615	0.827	1919	0.567	0.582	0.716	841	0.567	0.622	0.704	621
Soap cleans germs from hands	0.945	0.944	0.837	1920	0.947	0.945	0.819	841	0.947	0.953	0.807	621
Number of times hands washed	2.744	2.738	0.974	1920	2.593	2.631	0.274	841	2.593	2.591	0.299	621
Open defecation practiced	0.696	0.716	0.334	1920	0.652	0.633	0.444	840	0.652	0.583	0.209	621
Panel C: Mother												
Age (years)	31.670	31.698	0.891	1920	31.549	32.360	0.117	841	31.549	31.102	0.749	621
Education (years completed)	5.974	5.670	0.113	1919	6.154	6.415	0.383	840	6.154	6.850	0.124	620
Hindu	0.772	0.769	0.981	1919	0.613	0.620	0.503	840	0.613	0.551	0.010	620
General caste	0.291	0.260	0.084	1916	0.449	0.461	0.617	839	0.449	0.551	0.013	619
Age at marriage	16.333	16.398	0.530	1908	16.609	16.948	0.020	833	16.609	16.969	0.333	616
People listen	2.986	3.035	0.296	1919	3.032	3.101	0.285	840	3.032	3.031	0.938	621
Mother makes child health decision	3.478	3.355	0.119	1916	3.037	2.962	0.540	839	3.037	2.961	0.892	620
Panel D: Children below 11 years												
Age of child (months)	69.852	69.214	0.401	3205	68.485	69.459	0.245	1386	68.485	70.388	0.479	1050
Male child	0.503	0.500	0.844	3211	0.492	0.500	0.603	1386	0.492	0.442	0.333	1050
Height (cm)	105.143	105.147	0.932	3201	103.795	105.227	0.058	1383	103.795	106.014	0.157	1047
Weight (kg)	15.343	15.217	0.624	3201	14.920	15.041	0.497	1382	14.920	15.496	0.304	1048
Preventive check-up (no. of times 6 mo.)	0.768	0.779	0.702	1159	0.722	0.538	0.552	504	0.722	0.685	0.440	382
Sick doctor visit (no. of times 6 mo.)	1.678	1.827	0.422	1131	1.621	1.778	0.420	490	1.621	1.630	0.530	374
Had cold in the last two weeks	0.337	0.364	0.156	3208	0.400	0.430	0.378	1384	0.400	0.394	0.423	1045
Had cough in the last two weeks	0.082	0.086	0.713	3151	0.062	0.087	0.208	1384	0.062	0.074	0.309	1048
Had diarrhea in last two weeks	0.044	0.056	0.240	3208	0.057	0.059	0.942	1386	0.057	0.073	0.439	1050
Exclusively breastfed (no. of months)	4.520	4.653	0.337	2163	5.135	4.582	0.027	899	5.135	4.421	0.020	680

Notes: p-values computed in a regression of the variable on treatment assignment with village level fixed effects.

Table 2—: Balance table for analysis comparisons

	Extensive margin incentives				Intensive margin incentives				Monitoring			
	(1) MV1	(2) IV1	(3) p-value	(4) N	(5) IV1	(6) IV2	(7) p-value	(8) N	(9) MV1	(10) MV2	(11) p-value	(12) N
Panel A: Household												
Access to electricity	0.954	0.946	0.185	2076	0.930	0.950	0.032	360	0.969	0.971	0.260	244
Daily labor occupation	0.539	0.527	0.355	2077	0.554	0.557	0.534	361	0.488	0.549	0.156	244
Agriculture occupation	0.215	0.182	0.242	2077	0.188	0.199	0.621	361	0.173	0.193	0.511	244
Number of rooms	2.057	2.083	0.880	2074	2.173	2.153	0.376	360	1.953	1.943	0.731	244
Deep tubewell drinking source	0.561	0.543	0.534	2077	0.516	0.535	0.167	361	0.583	0.574	0.664	244
Distance to drinking source (min)	9.348	9.818	0.137	2075	9.333	9.565	0.895	361	10.528	9.947	0.081	243
Latrine	0.386	0.361	0.016	2076	0.274	0.310	0.072	361	0.488	0.447	0.949	244
Mobile	0.770	0.741	0.086	2077	0.704	0.726	0.423	361	0.795	0.770	0.781	244
Breakfast start hour	8.033	8.038	0.948	2069	7.984	8.058	0.290	359	8.118	8.168	0.565	244
Lunch start hour	12.926	12.954	0.549	2069	12.916	12.951	0.902	359	13.008	12.986	0.949	244
Dinner start hour	20.362	20.331	0.681	2075	20.333	20.316	0.599	361	20.327	20.381	0.510	244
Panel B: Hygiene and sanitation												
Cold can spread	0.610	0.633	0.579	2078	0.640	0.653	0.896	360	0.622	0.615	0.507	244
Soap cleans germs from hands	0.947	0.939	0.404	2078	0.930	0.953	0.066	361	0.953	0.943	0.396	244
Number of times hands washed	2.693	2.677	0.752	2078	2.737	2.737	0.695	361	2.591	2.623	0.838	244
Open defecation practiced	0.677	0.703	0.007	2078	0.785	0.742	0.103	361	0.583	0.615	0.817	244
Panel C: Mother												
Age (years)	31.608	31.319	0.281	2078	31.468	31.717	0.587	361	31.102	31.295	0.224	244
Education (years completed)	6.081	5.700	0.002	2076	4.914	5.493	0.059	361	6.850	6.680	0.854	244
Hindu	0.715	0.712	0.054	2076	0.823	0.751	0.002	361	0.551	0.594	0.116	244
General caste	0.351	0.361	0.055	2073	0.231	0.258	0.314	361	0.551	0.500	0.094	244
Age at marriage	16.454	16.578	0.885	2064	16.312	16.391	0.853	361	16.969	16.864	0.681	243
People listen	3.001	3.000	0.656	2077	2.978	3.044	0.542	361	3.031	3.070	0.226	244
Mother makes child health decision	3.328	3.131	0.123	2074	3.247	3.269	0.892	360	2.961	2.922	0.839	244
Panel D: Children below 11 years												
Age of child (months)	69.488	70.931	0.209	3459	71.285	70.072	0.110	625	70.388	68.610	0.783	395
Male child	0.497	0.473	0.995	3464	0.494	0.496	0.918	625	0.442	0.473	0.693	395
Height (cm)	104.823	105.374	0.860	3457	104.959	105.276	0.981	624	106.014	105.040	0.947	394
Weight (kg)	15.238	15.625	0.386	3458	15.708	15.483	0.513	624	15.496	15.144	0.823	394
Preventive check-up (no. of times 6 mo.)	0.750	0.777	0.376	1234	0.835	0.850	0.534	234	0.685	0.620	0.264	150
Sick doctor visit (no. of times 6 mo.)	1.663	1.829	0.320	1200	1.956	1.785	0.144	233	1.630	1.871	0.131	147
Had cold in the last two weeks	0.358	0.366	0.409	3457	0.348	0.363	0.939	625	0.394	0.426	0.574	392
Had cough in the last two weeks	0.076	0.084	0.545	3417	0.091	0.091	0.729	613	0.074	0.084	0.861	392
Had diarrhea in last two weeks	0.049	0.059	0.456	3464	0.051	0.051	0.963	623	0.073	0.063	0.261	396
Exclusively breastfed (no. of months)	4.678	4.724	0.490	2297	4.909	4.779	0.911	417	4.421	4.506	0.605	258

Notes: p-values computed in a regression of the variable on treatment assignment with village level fixed effects.

Table 3—: Balance table for analysis comparisons (cont'd)

	RA Incentives				RA Monitoring			
	(1) IV1	(2) IV3	(3) p-value	(4) N	(5) MV1	(6) MV3	(7) p-value	(8) N
Panel A: Household								
Access to electricity	0.968	0.964	0.883	490	0.969	0.957	0.014	361
Daily labor occupation	0.561	0.541	0.603	491	0.611	0.576	0.461	361
Agriculture occupation	0.212	0.220	0.449	491	0.214	0.213	0.310	361
Number of rooms	2.127	2.197	0.209	491	1.947	1.899	0.339	361
Deep tubewell drinking source	0.566	0.541	0.354	491	0.580	0.608	0.276	361
Distance to drinking source (min)	9.767	9.434	0.259	491	9.300	9.916	0.123	360
Latrine	0.349	0.338	0.825	491	0.397	0.435	0.633	361
Mobile	0.741	0.761	0.306	491	0.733	0.769	0.538	361
Breakfast start hour	8.119	8.096	0.107	487	8.188	8.072	0.715	358
Lunch start hour	13.005	12.956	0.830	487	12.988	12.960	0.858	358
Dinner start hour	20.285	20.387	0.059	488	20.410	20.416	0.633	358
Panel B: Hygiene and sanitation								
Cold can spread	0.644	0.605	0.897	490	0.580	0.582	0.571	361
Soap cleans germs from hands	0.979	0.950	0.047	491	0.939	0.945	0.880	361
Number of times hands washed	2.730	2.738	0.899	491	2.656	2.631	0.533	361
Open defecation practiced	0.693	0.690	0.350	491	0.649	0.633	0.887	360
Panel C: Mother								
Age (years)	32.042	31.788	0.397	491	31.641	32.360	0.552	361
Education (years completed)	6.116	5.964	0.591	491	6.466	6.415	0.565	361
Hindu	0.672	0.748	0.033	491	0.641	0.620	0.628	361
General caste	0.302	0.271	0.790	490	0.450	0.461	0.551	361
Age at marriage	16.519	16.432	0.212	488	16.785	16.948	0.649	358
People listen	3.106	3.057	0.497	491	3.099	3.101	0.791	360
Mother makes child health decision	3.303	3.397	0.086	490	2.939	2.962	0.575	360
Panel D: Children below 11 years								
Age of child (months)	68.587	68.412	0.942	836	66.493	69.459	0.039	562
Male child	0.505	0.503	0.907	837	0.517	0.500	0.536	562
Height (cm)	105.423	105.220	0.913	831	103.858	105.227	0.070	561
Weight (kg)	15.245	15.025	0.541	831	14.799	15.041	0.305	559
Preventive check-up (no. of times 6 mo.)	0.858	0.759	0.726	323	0.576	0.538	0.808	203
Sick doctor visit (no. of times 6 mo.)	1.661	1.780	0.153	321	2.122	1.778	0.597	197
Had cold in the last two weeks	0.380	0.370	0.665	836	0.455	0.430	0.575	562
Had cough in the last two weeks	0.092	0.084	0.352	822	0.096	0.087	0.842	560
Had diarrhea in last two weeks	0.052	0.058	0.556	834	0.052	0.059	0.652	562
Exclusively breastfed (no. of months)	4.654	4.552	0.883	560	4.625	4.582	0.857	359

Notes: p-values computed in a regression of the variable on treatment assignment with village level fixed effects.

Table 4—: Attrition trends in dispenser data

	(1)	(2)	(3)
	Whether household used at dinnertime (0-9 months)		
Attrited v. remaining sample	-0.0202	-0.0438	-0.0836
	(0.0150)	(0.0493)	(0.0584)
Monitoring treatment effect - attrited v. remaining sample		0.0380	
		(0.0497)	
Incentive treatment effect - attrited v. remaining sample			0.0636
			(0.0618)
Mean of remaining sample	0.4016	0.3282	0.4314
	[0.4902]	[0.4696]	[0.4953]
N	243665	98037	172316

Notes: Observations are at the household-day level. Robust standard errors in parentheses and clustered at the village level. All regressions include village and day fixed effects. is made up of the households who did have sensor data collected in the last month of the experiment; is made up of those that did not. Outcome variable is a binary variable that equals one if the household used the dispenser during dinner time and zero otherwise; these outcomes are drawn from the first nine months of the experiment. * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 5—: Persistence in rational habit stock

	(1) Day 0 to 59	(2) Day 0 to 116	(3) Day 60 to 329	(4) Day 117 to 329
	Likelihood of using during reported dinnertime:			
	During intervention		After withdrawal of intervention	
Anticipated tripled incentive (IV3.2 or IV3.3)	0.0200		-0.0247	
	(0.024)		(0.018)	
Anticipated monitoring (MV3.2 or MV3.3)		0.0272		0.0030
		(0.024)		(0.019)
Mean of pure control	0.619	0.376	0.351	0.184
	[0.486]	[0.484]	[0.477]	[0.387]
Comparison group				
N	24832	33767	76412	33365

Notes: "Treated" pools all households that received a dispenser. "Pure control" are households who did not receive a dispenser. p-values computed in a regression of the variable on treatment assignment with village level fixed effects. *Notes:* Observations are at the household-day level. Robust standard errors in parentheses and clustered at the village level. Standard deviation in brackets. All regressions include fixed effects for day. All regressions include fixed effects for day. Households in the one ticket daily incentive group are compared to households in the dispenser only group. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). One MHT family: the 8 coefficients of the effect of the IV1.2 treatment on handwashing outcomes (Columns 1-4 of Panels A and B).

Table 6—: Daily child diarrhea and ARI outcomes—ITT estimates

	(1)	(2)	(3)	(4)
	Whether child had diarrhea		Whether child had ARI	
Received dispenser	-0.0010		-0.0182***	
	(0.0008)		(0.0063)	
Incentives		-0.0003		-0.0128
		(0.0010)		(0.0077)
Monitoring		-0.0023*		-0.0287**
		(0.0014)		(0.0118)
Dispenser only		-0.0023		-0.0261
		(0.0017)		(0.0157)
Mean of pure control	0.0045	0.0045	0.1458	0.1458
	[0.0670]	[0.0670]	[0.3529]	[0.3529]
N	112737	112737	112737	112737

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. All regressions include village and day fixed effects, and the following baseline child health controls: child age, child sex, baseline height, baseline weight, baseline mid-arm circumference, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. Biweekly child health data spans February and March of 2016 (4-5 months after rollout). All treatment effects are estimated relative to the pure control group. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Four MHT families: the 2 coefficients of the effect of the Received dispenser treatment on the diarrhea and ARI outcomes (Columns 1 and 3); the 2 coefficients of the effect of the Incentives treatment on the diarrhea and ARI outcomes (Columns 2 and 4, second row); the 2 coefficients of the effect of the Monitoring treatment on the diarrhea and ARI outcomes (Columns 2 and 4, third row); and the 2 coefficients of the effect of the Dispenser only treatment on the diarrhea and ARI outcomes (Columns 2 and 4, fourth row). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 7—: Attrition trends for midline survey

	(1) Atritted* Treatment Household	(2) p-value	(3) N
Panel A: Household			
Access to electricity	0.0601**	0.0452	2884
Daily labor occupation	0.0672	0.2713	2886
Agriculture occupation	-0.0614	0.2179	2886
Number of rooms	-0.3252*	0.0810	2883
Deep tubewell drinking source	0.0551	0.1903	2886
Distance to drinking source (min)	-1.2872	0.2268	2883
Latrine	-0.0319	0.5606	2885
Mobile	-0.0312	0.5761	2886
Breakfast start hour	0.1206	0.2329	2887
Lunch start hour	-0.0381	0.6293	2887
Dinner start hour	0.0523	0.6017	2887
Panel B: Hygiene and sanitation			
Cold can spread	-0.0439	0.6434	2886
Soap cleans germs from hands	-0.0286	0.1040	2887
Number of times hands washed	0.0404	0.3217	2887
Open defecation practiced	0.0275	0.5983	2886
Panel C: Mother			
Age (years)	0.6434	0.5047	2886
Education (years completed)	-0.1617	0.6223	2885
Hindu	0.0089	0.8123	2885
General caste	-0.0531	0.2523	2881
Age at marriage	-0.3282	0.2659	2868
People listen	0.0809	0.5083	2885
Mother makes child health decision	0.3817	0.1033	2881
Panel D: Children			
Age of child (months)	0.5236	0.8623	4498
Male child	0.0066	0.9049	4504
Height (cm)	-0.4584	0.7620	4492
Weight (kg)	0.0917	0.8502	4492
Preventive check-up (no. of times 6 mo.)	-0.0517	0.8606	1677
Sick doctor visit (no. of times 6 mo.)	-0.1644	0.4341	1637
Had cold in the last two weeks	-0.0302	0.5719	4496
Had cough in the last two weeks	0.0015	0.9473	4447
Had diarrhea in last two weeks	-0.0134	0.6287	4501
Exclusively breastfed (no. of months)	0.3874	0.2869	3004

Notes: "Attrited" are those households we could not reach at endline. "Treated households" is any household that was randomized to a treatment arm (dispenser control, monitoring or incentives). p-values computed in a regression of the baseline value of the variable on treatment assignment, whether or not the household had attrited, and an interaction of the two with village level fixed effects and standard errors clustered at the household level. Only the interaction coefficient is reported. * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 8—: Child health after eight months disaggregated by treatment arm

	(1) Whether child had any loose stool in last two weeks	(2) Total days of loose stool in last two weeks	(3) Whether child showed any ARI symptoms in last two weeks	(4) Total days of ARI in last two weeks
Incentives	-0.0156	-0.0588	-0.0234	-0.1111
	(0.014)	(0.034)	(0.021)	(0.118)
Monitoring	-0.0354*	-0.1120*	-0.0664*	-0.3434*
	(0.019)	(0.051)	(0.030)	(0.182)
Dispenser only	-0.0594**	-0.1150*	-0.0974**	-0.3812*
	(0.025)	(0.068)	(0.039)	(0.239)
Mean of pure control	0.099	0.207	0.269	1.245
	[0.299]	[0.802]	[0.444]	[2.458]
N	3333	3342	3342	3342

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Data was collected six to seven months after rollout. "Whether child showed any ARI symptoms" equals one if the child experienced any of the following in the two weeks priod: runny nose, nasal congestion, cough (with or without sputum production), ear discharge, hoarseness of voice, sore throat, difficulty breathing or a prescription from a doctor for such. Baseline controls include: child age, child sex, baseline height, baseline weight, baseline mid-arm circumference, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Three MHT families: the 4 coefficients of the effect of the Incentives treatment on the diarrhea and ARI outcomes (Columns 1-4, first row); the 4 coefficients of the effect of the Monitoring treatment on the diarrhea and ARI outcomes (Columns 1-4, second row); and the 4 coefficients of the effect of the Dispenser only treatment on the diarrhea and ARI outcomes (Columns 1-4, third row). * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 9—: Child anthropometric outcomes after eight months disaggregated by treatment arm

	(1) Weight-for-age z-score	(2) Height-for-age z-score	(3) Mid-arm circ.-for-age z-score
Incentives	0.1155	0.1806	0.0054
	(0.079)	(0.102)	(0.060)
Monitoring	0.1773	0.2802	0.1454
	(0.124)	(0.190)	(0.104)
Dispenser only	0.1289	0.2627	0.2366
	(0.181)	(0.334)	(0.138)
Mean of pure control	-2.167	-1.866	-1.365
	[1.087]	[1.573]	[0.990]
N	852	851	847

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Dependent variables calculated using WHO anthropometric methodology. Sample is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Data was collected eight months after rollout. Baseline controls include: child age, child sex, baseline HAZ, baseline WAZ, baseline MAZ, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. "Incentives" is the pooled sample of all households in the standard incentive arm, surprised three ticket arm, and anticipated three ticket arm. "Monitoring" is the pooled sample of all households in the surprised monitoring arm and anticipated monitoring arm. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Three MHT families: the 3 coefficients of the effect of the Incentives treatment on the anthropometric outcomes (Columns 1-3, first row); the 3 coefficients of the effect of the Monitoring treatment on the anthropometric outcomes (Columns 1-3, second row); and the 3 coefficients of the effect of the Dispenser only treatment on the anthropometric outcomes (Columns 1-3, third row). * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 10—: Child anthropometric outcomes after eight months disaggregated by age

	(1) 1-12 months	(2) 13-24 months	(3) 25-36 months	(4) 37-48 months	(5) 39-60 months
Panel A: Weight-for-age z-score					
Treated household	-0.3416	0.1347	0.4018	-0.0450	0.0475
	(0.274)	(0.214)	(0.161)	(0.121)	(0.177)
Mean of pure control	-1.893	-2.151	-2.225	-2.152	-2.310
	[1.213]	[1.232]	[1.072]	[0.930]	[1.047]
With baseline controls					
N	86	177	198	260	131
Panel B: Height-for-age z-score					
Treated household	-0.5573	0.3896	0.3441	0.1061	-0.1043
	(0.490)	(0.376)	(0.170)	(0.155)	(0.260)
Mean of pure control	-2.118	-2.058	-1.907	-1.691	-1.680
	[2.119]	[2.039]	[1.289]	[1.113]	[1.449]
With baseline controls					
N	86	177	198	260	130
Panel C: Mid-arm circumference for age z-score					
Treated household	0.0868	0.3099	0.0905	-0.0112	0.0778
	(0.265)	(0.190)	(0.113)	(0.101)	(0.170)
Mean of pure control	-0.905	-1.353	-1.461	-1.369	-1.539
	[1.063]	[0.992]	[0.924]	[1.025]	[0.894]
With baseline controls					
N	85	177	199	259	127

Notes: Observations are at the child level. Standard deviation in brackets. Dependent variables calculated using WHO anthropometric methodology. Sample is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Data was collected eight months after rollout. Baseline controls include: child age, child sex, baseline HAZ, baseline WAZ, baseline MAZ, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). One MHT family: the 15 coefficients of the effect of the Household treatment variable on the anthropometric outcomes across different ages (Columns 1-3 of Panels A, B and C). * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 11—: Forecasted handwashing performance

	(1) Pre-change period	(2) Pre-change period	(3) Intervention period	(4) Intervention period	(5) Full experiment	(6) Full experiment
	Number of forecasted handwashing days in the upcoming week					
Anticipation of monitoring	0.2301*					
	(0.1241)					
Anticipation of triple tickets		0.0036				
		(0.0735)				
Monitoring intervention			0.0200			
			(0.0606)			
Triple ticket intervention				0.0354		
				(0.0663)		
Treated household					3.2977***	
					(0.0582)	
Incentives						3.2805***
						(0.0759)
Monitoring						3.3726***
						(0.0928)
Dispenser only						3.2121***
						(0.1266)
Mean of pure control	5.726	6.228	6.317	6.366	2.964	2.964
	[1.901]	[1.579]	[1.220]	[1.381]	[2.786]	[2.786]
Comparison group						
N	455	655	2125	1272	21019	21019

Notes: Observations at household level with forecasts collected approximately twice a month. Robust standard errors are in parentheses and are clustered at the household level. Respondents were asked during each biweekly health survey: . represents days -21 to -1. represents days 30 to 60 for triple ticket households and days 30 to 110 for monitoring households. Full experiment includes only those days for which we have forecasting data, which is days -21 to 110. Control group for column 1 is the group that was not anticipating monitoring; for column 2 is the group not anticipating a ticket boost; for column 3 is the dispenser control group; for column 4 is the standard one ticket incentive group; and for columns 5-6 is the pure control group. * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 12—: Alternative hygiene measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Observed hand cleanliness		Observed nail cleanliness		Whether handwashing habit was achieved		Whether household has non-project liquid soap	
Received dispenser	0.0403***		0.0589***		0.5187***		0.0459***	
	(0.0161)		(0.0166)		(0.0159)		(0.0093)	
Incentives		0.0286**		0.0700***		0.5598***		0.0622***
		(0.0203)		(0.0220)		(0.0195)		(0.0116)
Monitoring		0.0699**		0.0401*		0.4370***		0.0155
		(0.0291)		(0.0265)		(0.0291)		(0.0167)
Dispenser only		0.0388		0.0352		0.4641***		0.0193
		(0.0393)		(0.0390)		(0.0412)		(0.0265)
Mean of pure control	0.6550	0.6550	0.4453	0.4453	0.2566	0.2566	0.0544	0.0544
	[0.4755]	[0.4755]	[0.4972]	[0.4972]	[0.4369]	[0.4369]	[0.2270]	[0.2270]
N	2629	2629	2628	2628	2626	2626	2627	2627

Notes: Observations are at the household level in columns 1-4 and at the child-day level in column 5. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. "Received dispenser" is the pooled sample of incentive, monitoring and dispenser control households. Coefficients are reported from two separate regressions: the first pools al dispenser households ("Received dispenser" row); the second includes covariates for each treatment arm (Incentives, Monitoring and Dispenser only). All regressions include village level fixed effects. The relevant comparison group is the pure control. Observed hand and nail cleanliness are graded by the enumerator on a three-point Likert scale with 1 indicating no visible dirt, 2 indicating some visible dirt, and 3 indicating extensive visible dirt. Whether a handwashing habit was achieved is rated by the respondent on a five-item scale as follows: 0 ="How? You did not give us soap."; 1 ="No, not at all."; 2 ="No, not yet, but it is growing"; 3 ="Yes, mostly, but still needs time.";4 ="Yes, definitely, the habit has been established. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Four MHT families: the 4 coefficients of the effect of the Received dispenser treatment on the alternative hygiene outcomes (Columns 1, 3, 5 and 7); the 4 coefficients of the effect of the Incentives treatment on the alternative hygiene outcomes (Columns 2, 4, 6 and 8, second row); the 4 coefficients of the effect of the Monitoring treatment on the alternative hygiene outcomes (Columns 2, 4, 6 and 8, third row); and the 4 coefficients of the effect of the Dispenser only treatment on the anthropometric outcomes (Columns 2, 4, 6 and 8, fourth row). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 13—: Sanitation outcomes

	(1)	(2)	(3)	(4)
	Whether household defecates in open		Whether household treats drinking water	
Received dispenser	0.0005		0.0069	
	(0.0176)		(0.0102)	
Incentives		0.0235		0.0097
		(0.0215)		(0.0129)
Monitoring		-0.0406		0.0023
		(0.0327)		(0.0179)
Dispenser only		-0.0432		0.0004
		(0.0477)		(0.0299)
Mean of pure control	0.6481	0.6481	0.0857	0.0857
	[0.4777]	[0.4777]	[0.2800]	[0.2800]
N	2629	2629	2626	2626

Notes: Observations are at the household level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. ”Received dispenser” is the pooled sample of incentive, monitoring and dispenser control households. All regressions include village level fixed effects. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Four MHT families: the 2 coefficients of the effect of the Received dispenser treatment on sanitation outcomes (Columns 1 and 3); the 2 coefficients of the effect of the Incentives treatment on sanitation outcomes (Columns 2 and 4, second row); the 2 coefficients of the effect of the Monitoring treatment on sanitation outcomes (Columns 2 and 4, third row); and the 2 coefficients of the effect of the Dispenser only treatment on sanitation outcomes (Columns 2 and 4, fourth row). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 14—: Willingness to pay for soap at six months

	(1)	(2)	(3)
	Willingness to pay (Rs.)		
Treated household	-6.097***		
	(1.844)		
Incentives		-8.4836***	-7.4705***
		(2.2077)	(2.3802)
Monitoring		-2.1379	-2.7452
		(3.6382)	(3.7585)
Dispenser only		-1.0546	-0.9132
		(4.6829)	(4.9048)
Mean of pure control	54.498	54.498	53.190
	[56.646]	[56.646]	[55.772]
N	2667	2667	2439

Notes: Observations are at the household level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. "Received dispenser" is the pooled sample of incentive, monitoring and dispenser control households. All regressions include village level fixed effects. Column 3 restricts sample to those households who do not report having non-project related liquid soap in the household during the midline survey. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Three MHT families: the 2 coefficients of the effect of the Incentives treatment on willingness to pay (Columns 2-3, second row); the 2 coefficients of the effect of the Monitoring treatment on willingness to pay (Columns 2-3, third row); and the 2 coefficients of the effect of the Dispenser only treatment on willingness to pay (Columns 2-3, fourth row). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 15—: Spillovers in handwashing rates

	(1) Days -40 to -30	(2) Days 40 to 50	(3) Days 120 to 130
	Likelihood of using during reported dinnertime		
No. of monitored households	-0.0077	0.0155	0.0085
	(0.0065)	(0.0084)	(0.0087)
Mean of pure control	0.212	0.382	0.254
	[0.410]	[0.487]	[0.437]
Comparison group			
N	1107	1154	1009

Notes: Observations are at the household level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Sample is all dispenser control households. Independent variable is the number of monitored households within 1 km of the dispenser control household. All regressions include village and day level fixed effects. Comparison group is dispenser only households who have zero monitored households within a one-kilometer radius. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). One MHT family: the 3 coefficients of the effect of the No. of monitored households on handwashing outcomes (Columns 1-3). * p$\leq$ 0.10, ** p$\leq$ 0.05, *** p$\leq$ 0.01.

Table 16—: Health spillovers

	(1) Whether child had any loose stool in last two weeks	(2) Total days of loose stool in last two weeks	(3) Whether child showed any ARI symptoms in last two weeks	(4) Total days of ARI in last two weeks
No. of dispenser households	-0.0026	-0.0073	-0.0021	-0.0288
	(0.002)	(0.005)	(0.004)	(0.023)
Mean of pure control	0.078	0.177	0.215	1.468
	[0.270]	[0.874]	[0.414]	[3.304]
With baseline controls				
N	548	553	553	553
No. of dispenser households	-0.0007	-0.0015	-0.0054	-0.0317
	(0.003)	(0.006)	(0.004)	(0.018)
Mean of pure control	0.095	0.215	0.264	1.239
	[0.294]	[0.854]	[0.441]	[2.440]
With baseline controls				
N	1385	1386	1386	1386

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Sample is composed of the children in pure control households in each type of village (monitoring village or incentive village). Independent variable is the number of households who received a dispenser (monitoring and dispenser only households for monitoring villages; incentivized households for incentive villages) within 1 km of the pure control household. Comparison group is made up of pure control households who have no dispenser receiving households within a one km radius of itself. p-values adjusted for multiple hypothesis (MHT) testing using Benjamini et al. (2006). One MHT family: the 4 coefficients of the effect of the No. of dispenser households on diarrhea and ARI outcomes (Columns 1-4). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 17—: Learning about health (midline data)

	(1) Dispenser only	(2) Dispenser only	(3) Monitoring	(4) Monitoring	(5) Incentives	(6) Incentives
	Average likelihood of handwashing at dinnertime one month after withdrawal of interventions					
Health index (Incidence)	-0.0115		-0.0058		0.0066	
	(0.0388)		(0.0136)		(0.0088)	
Health index (Anthro)		-0.0479		0.0640		-0.0072
		(0.0530)		(0.0609)		(0.0639)
Mean of dep. var.	0.2668	0.2668	0.3084	0.3084	0.6399	0.6399
	[0.2654]	[0.2654]	[0.2894]	[0.2894]	[0.2794]	[0.2794]
N	123	33	309	48	680	105

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. All regressions include village level fixed effects and controls for the average likelihood of washing during dinnertime during the course of the intervention, baseline health index, child sex, child age, number of months the child was breastfed, household occupation, number of rooms, mother's age at marriage, and mother's education. Health index is constructed using Anderson (2008); the "Incidence" index is constructed as a weighted average of the child being free of loose stool or ARI in the two weeks prior to surveying and the number of days she was free of these illnesses; the "Anthro" index is constructed using child height, weight and mid-arm circumference z-scores. Therefore, a higher health index implies better health. The dependent variable is the average likelihood of the dispenser being active during dinnertime over the course of the one month after the withdrawal of monitoring or incentives (the time frame for monitoring is also applied to the dispenser only group). Columns 1, 3 and 5 include all children below the age of 14 years; columns 2, 4 and 6 include only children 60 months and below. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Three MHT families: the 2 coefficients of the effect of the health indices on handwashing among dispenser only households (Columns 1-2); the 2 coefficients of the effect of the health indices on handwashing among monitoring households (Columns 2-3); and the 2 coefficients of the effect of the health indices on handwashing among incentives households (Columns 4-5). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 18—: Learning about health (panel data)

	(1) Week 7	(2) Week 8	(3) Week 9	(4) Week 10	(5) Week 11	(6) Week 12
Panel A: Dispenser only and monitored households						
Sick in previous week	-0.1296	0.2139	-0.2342	0.0460	0.8228	-0.0857
	(0.2156)	(0.3039)	(0.3961)	(0.3657)	(0.4724)	(1.0545)
N	359	341	337	304	235	258
Panel B: Dispenser only and incentivized households						
Sick in previous week	-0.3286	-0.3718	0.5577	0.8721	-0.0657	0.4690
	(0.3254)	(0.4468)	(0.4013)	(0.3750)	(0.3916)	(1.0662)
N	577	563	572	496	454	427

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Sample is restricted in each specification to those children who experienced a sickness either in the week prior to handwashing observation or the week after handwashing observation (but not both). All regressions include village level fixed effects and controls for whether or not the child experienced ARI in the week that the handwashing outcome is observed, the total number of ARI incidences up to the week before observation, and the total number of days the dispenser was used up to the week before observation. The dependent variable is the total number of days the dispenser was active during dinnertime during the week of observation. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Two MHT families: the 6 coefficients of the effect of being sick in previous week on handwashing among dispenser only and monitored households (Columns 1-6 of Panel A); and the 6 coefficients of the effect of being sick in previous week on handwashing among dispenser only and incentivized households (Columns 1-6 of Panel B). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

REFERENCES

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

Bennett, Daniel, Asjad Naqvi, and Wolf-Peter Schmidt. 2018. "Learning, Hygiene and Traditional Medicine." *The Economic Journal*, 128(612): F545–F574.

Halder, Amal K., Carole Tronchet, Shamima Akhter, Abbas Bhuiya, Richard Johnston, and Stephen P. Luby. 2010. "Observed Hand Cleanliness and Other Measures of Handwashing Behavior in Rural Bangladesh." *BMC Public Health*, 10(1): 545.

Luby, Stephen P., Amal K. Halder, Tarique M. N. Huda, Leanne Unicomb, and Richard B. Johnston. 2011. "Using Child Health Outcomes to Identify Effective Measures of Handwashing." *The American Journal of Tropical Medicine and Hygiene*, 85(5): 882–892.

Ruel, Marie T., and Mary Arimond. 2002. "Spot-Check Observational Method for Assessing Hygiene Practices: Review of Experience and Implications for Programmes." *Journal of Health, Population and Nutrition*, 20(1): 65–76.