Handwashing and Habit Formation

Reshmaan Hussam*†, Atonu Rabbani‡, Giovanni Reggiani§, and Natalia Rigol¶

September 19, 2016

Abstract

Regular handwashing with soap is believed to have substantial impacts on child health in the developing world. Most handwashing campaigns have failed, however, to establish and maintain a regular practice of handwashing. Motivated by the idea that handwashing is habitual, we design, implement and analyze a randomized field experiment aimed to test the main predictions of the rational addiction model. To reliably measure handwashing, we developed and produced a novel soap dispenser, within which a time-stamped sensor is embedded. The experiment has three arms: a dispenser arm, in which households are provided a soap dispenser; a monitoring arm, which adds biweekly performance reports shared with each household; and an incentive arm, which adds daily incentives. Moreover, a subset of households are informed of a future boost in monitoring or incentives, enabling a test of whether agents behave rationally regarding the habit-forming nature of handwashing. We find that (1) both financial incentives and monitoring without incentives increases handwashing, compared to just receiving a dispenser; (2) these effects persist after incentives or monitoring are removed; (3) the anticipation of monitoring increases handwashing, but not the anticipation of incentives; and (4) all the treatments (even the dispensers alone) increase hand cleanliness and have substantial impacts on child health, confirming previous suggestive literature that handwashing has positive impacts on child health in resource-poor settings. The results are consistent with a model which exhibits both rational addiction and a distinction between intrinsic and extrinsic motivation.

*We are indebted to research assistant Sami Safiullah for his excellent field work and continual valuable insights. We thank the Birbhum Population Project (BIRPOP) of the Society for Health and Demographic Surveillance (SHDS), a Health and Demographic Surveillance Site in West Bengal, India, for support in the field throughout: in particular, the hard working team of surveyors, survey managers, field monitors, data entry operators, data managers, and the research team. We are extremely grateful to Nan-Wei Gong, who designed and developed the dispensor sensors, as well as AQS, the manufacturing company that manufactured them for the field. We received guidance from Esther Duflo, Abhijit Banerjee, Frank Schilbach, Chris Udry, Dean Karlan, Mushfiq Mubarak, Dan Keniston, Nick Ryan, Supreet Kaur, Lasse Brune, Hannah Trachtman, David Jimenez-Gomez, and David Levine, as well as helpful comments from seminar and workshop participants at MIT, Yale, and the University of Cape Town. We are grateful for financial support from the J-PAL Urban Services Initiative, the Weiss Family Fund, the Schultz Fund, USAID, and the Evenson Fund.

†Yale University, Economic Growth Center; corresponding author: reshmaan.hussam@yale.edu

‡Dhaka University Dept. of Economics

§MIT, Dept. of Economics

¶Harvard University, Center for Population and Development

"In the acquisition of a new habit, or the leaving off of an old one, we must take care to launch ourselves with as strong and decided an initiative as possible. Accumulate all the possible circumstances which shall reinforce the right motives; put yourself assiduously in conditions that encourage the new way; make engagements incompatible with the old; take a public pledge, if the case allows; in short, envelop your resolution with every aid you know. This will give your new beginning such a momentum that the temptation to break down will not occur as soon as it otherwise might; and every day during which a breakdown is postponed adds to the chances of its not occurring at all."

- William James, *Habit*, 1914

1 Introduction

Bacterial and viral contamination, resulting in anemia, diarrheal disease, and acute respiratory infection, end the lives of nearly three million children per year and contribute to the severe stunting of millions more. Handwashing with soap is widely regarded as "the most cost effective vaccine" against such deaths (Kremer and Zwane 2007), as it decreases person-to-person transmission and protects the last point of contact between the body and germs (Barker et al. 2004, Sanderson and Weissler 1992, WHO 2009). Despite enormous policy interest and funding invested in hand hygiene campaigns over the last thirty years, however, we know little about how to improve hygiene behavior sustainably. Most public health interventions find zero impact on behavior or health (WSP 2012, WSP 2013, Galiani et al. 2015). The few who do are intensive 'omnibus' interventions (including information, resources, community involvement, and monitoring, and other hygiene and sanitation tips) which make them difficult to replicate in practice or to understand the key mechanism at work (Luby et al. 2005, Bennett et al. 2015, Haggerty et al. 1994, Han and Hliang 1989).

One feature of handwashing that may explain why sustained change is difficult is that, in order to be repeated as often as needed, the behavior must become a habit. For example, 57% of households in our sample (in rural West Bengal) articulate, unprompted, that they do not wash their hands with soap because *"Obhyash nai,"* or "I do not have the habit." These are not unique features of handwashing: preventative health behaviors often require routines. Water should be treated daily; clean cookstoves utilized per meal, medicine consumed at regular intervals, and handwashing engaged in during the same critical moments each day. Agents incur repeated costs from engaging. As such, agents can benefit from these behaviors becoming matters of habit. Most of these preventative health behaviors suffer low rates of takeup in the developing world despite their affordability (Dupas and Miguel 2016). Across the board, neither information provision nor materials and/or infrastructure provision appear to generate sustained improvements in such practices (Clasen et al. 2014, Kremer and Zwane 2007, Banerjee et al. 2010, WSP 2012, WSP 2013).

In this study, we take seriously the hypothesis that handwashing may be habitual, and explore the implications for the design of effective public health interventions. Motivated by the economic theory of rational addiction (Becker and Murphy 1988), we set up and design an experiment that tests the main implications of this model, enriched to take into account a key implication of the fact that the costs and the

returns to the activity are uncertain and individuals may try to infer both from contextual cues (Benabou and Tirole 2003).

Along the way, we develop a novel technology to accurately measure handwashing. In partnership with the MIT Media Lab, we designed a time-stamped sensor technology embedded in a liquid soap dispenser, which we then produced at scale in China at the cost of approximately $30 USD a dispenser. This technology addresses the key problems of standard handwashing outcomes: desirability bias (hand hygiene is conspicuously observed by enumerators), subjectivity (hand cleanliness outcomes are subjectively graded by enumerators), noise (metrics are broad and data collected infrequently), and nonspecificity of behavior (presence and use of barsoap, a common outcome measure, is often due to bathing and laundry rather than handwashing). Our novel sensor is neither visible nor accessible to household, yielding more objective data; it is precise, measuring use at the second level and allowing us to connect observed use with crucial times of use (such as prior to mealtime); and it tracks the use of liquid soap, which is uniquely associated in our study context with handwashing rather than bathing or laundry.

In our conceptual framework, we define habit formation as intertemporally linked preferences in consumption: the more is consumed in the past, the easier or more likely is consumption in the present. This framing was popularized by Becker and Murphy (1988), who also posit the theory of "rational addiction." Rationality implies that agents are aware of the habit-forming nature of a behavior, foresee their future consumption path given the intertemporal complementarities in behavior, and make the decision of whether or not to engage accordingly. If handwashing is indeed a habit-forming activity but agents fail to internalize this feature in their consumption decisions, they will underinvest in the behavior, justifying short-term subsidies to boost usage. Understanding the nature of the behavior (habitual or not) and how an agent conceptualizes the behavior (rationally or otherwise) is thus important for the design of interventions.

A second feature of preventative health behaviors (particularly those with low takeup) is that neither the cost nor the returns to the activity are immediately obvious, and agents may try to infer both from the information in their environment. In this setting, a campaign that incentivizes agents can have countering effects: the incentives, by increasing and solidifying uncertain returns, may increase takeup; however, by potentially signaling through their provision that the activity is a costly one, they may decrease takeup (Benabou and Tirole 2003). Each has distinct implications on the short and long run engagement in activities that require perpetual repetition.

Our experiment is designed to test these two features within the context of handwashing with soap. We draw from the psychology literature on habit formation to make handwashing as amenable to habituation as possible, using the classic habit loop: a trigger (the evening meal), a routine (handwashing) and a reward (incentives or feedback) (Duhigg 2011).

Specifically, we distribute handsoap dispensers with liquid soap and sensor technology to a random

subset of households in our sample. Within this group, the experiment has two arms: in one arm, we offer daily incentives for handwashing in the form of tickets that can be redeemed for household goods; in the other arm, we offer to a subset of dispenser households the same regular feedback on performance, but no incentives. In both cases, feedback and incentives are removed after four months, and we continue to track behavior. Finally, to test whether agents are rational about the habitual nature of the activity, we announce a future increase in incentives (re. a future introduction of monitoring) to a randomized subset of the incentive (re. monitoring) treatment group. A present reaction to anticipated changes in future handwashing behavior would confirm that individuals are aware of the intertemporal complementarities between performance today and performance tomorrow.

The incentives treatment raises short run handwashing rates by up to 70%. After their withdrawal, higher handwashing rates persist for several months, implying that handwashing is indeed a habit-forming activity. However, despite decisive evidence that incentives increase the value of handwashing and are habit-forming in nature, we find no anticipated reaction: households anticipating a tripling of tickets wash no more than those not anticipating this future change.

Two features of the incentives experiment on its own preclude a clean interpretation of the incentives effect and the absence of the rational habit formation effect. Daily incentives are implicitly a combination of two interventions: (1) rewards for performance and (2) monitoring and feedback on performance.[1] *A priori*, we cannot know whether or what fraction of the positive treatment effect is due to the reward itself versus the direct feedback (and the fact that agents know that they are observed). Moreover, in an environment where costs and returns are unknown, agents may infer from the increased incentives that the costs are higher than they thought, thereby countering any effort to invest today in anticipation of a future increase in the incentive. This would diminish our ability to observe a rational habit formation effect even if agents are rationally habit-forming.

The parallel monitoring intervention allows us to disentangle the mechanisms behind the incentive effects we find. As with incentives, our results on habit formation are encouraging. We find that monitoring succeeds in raising short run handwashing rates by up to 23%. These higher handwashing rates persist for several months after the withdrawal of the service. While the short run impact of incentives is significantly higher than that of monitoring, the long run decay rate is shallower amongst [formerly] monitored households. Additionally, we find compelling evidence of rational habit formation among households who were initially anticipating the monitoring service: these households wash 39% more than their nonanticipating counterparts, with the difference increasing as the date of the monitoring service approaches. It appears, therefore, that households do indeed recognize the habit-forming nature of handwashing, and additionally act upon this knowledge in the monitoring setting. In contrast, the absence of evidence of rational habit for-

[1] This applies more broadly to any incentive treatments in the literature where a [conscious/sentient] principal is rewarding an agent for his behavior.

mation in the incentive intervention and the rapid decay after withdrawal suggest an adverse signal regarding the cost of handwashing from the increase in incentives.

Lastly and perhaps most importantly, we examine child health outcomes to establish the causal link between handwashing and child health. We find strong effects on child health, confirming that handwashing alone has substantial returns in resource-poor settings. Children in households that received a dispenser and soap (regardless of whether they also received feedback and incentives) report 39.5% fewer days of loose stool (a proxy for diarrhea episodes) and 23% fewer days of acute respiratory infection (ARI) eight months after the distribution of the dispensers (intent to treat estimates).[2] These effects rise to 83% fewer days of loose stool and 31.5% fewer days of ARI when we examine the impact of the treatment on the treated, where 'treated' is defined as those for whom the dispenser was active during their reported evening mealtime. These reductions in morbidity translate to improvements in child weight-for-age and height-for-age: treated children experience a 9-19% increase (ITT-TOT estimates) in their height-for-age and weight-for-age Z-scores eight months after dispenser distribution.

This study makes four contributions. First, to our knowledge, this is the first field experiment designed to test for rational habit formation. Existing literature that tests for rational habit formation employ time series data and suffer serious identification issues: price instruments are endogenous, knowledge of future changes in price are implausible, and serial correlation in prices yields false positives in evidence of rational addiction (Auld and Grootendorst 2004). The experimental design of our study systematically addresses each issue that has previously challenged causality. This is also the first experiment to examine rational habit formation in the context of *good* habits, an important feature of preventative health behaviors.

Second, ours is the first study (experimental or otherwise) to examine how two theoretically distinct interventions, both of which alter behavior when implemented, may have different implications for rational habit formation. This enriches the existing theory on rational addiction, which predicts that agents should respond consistently as long as the future value of the behavior is altered in the same direction. We present compelling evidence to the contrary, and discuss how this is consistent with behavioral theories on the hidden costs of extrinsic rewards. Relatedly, by identifying the marginal impacts of incentives, monitoring, and dispenser and soap provision, this study sets an important precedent for the design of public health campaigns, which regularly pool multiple interventions together and are unable to disentangle the causal effects of each, often theoretically distinct, dimension of the program. For example, in Luby et. al (2005), the highly-cited study used as the hallmark of a successful handwashing campaign, community volunteers visit households twice weekly, deliver soap, instruct and monitor households' handwashing practices, and also advise households on other hygiene and sanitation behaviors. While the authors find a sustained effect of the intervention on child incidence of diarrhea and respiratory infection, they are unable to identify which

[2] We cannot reject that treatment effects are statistically equivalent across the sub-treatment arms of incentives, monitoring, and dispenser only, so we report pooled estimates here.

aspect of the intervention led to the health improvements.

Third, our data quality is unprecedented within the hygiene and sanitation literature. The objective, high-frequency data of the dispenser sensors allows us the first opportunity to design an experiment which disentangles the various behavioral mechanisms that may lead to poor handwashing takeup. This time-stamped data is also rare in the broader literature of adoption of preventative technologies. It complements the recent collection of energy conservation studies in developed countries that utilitize household-level meter data from energy utility companies to examine how various informational interventions affect household energy consumption (Alcott and Rogers 2014, Ito et al. 2014, Jessoe and Rapson 2015, Alcott and Kessler 2015, among others). Importantly, these studies have as yet been unable to disentangle the mechanisms which lead to reduced energy consumption, whereas the sensor data and design of the present study permit a direct link between increased dispenser use and handwashing with soap during the evening mealtime.

Finally, this study offers the first *treatment on the treated* estimate of the impact of handwashing on child health. In a literature that is plentiful in health impacts of zero, occasional in impacts that are positive yet unable to identify the cause of improved health, and scarce in causal estimates which still say nothing of the ratio of input (handwashing) to output (health), this study offers a significant step forward in establishing the magnitude of impact that handwashing alone can have on health. This helps us build a more precise production function of child health as it relates to preventative behaviors in low-resource settings, which is essential for the more efficient allocation of research and policy dollars.

The remainder of the paper proceeds as follows. Section 2 outlines the conceptual framework motivating our experimental design; Section 3 describes the study sample and design; Section 4 specifies our outcomes of interest and the empirical strategy; Section 5 presents results on handwashing behavior; Section 6 presents child health results; and Section 7 concludes.

2 Conceptual framework

Our framework for habit formation builds upon the seminal work of Becker and Murphy (1988) on rational addiction. They and others in their spirit have focused on characterizing and testing the implications of rational addiction in the context of bad habits. We articulate the same and expand to the context of good habits, of which handwashing with soap before mealtime is our focus. Substantively, the shift from a bad habit to a good habit is equivalent to the shift from an activity in which the user experiences positive gains in the present but incurs costs in the future to an activity in which the user incurs costs in the present but experiences positive gains in the future. This model is formalized in Section 3. Throughout our discussion, we use 'addiction' and 'habit formation' interchangeably, as their underlying mechanisms are identical.

Habitual behaviors share two defining properties: (1) reinforcement, or the development of a craving,

such that the more one engages in the behavior, the more one wishes to engage in it; and (2) tolerance, such that the more one engages in the behavior, the easier it becomes (the lower the cost incurred in engaging).[3]

Reinforcement and tolerance are intrinsic properties of a habit, to be experienced by the user by nature of the activity. *Rational* habit formation (what Becker and Murphy (1988) term rational addiction) is the recognition of these properties: the user is aware of the habit forming nature of the activity, and is thus aware not only of the present cost and future return, but of the craving and tolerance developed through continual engagement, and chooses to engage conditional on this knowledge. The key tradeoff that a rational individual engaging in a habit forming activity faces is between the drop in utility from consumption today and the increase in long-run utility from the accumulation of the stock in the addictive good.

2.1 Empirical evidence on rational habit formation

Existing empirical literature revolves around bad habits (eg. smoking and alcohol consumption). The standard empirical test of rational 'addiction' involves regressing present consumption on past and future consumption and other demand shifters, instrumenting for the lag and lead of consumption using lag and lead of prices or tax rates.

$$c_t = \theta \mathrm{c}_{t-1} + \beta\theta \mathrm{c}_{t+1} + \delta \mathrm{p}_t + \epsilon_t$$

where a positive coefficient θ is evidence of addictiveness and a positive coefficient $\beta\theta$ is evidence of rational addiction. The ratio of the latter to the former yields the discount rate β (Becker et al. 1994).

The vast majority of the literature rests in favor of rational addiction. However, Auld and Grootendorst (2004) describe the implausible variation in discount rates, unstable demand, and low price elasticities implied by such literature (Becker et al. 1991). They go on to demonstrate that entirely non-addictive goods such as milk display the same positive and significant coefficient on future consumption as cigarettes under the standard empirical test, using this supposed rational addictiveness of milk as evidence for the abundance of false positives in the empirical literature. The authors demonstrate how high serial correlation in the commodity of interest, endogeneity in the price instruments, and overidentified IV estimators can all contribute to positive coefficient on future consumption incorrectly interpreted as evidence of rational addiction .

The field as such has advanced little in the last thirty years, as the datasets employed are aggregate time series data which suffer from the problems described above. Of significant added concern is the implausibility of consumer knowledge of future price changes in the contexts explored in the literature.[4] We therefore find

[3]Note that for bad habits, reinforcement follows the same direction, but tolerance moves in the opposite direction: the more one engages in the bad behavior, the lower ones future utility given the amount of future consumption. In contrast, the more one engages in the good behavior, the less ones utility drops from each unit.

[4]Gruber and Koszegi (2001) seek to address the problems of endogeneity and implausibility of future changes by employing state specific time trends and using announced but as of yet unenforced tax rate increases (rather than far future sales data)

the literature well poised for the introduction of a field experiment. A field experiment addresses each shortcoming described above: (1) prices are imposed exogenously so there is no concern for endogeneity between prices and consumption; (2) future prices are explicitly announced so consumer knowledge of future price changes is confirmed; (3) differential time trends are no longer a concern given randomization; (4) endogenous misreporting is not a concern if we have an objective measurement device; (5) serial correlation is no longer a concern as we are comparing outcomes across groups rather than over time.

2.2 Model of rational habit formation

We present a model of rational habit formation for positive behaviors, adapted from O'Donoghue and Rabin's (2001) discrete time exposition of rational addiction for bad habits.

Consider a discrete time model with periods 1,...,T. In each period, an agent can wash her hands before dinnertime such that consumption $w_t = 1$, or refrain from handwashing such that $w_t = 0$. Define k_t as the 'habituation level' of the agent in period t:

$$k_t = \gamma k_{t-1} + w_{t-1}, \gamma \in [0, 1) \tag{1}$$

Habituation is a recursive function which is dependent on the agent's habituation to handwashing in the previous period, k_{t-1}, the level of decay the behavior is subject to, γ, and whether the agent washed in the previous period, w_{t-1}.

Define the agent's instantaneous utility function in period t as

$$u_t(w_t, k_t) = \begin{cases} (\alpha + \beta)k_t - x_t & \text{if } w_t = 1 \\ (\alpha - \sigma)k_t - y_t & \text{if } w_t = 0 \end{cases} \tag{2}$$

where $x_t \geq 0$ is the cost associated with handwashing before dinnertime and $y_t \geq 0$ is the cost associated with refraining from the behavior.

Define the agent's 'desire' to handwash, $d_t(k)$, as the instantaneous marginal utility of washing:

$$\begin{aligned} d_t(k) &= u_t(1, k) - u_t(0, k) \\ &= (\beta + \sigma)k_t - [x_t - y_t] \end{aligned} \tag{3}$$

The model has three key features. First, a good habit generates positive internalities: the more one has

as instruments for future consumption. However, they are still vulnerable to the endogeneity of prices to consumption yielding spurious results in favor of rational addiction. Furthermore, although the announced tax rate change is an improvement upon previous work, there is no way to verify whether consumers are aware of the future tax rate, and the likelihood is low given the year or more between the observed consumption decision and the tax enactment.

washed her hands in the past, the greater current well being will be ($\alpha > 0$). This is independent of whether one washes today or not, since health benefits are always realized in the future. Second, the process may generate learning: the more one has washed her hands in the past, the easier it will be to wash at present ($\beta > 0$). Finally, the behavior must be habit forming: the more one has washed her hands in the past, the greater her desire to wash at present ($d_t'(k) > 0$, or $\sigma \geq -\beta$). σ paramaterizes the craving generated through habit formation.

For good habits, an agent who washes her hands with soap chooses to incur the marginal cost of washing but benefit from the ease of washing (β) and from avoiding the utility hit of the craving (σ), both of which are a function of past behavior. The desire to engage $d_t(k)$ is positive iff

$$x_t - y_t \leq (\beta + \sigma)(\gamma k_{t-1} + w_{t-1}). \tag{4}$$

In a world with no learning and where agents discount the future fully, an agent will only wash her hands today if the marginal cost of washing is less than the craving (eg. anxiety, discomfort from breaking a routine) incurred when not washing. The more she has washed in the past, the greater the impact of the craving on the marginal utility of consumption and the more likely she is to wash in the present. This intertemporal complementarity in consumption is the essence of habit formation.

What levers can be shifted to generate a positive desire to handwash? For agents who have not yet accumulated handwashing stock (most households in our setting), neither γ, β, nor σ initially offer leverage, because $k_{t-1} = 0$ and $w_{t-1} = 0$. To facilitate the accumulation of stock, we must focus first on lowering the cost of handwashing, x_t, or increasing the cost of *not* handwashing, y_t. We consider doing the former with incentives (I) and the latter with third party monitoring (M).

Having generated a positive amount of handwashing stock, these interventions can be complemented with an environment which facilitates maximum retention of handwashing stock. For example, we can maximize the size of the craving generated, σ, by framing the behavior as part of a habit loop: the handwashing routine can be supported on the front end by the trigger of mealtime and the back end by incentives or monitoring feedback. Given our limited sample size, we choose not to experimentally vary the type of trigger administered (it remains the dinner mealtime for all households), but do vary the feedback on the back end as described further in the following section on experimental design.

Lastly, if our objective was to maximize the impact of past consumption on present consumption, we could consider a behavior for which learning by doing (β) is important. However, the purpose of this experiment is to identify the effects of 'habit formation' as interpreted through the channel of the craving only. Therefore, we minimize the amount of learning required to engage in the activity. We do this by defining our behavior as the specific and limited task (a 'baby step') of washing before a particular mealtime, and we maximize the convenience and accessibility of the dispenser and soap by positioning the dispenser near the

household's place of eating and providing soap regularly.

Thus far we have considered the instantaneous utility function which an agent faces for habit-forming behaviors. In a world where agents are forward thinking, the long run utility function is as follows:

$$U_t(k_t, w) = \begin{cases} [(\alpha+\beta)k_t - x_t] + \delta U_{t+1}(\gamma k_t + 1, \alpha, \beta) & \text{if } w(k_t, t) = 1 \\ [(\alpha-\sigma)k_t - y_t] + \delta U_{t+1}(\gamma k_t, \alpha, \sigma) & \text{if } w(k_t, t) = 0 \end{cases} \tag{5}$$

where $\delta \leq 1$ is the agent's discount factor. A rational habit former is one who *recognizes* the intertemporal complementarities in utility from consumption. If she is aware that her stock of handwashing in the past affects her likelihood of engaging today, then she is similarly aware that her likelihood of engaging in the future will be affected by her engagement today. Therefore, if an exogenous shock changes her likelihood of engaging in the future, she should update accordingly her likelihood of engaging today.

Note that for an agent who is *not* rational regarding the habit forming nature of handwashing, the long run utility function she will be operating under is as follows:

$$U_t(w) = \begin{cases} [\alpha w_t - x_t] + \delta U_{t+1}(w_t, \alpha) & \text{if } w_t = 1 \\ [\alpha w_t - y_t] + \delta U_{t+1}(w_t, \alpha) & \text{if } w_t = 0 \end{cases} \tag{6}$$

In other words, while a change in the value of handwashing will have an impact on future utility, it has no impact on the instantaneous utility that an agent believes she will gain from washing today, and should therefore not affect her decision to wash or not today.

In summary, the model yields the following testable implications.

1. Health: $\alpha \geq 0$. Handwashing, as a good habit, generates positive health internalities.
2. Incentives: $\frac{\delta x_t}{\delta d_t} \geq 0$. Reducing the cost of handwashing (by increasing the value of handwashing) raises handwashing rates.
3. Monitoring: $\frac{\delta y_t}{\delta d_t} \geq 0$. Increasing the cost of not handwashing raises handwashing rates.
4. Habit formation: $\frac{\delta d_{t-1}}{\delta d_t} \geq 0$. A rise in past handwashing rates increases current handwashing rates.
5. Rational habit formation: $\frac{\delta d_{t+1}}{\delta d_t} \geq 0$. An anticipated [and actual] rise in future handwashing rates is associated with an increase in current handwashing rates.

2.3 Conceptual framework for incentives and monitoring

Importantly, the incentives and monitoring interventions can shift not only the cost parameters, but alsogenerate variation in the decay parameter γ. An incentive intervention may raise x_t (and thereby the likelihood

of handwashing) more than a monitoring intervention lowers y_t, but the extrinsic motivations behind incentives may yield a more rapid decay than those behind monitoring. This follows from a rich psychology literature (that has yet to be deeply explored in economics) on the negative effects of extrinsic rewards on future re-engagement (i.e. persistence) of a behavior (Condry and Chambers 1978; Gneezy et al. 2011).

Figure 1 plots the likelihood of washing (equated here with the desire to wash) over time for varying parameters of x_t, y_t, and γ under potential incentives and monitoring regimes. Period six marks the end of the interventions. Note that the rates of decay are higher after the interventions have been withdrawn, a parameter choice we make to most closely approximate reality. The blue and red lines plot monitoring and incentive performance, respectively, for identical rates of decay but varying costs of engaging. The green line plots the incentive performance for higher rates of decay. Because changing the cost of handwashing shifts the performance curve, but the decay parameter alters the slope of the curve, a monitoring intervention which is less successful than an incentive intervention in the short term may yield higher handwashing rates in the long term.

In fact, the hidden costs of incentives can affect handwashing rates throughout the experiment relative to monitoring. The two interventions have three important distinctions which predict poorer performance under the former. The monetary nature of an incentive has been associated with poorer behavioral outcomes than more socially motivated interventions like monitoring (Heyman and Ariely 2004). Incentives are also a positive performance-contingent intervention while monitoring is a [potentially] negative performance-contingent intervention to which agents may be more responsive depending on their degree of loss aversion. Finally, in an environment where the principal (the surveyor) may know more about handwashing than the agent (a not unlikely scenario given the low rates of handwashing with soap), a positively priced incentive can serve as a signal of the difficulty of the behavior (Benabou and Tirole 2003). While the size of the incentive may be sufficient to increase handwashing rates *during* the incentive regime, handwashing rates may fall to below pre-regime rates outside of the incentive regime because the agent's estimation of the cost of engaging has increased. This is in contrast to the monitoring intervention which, by focusing on matters of self image and shared information, is less likely to convey a signal that handwashing is more difficult than the agent anticipated at baseline.

These differences not only predict more rapid decay after the withdrawal of incentives, but also a potential dilution of a rational habit formation effect prior to the introduction of incentives. While rational habit formation would predict an increase in rates in anticipation of a higher return (incentives) in the future, the extrinsic nature of incentives may generate a decrease in rates given the signal of higher cost to the activity of handwashing. In contrast, the anticipation of a monitoring intervention, which motivates intrinsic feelings of self-worth, knowledge, and expectations, will not have this dampening effect on rational habit formation.

In summary, the above discussion yields the following additional testable implications.

1. $\delta\gamma_I \geq \delta\gamma_M$: An extrinsically motivated incentives intervention will decay at a more rapid rate than an intrinsically motivated monitoring intervention.

2. Extrinsic interventions and anticipated costs: $\frac{\delta x_{t-1}}{\delta d_t} \geq 0$; $\frac{\delta y_{t-1}}{\delta d_t} = 0$ An anticipated extrinsically motivated incentives intervention will increase the apparent cost of handwashing. An anticipated intrinsically motivated monitoring intervention will have no effect on the apparent cost of handwashing.

3. Extrinsic interventions and future costs: $\frac{\delta x_{t+1}}{\delta d_t} \geq 0$; $\frac{\delta y_{t+1}}{\delta d_t} = 0$ An extrinsically motivated incentives intervention will increase the cost of handwashing after the incentive subsidy is withdrawn. An intrinsically motivated monitoring intervention will have no effect on the cost of handwashing after the monitoring is withdrawn.

3 Experimental Design

3.1 Study sample and context

Our sample population is made up of 2943 peri-urban and rural households containing 3763 children below the age of seven across 105 villages in the Birbhum District of West Bengal, India. Table 1 presents sample means for a host of household, mother, and child characteristics, as well as measures of the mother's hygiene knowledge and practice at baseline. The average mother is just above 30 years and was married at age sixteen with six years of education. 55% of households in our sample are day laborers and 20% work in agriculture. 40% have a latrine, although 68% continue to practice open defecation. Respondents appear to know a substantial amount regarding hand hygiene: 95% are aware that soap cleans hands, and 79% articulate without prompting that soap cleans germs. However, hygiene practice is poor. Despite more than 96% of respondents reporting that they rinse their hands with water before cooking and eating, only 8% report using soap before cooking and 14% before eating. They report that their children wash with soap before eating 30% of the time.[5] This failure to use soap cannot be due to lack of soap availability: 99.8% of households report having soap in the home.

Our partner organization is the Society for Health and Demographic Surveillance (SHDS), a public health organization with a strong presence in the Birbhum District. SHDS had been conducting a variety of public health surveys and initiatives within the sample region over the previous ten years, and surveyors

[5] While these numbers are low, they are likely to be overestimates: in a pilot done across 50 representative households from our sample region four months prior to baseline, child soap use before eating was reported at 8%. It is likely that, over the course of the four months between the pilot and the baseline, households heard more about the upcoming project and catered their baseline answers accordingly (our surveyors, as part of a large public health organization which works regularly in our sample fields, would have shared the upcoming project with excitement).

had already been visiting all households in our sample biweekly (twice monthly) for one year prior to this study's baseline in order to collect child health data. Therefore the biweekly surveyor visits of the present study were nothing new to our study population.

3.2 Dispenser and soap features

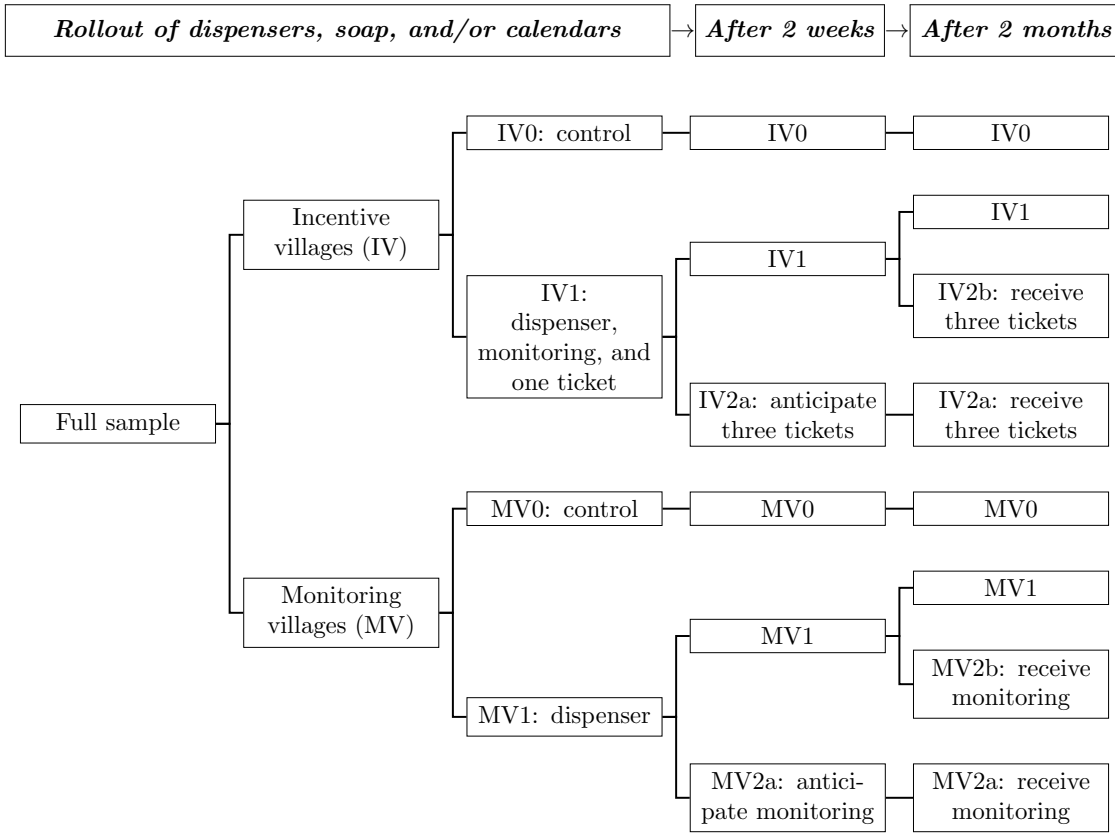
We employed a standard wall-mounted dispenser as depicted in the left picture of Figure 2, which was outfitted with a specially designed sensor. The dispenser is opened with a unique key that was not supplied to the households during the course of the experiment. Soap was loaded in a one liter plastic container inside the dispenser and refilled as needed throughout the course of the experiment (during the surveyors' bimonthly visits). The sensor module is fit between the container and the soap spout, as shown in the right picture of Figure 2. The circuitry is protected by a waterproof casing, an essential feature for the oft wet environment of West Bengal and broadly for outdoor environments. Each push of the outer black button is registered in the sensor, which records the time of each push to the seconds unit. The unit is powered by a small rechargeable 3.7V lithium ion battery which can last up to two months in the field before requiring recharging; this was essential given the lack of electricity in many of our rural households. The sensor is a modular unit, easily removed and refitted into the dispenser; this design permitted surveyors to replace the modules with fully charged versions on their biweekly visit with ease. Each soap dispenser cost approximately $4 USD, and each sensor module cost approximately $26 USD at a quantity of 1200 pieces; this cost drops sharply with higher production given the substantial fixed cost of designing the mold for the waterproof casing. This is the first time-stamped sensor technology to be designed for the purpose of handwashing in outdoor, off-grid environments and successfully implemented at scale.

The dispenser was installed near the dining space or water station as chosen by the household. Figure 3 depicts a typical setting for the dispenser: families usually eat on a mat on the veranda or just inside the front door. We chose a wall-mounted dispenser after repeated prototypes of sensor-embedded tabletop dispensers revealed that (1) the tabletop dispenser was at greater risk of being lost or stolen given its size and mobility, and (2) creating a permanent 'handwashing station' through mounting in a prominent place made it easier for households to remember to wash, potentially enhancing the physical trigger in the habit loop[6]. The dispenser was positioned at a height reachable by young children as shown in Figure 3.

Identifying an appropriate soap likewise required extensive piloting. We experimented with several scents and consistencies which revealed that households preferred: (1) unscented or lightly scented soap that would not interfere with their eating experience; (2) soap of a thinner consistency; and (3) soap that lathered easily. We thus chose a foaming soap with a subtle scent approved by pilot households. We preserved some

[6]Pilot households motivated their valuation for the dispenser with the phrase "chokhe pore," literally meaning that it falls upon the eyes, making soap use easy to remember.

Figure 4: Randomization map



scent as the olfactory system is a powerful sensory source of both memory and pleasure and thus easily embedded into the habit loop (Duhigg 2012).

3.3 Timeline and treatment groups

Figure 4 provides a map of all treatment arms and the time-contingent randomization process.

The randomization was conducted in three stages. First, the 105 sample villages were randomized into Monitoring Villages (MV) and Incentive Villages (IV). Households in MV were randomized into two groups: (MV0) control and (MV1) dispenser. Households in IV were likewise randomized into two groups: (IV0) control and (IV1) dispenser + monitoring + incentive (henceforth referred to as 'incentive households'). Households were first randomized at the village level in order to limit the scope for inter-household tension: surveyors expressed concern that control households would be angered if they had some neighboring house-

holds who received a dispenser and others who received a dispenser and incentives. It would be easier to justify the interventions through the limited resources lottery framework if all dispenser-receiving households within a village received a consistent package of goods (i.e. the dispenser either always came paired with incentives or never did).

At rollout, all households received a basic information campaign regarding the importance of washing hands with soap, especially prior to eating. They also received a calendar with the SHDS logo as a token for participation. They were notified that they would be visited biweekly for one year to collect information on child health and (for those who received dispensers) check and replenish soap supplies.

The remainder of the randomizations were conducted at the household level. Each treatment arm is described in detail below.

(MV0 and IV0) Control: Households were given a simple informational lecture on the importance of washing hands with soap, with stress placed on the responsibility of the mother to do so and encourage her household to do so for the sake of her children's health. They also received a calendar with the partner organization logo on it as a token of appreciation.

(MV1) Dispenser: Households were given a liquid handsoap dispenser. The product was described as a high quality soap dispenser that would make it easier to wash hands. Households were informed that there was a switch inside the dispenser that, if turned on, would track their behavior. SHDS wished to offer a monitoring service to the households in which household handwashing behavior would be tracked on their calendar and reported biweekly. Because resources were limited, the service would be administered in lottery fashion. If they did not get selected for the service, their switch would not be turned on and their behavior would not be monitored.[7]

(MV2a) Announced monitoring: Two weeks after dispenser distribution, these households were informed that they were selected in the lottery: the internal switch would be turned on, and the device would record the time and frequency with which they washed their hands with soap.[8] The surveyor would be carefully observing this data every two weeks to track their behavior and would provide the household with a biweekly report of their daily behavior, marking the household's calendar in the presence of the mother. This arm can therefore be regarded as a combination of a third party-monitoring and self-monitoring intervention. The service would begin two months after dispenser distribution on a date circled clearly by the surveyor on their calendar. This date was reannounced at each proceeding surveyor visit.

[7] These lotteries were publicly announced in order to equalize the expected value of the monitoring or incentive boost across receiving and non-receiving households; it preempted the possibility that a household would update its valuation of handwashing because, for example, the provision of an additional incentive was a signal that they should value the behavior more.

[8] Households could choose whether or not they wanted to receive this program; in practice, all selected households chose to accept it.

(MV2b) Unannounced monitoring: Two months after dispenser distribution, these households were surprised with an identical monitoring service to those in MV2a, effective immediately.[9]

(IV1) Incentives: At the point of dispenser distribution, these households were informed that their dispenser tracked the frequency and time of use and that their behavior would be tracked. They were then given a small coin purse and told that they would receive one ticket for every day in which the device was active prior to their stated dinnertime, which they should accumulate in their purse. These tickets could be exchanged for various household and child prizes as detailed on a prize catalog.[10] These incentive payments would last for four months. They were also told that SHDS anticipated receiving additional funding from the government for the project in the near future, at which point SHDS hoped to increase the reward for handwashing by three-fold. Because the future funds were limited, households would be entered into a random lottery to see who would receive the future increase in reward. They would be notified of the results of this lottery within two weeks.[11]

(IV2a) Announced incentives with boost: Two weeks after dispenser distribution, these households were informed that they had been selected in the lottery for the incentive boost and could anticipate receiving triple the number of tickets for every day in which the device was active prior to their stated dinnertime for thirty days. The boost would begin two months after dispenser distribution on a date circled clearly by the surveyor on the calendar. This date was reannounced at each proceeding surveyor visit.

(IV2b) Unannounced incentives with boost: Two months after dispenser distribution, these households were surprised with an identical incentive boost to those in IV2a, effective immediately.

3.4 Identification of effects

The effect of receiving the dispenser and soap alone is captured in the comparison of households in MV1 to MV0.

A higher take-up of handwashing behavior in MV2a relative to MV1 and IV2a relative to IV1 (*before* the price change) demonstrates the presence of rationally habit forming behavior: households who increase

[9] As with the households in MV2a, households could refuse to be monitored. In practice, all households accepted the service.

[10] The ideal incentive requires three conditions: (1) the incentive must be divisible; (2) the daily amount offered must be sufficiently high to induce behavioral change on a daily basis, which is key to habit formation; and (3) the marginal value of the units accumulated as the process of habit formation continues must also remain sufficiently high to continue inducing behavioral change. Tickets exchanged for goods provides satisfies all three conditions while also offering flexibility in the types of goods that a household may find appealing. Prizes were selected to focus on child health and schooling and adult household goods.

[11] It was important that we provide all households with an incentive from the beginning (prior to the increase in incentives) in order to establish an understanding of the nature of the incentives and trust between the surveyors and the households that the future increase would indeed by fulfilled. In additional to such logistical concerns, we also chose to design the experiment to examine the effects of an intensive margin change in incentives in order to most closely mimic the existing literature on rational addiction, all of which examines future intensive margin price changes (all significantly smaller than ours in percentage terms) on current habitual behavior.

take-up today due to an increase in the future value (or decrease in cost) of handwashing must recognize that higher take-up today will generate a greater accumulation of the positive internalities and craving stock over time, making it easier to reap the benefits of the future rewards to the behavior.

A zero difference in take-up between households in MV2a versus MV1 and IV2a versus IV1 prior to the price change could be due to three reasons: (1) households are not rational habit formers in handwashing; (2) the future price change was not sufficiently compelling to induce behavioral change, even for forward-looking individuals; or (3) handwashing is not a habit-forming activity. The second possibility is eliminated if households do indeed respond to the price change when it is enacted. This pure price change effect can be identified by comparing households in MV1 to those in MV2b and households in IV1 to those in IV2b after the price change, as the only difference between these sets of households is the price change itself, with no behavioral response to anticipation. This comparison gives us the pure effect of the incentive boost or the monitoring service on handwashing behavior.

The third possibility is eliminated by comparing persistence in behavior across all arms after the withdrawal of all interventions. For households in arms IV1, IV2a, and IV2b, all incentives [and monitoring] services were discontinued approximately two months after the price change. For households in arms MV2a and MV2b, all monitoring services were discontinued approximately four months after their introduction.[12] In practice, households were informed that the switch in their machine had been "turned off," that surveyors would no longer be observing their behavior but would continue to visit monthly to collect child health data, and that they would no longer provide reports on household handwashing performance (nor tickets for incentive households).[13] A comparison of each treatment arm to MV1 households, who were never exposed to any interventions beyond the provision of the dispenser and soap, quantifies the extent to which a handwashing habit was formed due to the temporary incentives or monitoring interventions. We track household handwashing behavior for one year after rollout.

By maintaing the same incentive stream across both groups, a comparison of MV2a to MV2b and IV2a to IV2b over the course of the experiment after the price change allows us to identify the effect of forward looking, rationally addictive behavior on habit formation (conditional on finding evidence of rational addiction prior to the price change). In other words, a long term comparison of take-up between the 2a and 2b groups demonstrates whether forward-looking behavior in fact facilitated the formation of the handwashing habit.

Finally, comparison of MV2a and MV2b to IV2 (or IV1, IV2a, and IV2b pooled) offers the first estimate in the literature of the marginal value of (positive) monetary rewards on top of (potentially negative) feedback

[12] This difference in date of discontinuation was implemented to equalize the exposure of households to each treatment, since incentive households had already been receiving incentives for nearly two months prior to the price change.

[13] As is true for MV1 (dispenser only) households as well, this practice of informing households that the switch in the machine was "turned off" constitutes deception. The practice was cleared by IRB boards at both MIT and IFMR (our Indian research organization counterpart) prior to implementation and was permitted given the scientific value and policy relevance of estimating the effect of third party monitoring and persistence after the withdrawal of interventions

on daily behavior.[14] [15]

4 Outcomes of Interest

The primary outcomes of interest encompass behavioral change in households and child health. We capture behavioral change through (1) recorded total daily handwashing rates and (2) recorded dinner time-specific daily handwashing rates. Sensor measures of handwashing rates could only be collected for those households with dispensers, so we do not have data from the pure control households on these metrics. We collect child health data in the form of (5) self-reported biweekly incidence of child diarrhea and respiratory illness and (6) anthropometric measures of height, weight, and midarm circumference. Each is defined in detail below.

A. Household handwashing behavior

Handsoap dispenser data was collected every two weeks during surveyor visits. Although it was not possible to identify the identity of the user at any given press, we proxy for separate users by collapsing presses that happen two or fewer seconds apart into a single press. In other words, if the device is used in seconds 34, 35, 37, 45, and 46, the first three presses are considered a single use by one household member and the latter two presses as a single use by another member. Though not exact, observations from pilots made clear that users press several times in quick succession and rarely return for more soap during a single handwashing event, since the water source (usually a bucket right outside the front porch) is not within reach of the dispenser (unlike the familiar setting of sink, soap, and running water common to more developed contexts).

Mealtime-specific handwashing rates are calculated as the total number of 'individual' uses in the interval of 90 minutes before and after the household's reported start of the evening meal time. If a family reported eating dinner every day at 8:00 PM, for example, this outcome would be the sum of all individual presses observed between 7:00 PM and 8:30 PM.

Binary use at mealtime is derived from the above and is a binary variable which equals one if at least one 'individual' use was observed in the dinnertime interval. This is our primary outcome measure

[14] The experimental design precludes perfectly capturing the effect of incentives on top of monitoring, although it is quite close. Monitoring was introduced (MV2a and MV2b) 60 days after rollout, while incentives (IV2) were introduced immediately after rollout. We were deliberate in this choice: monitoring was delayed in order to increase our sample size on the rational addiction test, with the tradeoff of a loss in the perfect comparison between monitoring only and monitoring+incentive households. Given the habitual nature of handwashing (or more broadly, dispenser use), the delay in introducing monitoring may have reduced the malleability of the behavior and therefore the potential effect of the treatment relative to that of incentives. This is in fact precisely why we did not delay the introduction of incentives to parallel the introduction of monitoring: this would mean a 75 day delay in the introduction of the future price change, reducing the likelihood of finding a rational addiction effect.

[15] Marshall, 1980: "There is, however, an implicit condition...which should be made clear. It is that we do not suppose time to be allowed for any alteration in the character or tastes of the man himself. It is, therefore, no exception to the law that the more good music a man hears, the stronger is his taste for it likely to become; that avarice and ambition are often insatiable; or that the virtue of cleanliness and the vice of drunkenness alike grow on what they feed upon. For in such cases our observations range over some period of time; and the man is not the same at the beginning as at the end of it."

of handwashing at dinnertime.

Daily handwashing rates are calculated as the sum of all 'individual' uses over the course of each twenty-four hour period.

Alternative hygiene measures such as direct observation of respondent hand and nail cleanliness, respondents' ratings of own handwashing habit formation, the presence of non-project liquid soap in the household, and the quantity of soap used (as proxied by total daily uses of the dispenser) were collected at the eight-month mark. We also collected measures of household sanitation, such as whether the household practices open defecation and whether they treat their water, to explore complementarities in behavior change and alternative mechanisms through which child health may be affected.

B. Child health

Incidence of child diarrhea and respiratory illness was collected every two weeks by surveyors, consisting of self reports in which mothers were asked how many days each child had experienced diarrhea, loose stool motion, or the symptoms of respiratory illness in the past two weeks.

Anthropometric outcomes were collected at baseline and again at the eight month mark. These include child weight, height, and mid-arm circumference, as measured by trained surveyors. We supplement self-reported incidence data with anthropometric outcomes to reduce the likelihood that any observed effects are driven by desirability bias on the part of mothers. Repeated diarrheal disease can affect child weight and height by reducing a child's ability to absorb sufficient nutrients from her food and thereby stunting her growth (McKay et al. 2010). We convert these measures into standardized height-for-age, weight-for-age, and midarm-circumference-for-age Z-scores (HAZ, WAZ, and MAZ, respectively) using the methodology provided in the WHO anthropometric guidelines; these Z-scores are calculated (as per WHO methodology) only for children five years and below (WHO, 2006).

4.1 Temporality of outcomes

Because various interventions were phased in and out at various times, below we define the time period for each effect of interest.

Baseline behavior is defined through the baseline survey, which was conducted four months prior to rollout.

Rational habit formation period is defined as the time between dispenser distribution and the service/price change. We also zoom in on the three week period prior to the price change. This is because (1) we showed a video three weeks prior to the price change date to all dispenser-receiving households

in order to increase and standardize comprehension regarding which treatment group each household was in; and (2) any rational habit formation effect should increase as the date of the anticipated change approaches, so we focus on the weeks nearest the price change.

Intervention period is defined as the two months following the price change for incentive households and four months following the monitoring service for monitoring households. For IV2 households only, this period is defined as the two months after rollout. These are the times during which pure intervention effects can be measured.

Habit formation period is defined as the period of 2-6 months after the price change for incentive households and 4-6 months after the monitoring service change for monitoring households.

4.2 Empirical strategy

Our preferred specification for primary outcomes is as follows:

$$Y_{hvt} = \alpha_{hvt} + \beta Treatment_{hvt} + \gamma_t + \theta_v + \epsilon_{hvt} \tag{7}$$

in which Y_{hvt} represents the household behavior outcomes specified in Section 6, *Treatment* is the assigned treatment for each subset of comparisons described above, γ_t is day fixed effects, and θ_v is village fixed effects. The latter two are included in all but those regressions comparing treatments across Monitoring and Incentive Villages (we omit village fixed effects in these regressions since randomization to MV or IV was at the village level). Standard errors are clustered at the household level except in cross-IV-MV comparisons, in which they are clustered at the village level. For analyses utilizing the midline survey, which is cross-sectional data collected eight months after roll-out, we omit day fixed effects.

5 Behavioral results

Table 2 presents a comparison of means between treatment and control households for an extensive set of baseline characteristics at the household, mother, and child levels. Treatment households are the pooled sample of all households who received the dispenser and soap; control households are pure control, or households who received no dispenser or soap. Appendix Table 1 presents the same set of comparisons for each treatment arm individually. Households are balanced across the majority of observables. Treated respondents are 0.4 minutes farther from their drinking water source, 3 percentage points less likely to be Hindu, marry 0.2 years later, rate themselves higher on whether people listen to them but lower on whether they make their children's health decisions, have taken their child to the doctor for an illness in the last two weeks 0.14 times more, and are 3 percentage points and 1 percentage point more likely to have a child

experience a cold or diarrhea in the last two weeks, respectively. While the imbalance on the latter three child health metrics may be concerning, the difference points in the opposite direction of the effects of interest, and we control for baseline health incidence in all forthcoming health regressions. The disaggregated comparisons of Appendix Table 1 show no obvious patterns in differences across treatment arms and control either; nor are any of these differences suggestive of imbalance on unobservables in a direction that will lead us to overstimate our effects of interest.

We next present our main results on the impact of each treatment on handwashing behavior. A note to the reader: the description of time in all figures and tables henceforth will be relative to the date of the incentive price change or introduction of the monitoring service, denoted as Day 0. This helps reframe the experiment to align with the standard field experiment that typically begins when the intervention commences. In this setting, we begin our experiment 70 days *before* the key interventions of interest are implemented, permitting the exploration of whether agents are rational about the [habitual] behaviors they choose to engage in.

5.1 Main treatment effects

5.1.1 Incentives

Table 3 presents results on the impact of the *extensive* incentives margin on handwashing behavior by comparing households in IV2, who received one ticket per day that they washed at dinnertime (beginning on the day of rollout), with households in MV2, who received no interventions outside the dispenser. Columns 1-3 demonstrate that incentives worked as intended: after two months of incentives, incentivized households use the dispenser 1.7 more times over the course of the day than control dispenser households (Column 1), but this increase is not born out during the daytime (Column 2); rather, the bulk of the change in handwashing occurs around dinnertime (Column 3). A similar pattern holds after four months of incentives (Columns 5-7).

Figure 5a plots the raw time trend of handwashing during the daytime and the evening across incentivized and control dispenser households over the four months that households were offered the one daily ticket incentive. While the response to incentives increases evening handwashing by approximately one press more per day relative to the control counterparts, there is no parallel trend in daytime handwashing. A closer look suggests that evening handwashing may first complement and then substitute for daytime handwashing (with the switch at day zero), but these differences are not statistically significant. By and large, households appear to regard each handwashing event as an independent act. This underscores the importance of defining [potentially] habitual behaviors with precision in behavioral change campaigns: to "wash hands before dinnertime" is a more tangible, manageable, and trigger-centric instruction than the more widely promoted direction to "wash hands before eating, before cooking, and after defecation."

Column 4 and 8 use the preferred binary outcome variable of whether or not the dispenser was active during the household's stated dinner time. Results show that incentivized households are 24 percentage points more likely than control households to wash at least once during their reported dinner time, both after two months (Column 4) and four months (Column 8) of incentives. By the fourth month of incentives, just before the withdrawal of the intervention, incentivized households are washing their hands during their reported dinnertime 63% of the time.

Figure 5b plots the time trend of dinnertime handwashing rates across incentive and control households. The vertical red lines represent the average dates of surveyor visits, during which incentive households received markups of their calendars and tickets based on their performance from the last batch of data collected. The time trend tells an important story. Households were first visited on day -70: dispensers were delivered and incentive households were told about their daily ticket rewards. They were next visited on day -54, during which surveyors collected the first batch of handwashing data form the dispensers. On the third visit on day -38, surveyors returned with the results of the first batch of data and the tickets the household had earned from this batch. Only upon *receiving* these tickets did households react to the incentive treatment. The reaction is followed by a steep descent, which is again buoyed by the next round of surveyor visits and tickets. Each of the third, fourth, and fifth visits prompt a sharp rise in handwashing, followed by an increasingly shallower decay. By the sixth round, despite continuing surveyor visits, household performance stabilizes.

This pattern is consistent with two stories. First, households may be building trust in the intervention. This is likely at the third visit but unlikely by the fifth. A complementary explanation is that surveyor visits serve as reminders or motivation to engage in handwashing. Motivation is particularly useful (as measured by the response to the visits) when the stock of handwashing that a household has accumulated is low in the early rounds. However, it becomes progressively less effective as the stock builds and the behavior becomes habitual. This pattern is replicated in Alcott and Rogers (2014) in the tracking of household energy usage against the date of letters sent regarding energy consumption and is consistent with a key prediction of Taubinsky's (2014) model of inattentive choice and the substitutability of reminders and habituation.

We next move to the study of rational habit formation. In order to measure rational habit formation, we must first empirically establish two features of handwashing. The first is that handwashing can be moved by our chosen interventions of monitoring and incentives, both on the extensive margin and the intensive margin of an incentive boost. If agents do not respond to these interventions, then the interventions have failed to change the value of the behavior and agents have no reason to respond to the *anticipation* of these interventions. The second feature is that handwashing must be a behavior that can become habitual. If we find no intertemporal complementarities in the behavior (i.e. persistence after the withdrawal of interventions), agents gain no utility from accumulating handwashing "stock" prior to the introduction of

the interventions.

Only after we have established these two features of handwashing can we examine whether agents are *rational* about their habit formation in our setting. Therefore, we present our results from the intervention period first, then the persistence period, then return to the pre-intervention period to examine evidence of rational habit formation. Though temporally out of order, this permits a clearer construction of the story we observe.

5.1.2 Intensive margin incentives

We first examine the impact of an intensive margin shift in incentives on handwashing. Column 1 and 2 of Table 4 presents the results for the comparison between households who received a three ticket boost in incentives to those who remained with the one ticket incentive at Day 0. We report results both for the full 60 days during which households were technically exposed to the boost, as well as a lagged time frame of Days 30 to 59. The lagged time frame is relevant because Day 30 is the first day in which households who were eligible for tripled tickets on Day 0 physically received them. Households responded positively and meaningfully to the tripling of daily tickets, washing 5.3-6.3 percentage points (9-10%) more than their standard incentive counterparts.

The treatment sample can be further narrowed to households who were surprised by the triple ticket boost; these households provide a measure of the 'pure' effect of tripling tickets, as they are not confounded by the implications of anticipation. Here we find a 4.4 percentage point increase in handwashing rates, which rises to a statistically significant 6.8 percentage points (11%) between days 30-59.

Figure 6a presents the three-day moving average of dinnertime handwashing rates for the tripled incentive arm relative to the standard incentive arm during the period of the incentive boost.

5.1.3 Monitoring

Columns 5 and 6 of Table 4 estimate the impact of the monitoring service on household handwashing behavior as compared to dispenser control households.[16] Column 5 presents results for the full tenure of the monitoring service, while Column 6 presents the lagged results of Days 30 to 116. The monitoring service increases handwashing rates by 7.1-8.4 percentage points (21-23%) over the duration of service provision. Results are robust to restricting our sample to the households who were surprised by the monitoring service on Day 0 (Columns 7-8).

[16] Recall that the monitoring service lasted from Day 0 to Day 116, which is two months longer than the length of the triple incentive boost in incentive villages. This was implemented to compensate for the two months of incentives that all incentive households had already received prior to the boost, and thereby permit a closer comparison between the long run effectiveness of incentives relative to monitoring.

Figure 6b presents the three-day moving average of dinnertime handwashing rates for monitored households relative to those who received the dispenser only. This graph highlights how household behavior to the monitoring arm reacts most strongly on the day of the first calendar receipt (Day 30), suggesting the important role of a public feedback mechanism in the effectiveness of the monitoring service.

5.2 Persistence

Section 5.1 establishes that the experiment succeeded in exogenously increasing the value and consequently the 'consumption stock' of handwashing in each treatment arm, addressing our second and third testable implications: $\frac{\delta x_t}{\delta d_t} \geq 0$ and $\frac{\delta y_t}{\delta d_t} \geq 0$. We now explore whether this exogenous shift in stock had an impact on subsequent handwashing behavior after the interventions ceased. Such persistence in behavior is evidence of habit formation: because the interventions that increased consumption stock are no longer active in this later time frame, any difference in performance between a treatment household and its relevant control must be due to intertemporal complementarities in handwashing behavior.

Table 5 presents the results on persistence. Results are separated into the first month after intervention withdrawal (Columns 1, 3, and 5) and all following months (Columns 2, 4, and 6). Columns 1 and 2 show that households who received the standard incentive continue to wash their hands during dinnertime 22.5 percentage points more than their dispenser control counterparts during the first month after incentive withdrawal; this drops to 12 percentage points more than control over the following two months. The intensive margin of incentives, on the other hand, has no lasting effect: formerly triple-ticketed households continue to wash their hands slightly more (3 percentage points) than their single-ticketed counterparts in the month after withdrawal, but this is statistically indistinguishable from zero and disappears entirely by the second month. Finally, Columns 5 and 6 demonstrate that, like the incentives on the extensive margin, the stock built from the monitoring intervention also persists: households are 9.6 percentage points more likely to wash than their dispenser control counterparts in the first month after the monitoring service is halted; we are continuing to collect data and will soon report persistence for later months. These results confirm our fourth testable implication: $\frac{\delta d_{t-1}}{\delta d_t} \geq 0$.

Figures 7a and 7b present the three-day moving average results for [formerly] incentivized and monitored households, respectively, relative to the dispenser control.

While both incentives and monitoring interventions generate persistent handwashing behavior, the degree of persistence differs between the two. Table 6 presents estimates of the decay rate in handwashing over the days following intervention withdrawal. Columns 1-2 examine the slope of handwashing for households who were formerly incentivized (pooling standard incentive, anticipated triple, and surprised triple); Columns 3-4 examine the slope for households who were formerly provided the monitoring service. Though the estimates are noisy, the pattern is clear: relative to the dispenser control, handwashing behavior in incentivized

households decays more rapidly than that in monitored households, despite the fact that incentives were significantly more effective at inducing initial behavior change than monitoring. This is further elucidated in Figure 8, which presents the seven-day moving average across incentive, monitoring, and dispenser control households for the full length of the experiment. This addresses our fifth testable implication: the decay parameter γ must also be a function of the type of intervention utilized to promote a habitual behavior. In our setting, intrinsically motivated interventions such as monitoring show greater persistence, or smaller decay, than extrinsically motivated interventions such as incentives.

5.3 Rational habit formation

Having established that handwashing is a habitual activity and that the interventions change the value of handwashing, we now turn to the question of whether agents are rational about the habit-forming nature of this behavior. Results are presented in Table 7. We first examine the pre-change period. Recall that during this period, no incentive households had received the tripled tickets and no [potential] monitoring households had received a monitoring service. Rather, a portion of them had been notified on Day -54 (two weeks after rollout) that they should expect such a change to take place at a future date as circled on their calendar (Day 0). We compare the behavior of these anticipating households to households who were not told to expect any change in the future. Results are presented both for the full period of anticipation (Day -54 to Day -1) as well for the final three weeks before the date of change (Day -21 to Day -1).

Columns 1-2 present the results for households anticipating a future tripling of tickets relative to those who are not. Although the coefficient of interest is positive, it is small and imprecise, offering no evidence that anticipation of a future price change affects current behavior. In fact, the coefficient becomes smaller as the date nears the date of change, further rejecting the presence of rational habit formation.

Columns 3-4 present the results for households anticipating a monitoring service relative to those who are not. In contrast to the incentives setting, households anticipating monitoring are 5.2 percentage points more likely to wash their hands during dinnertime than their unanticipating counterparts (22.5%); this rises to a substantial 8 percentage point difference (39%) in the final three weeks before the monitoring commences, significant at the one percent level.

The remainder of Table 7 explores whether these patterns continue throughout the rest of the experiment. Columns 5-8 examines the behavior of (now formerly) anticipating household relative to their surprised counterparts over the course of the triple ticket or monitoring regimes, and Columns 9-12 examine their behavior after the withdrawal of the interventions. Consistent with the theory, anticipating triple ticket households, who accumulated no more stock than their nonanticipating counterparts in the pre-change period, show no differences in behavior during the triple ticket regime nor after incentive withdrawal. This is clearly shown in Figure 9a, which plots the five-day moving average of handwashing behavior between antici-

pating and nonanticipating triple ticket households. Household behavior follow essentially identical patterns over the course of the experiment. On the other hand, Figure 9b, which plots the same for anticipating and nonanticipating monitoring households, suggests that the effects of the handwashing stock accumulated by anticipating monitoring households in the pre-change period persists. Although the estimates are noisy, formerly anticipating households wash 3.1 percentage points more than their surprised counterparts during the monitoring regime (Columns 7-8), decreasing to 2 percentage points and eventually zero two months after the monitoring service has stopped (Columns 11-12).

Pair the results of Tables 6 and 7 and we see that, not only is the weight of past behavior on present behavior greater amongst those who have been monitored (persistence), but the weight of future behavior on present behavior is also greater for those who expect to be monitored (rational habit formation). Our results therefore confirm our sixth testable implication, $\frac{\delta d_{t+1}}{\delta d_t} \geq 0$, for the monitoring intervention, but reject any evidence of rational habit formation for the incentives intervention.

5.4 Discussion

Are there competing explanations for why households anticipating the monitoring service responded so substantially in anticipation, while those anticipating a tripling of tickets did not? One possibility is confusion: households may have believed that they were being monitored starting on Day -54 rather than Day 0. This is unlikely for two reasons: first, households were reminded of their treatment assignment on every surveyor return date (Day -54, Day -38, and Day -21): as an anticipating household, one was reminded of the impending date of change; as a non-anticipating household, one was reminded that the surveyor would continue to return every two weeks to collect child health data. Households were shown a video on Day -21 clarifying their treatment assignment; the videos were met with much interest by both mothers and their children and involved interactive comprehension questions, which should have further reduced the possibility of confusion.

Instead, our set of behavioral results appear to be consistent with the two theories considered in our conceptual framework. Agents in the triple ticket incentives treatment may have indeed been rationally habit forming, but any increase in washing due to the anticipation of a future price boost may have been countered by a corresponding decrease in washing due to an updated belief about the difficulty (or hassle) of the activity. However, given that these agents had already been engaging with the dispenser regularly under the standard (single ticket) regime, it is not obvious that they would have perceived any remaining information asymmetry between themselves and the surveyors regarding the nature of handwashing (and it is thereby not obvious that they would have used the announcement of the tripling as a signal of the difficulty of handwashing). Furthermore, since the possibility of triple tickets was conveyed to all households (both anticipating and non-anticipating in the form of a lottery at the beginning of the experiment), a difference in anticipatory behavior due to a difference in cost perception is unlikely. As such, loss averse preferences

are a plausible supplement to this story: agents anticipating a future loss in self-respect may have been more pressed to prepare through practice than agents anticipating a future gain in prizes.

6 Health results

Does handwashing generate positive health internalities, $\alpha > 0$? We examine three sets of data. The first utilizes day-level reports by mothers of child diarrhea and ARI incidence as collected by surveyors every two weeks during the first five months of the experiment. We examine health data from months four and five only, as this encompasses the peak of handwashing performance across treatment households.[17]

Our second set of outcomes utilizes two-week incidence reports from the midline survey conducted between months seven and eight. This midline survey revised the manner in which we collected data on child diarrhea and ARI. The restructuring was motivated by concerns from the field that surveyors were missing incidence cases. For example, for diarrhea, reporting mothers (1) felt diarrhea was a serious illness that their children could not suffer from unless the child was visibly sick, and (2) often did not know whether their children had experienced regular loose stool motions since their children played outside most of the day and defecated in open fields away from the house. We therefore revised the questions to cast a wider net on illnesses and we required surveyors to have the child present during the time of surveying.[18]

Our final set of health outcomes also comes from the midline survey, in which we recollected the baseline anthropometric measures of child height, weight, and mid-arm circumference.

Table 8a presents the intent-to-treat estimates from the child-day level incidence reports. All regressions include day and village-level fixed effects as well as a full set of child health baseline controls, although results are robust to excluding baseline controls (not shown). Columns 1 and 5 report results for the pooled sample of all treated households relative to households in the pure control group respectively for diarrhea and ARI incidence. Columns 2-4 and 6-8 disaggregate this sample into each treatment group: incentives, monitoring, and dispenser control households.

[17] Although this time restriction was not specified in the pre-analysis plan, we did not explore any other time-frame for the health outcomes during our analysis to avoid the concern of multiple hypothesis testing.

[18] The wider net was cast as follows: mothers (and children directly) were asked whether the child had experienced *any* loose stool motion in the last two weeks. If so, the days they experienced it were recorded. This is in contrast to the previous five months of incidence data collection, during which mothers (and not chilren) were asked whether their child had experienced loose stool motion at least three times in a day, the clinical definition of diarrhea. As is evident in Table 8a, this yielded too few cases for statistically significant movement to be detectable. We acknowledge that a single loose stool motion is not necessarily reflective of diarrhea; however, a single reported motion is likely to be a signal for more actual motions in a day (given the recall problem for young children and the lack of supervision problem for mothers). We report the results, however, as 'loose stool' and not as 'diarrhea' and leave the reader to interpret. For the ARI question, mothers (and children directly) were asked whether the child had experienced *any* of the symptoms of ARI in the last two weeks, and the surveyor listed the following: 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. If the respondent answered yes to any of them, the surveyor then asked how many days they had experienced these symptoms. This is in contrast to the previous five months, during which surveyors asked whether the child had suffered from any two of the three symptoms of a runny nose, cough, or fever.

While estimates for the impact of treatment on diarrhea are consistently negative, they are noisy and close to zero. This is not surprising, as the reported likelihood of a child in the pure control group suffering from diarrhea on a given day is only 0.4 percent. Results on ARI are clearer: children in treated households are 2.2 percentage points (15.3%) less likely to be suffering from ARI on a given day than their untreated counterparts; this effect size, significant at the one percent level, is relatively evenly distributed across the treatment groups, with monitoring households seeing the largest drop in ARI incidence of 23.4%.

Table 8b presents the intent-to-treat estimates from the restructured midline survey with and without baseline controls. Mean two-week incidence of loose stool in the pure control group is 10.4%. A child in a treated household is 3.2 percentage points (30.4%) less likely to experience loose stool motion in the previous two weeks. Similarly, the average treated child experiences .08 fewer days of loose stool (39.4%) per two weeks, significant at the one percent level. Therefore, when we broaden the net to any loose stool, the impact of handwashing is clear.

ARI results remain consistent in percent magnitude with those in Table 8a. A child in a treated household is 3.7 percentage points (13.6%) less likely to show any symptoms of ARI in the last two weeks, 2.1 percentage points (23.3%) less likely to have experienced ARI as defined by a cough and runny nose, and experiences 0.2 fewer days (14.9%) of ARI per two weeks. Appendix Table 2 disaggregates these results into each treatment arm; treatment effect sizes remain broadly consistent across arms, but we do not focus on these given concerns over multiple hypothesis testing.

Table 9 presents the intent-to-treat estimates on child anthropometric outcomes. Both HAZ and WAZ scores increase by approximately 10% in treated households (0.20 and 0.19 points, respectively). Effect sizes for MAZ are also positive, but the estimates are imprecise. To get a sense of the magnitude of these results, consider that children ages five years and below in treated households are approximately 0.38 kg heavier than those in pure control households. At a conversion rate of 7780 calories per kilogram (Wishnofsky 1958) and given that the dispensers have been in use for eight months at the point of data collection, treated children are able to absorb approximately 12 more calories per day than children without a dispenser.[19]

Since the average rate of handwashing at dinnertime among treated households is 47%, we now consider estimates of the treatment on the treated (TOT), where "treated" is a child (household) who uses the dispenser at dinnertime (restricted to performance in the month of and month prior to outcome data collection). As dispenser use is endogenous, we instrument for it using treatment assignment, utilizing the three treatment groups of incentives, monitoring, and dispenser households. Note that this definition of 'treated' implies that all pure control households are compliers: because no pure control household received a dispenser, none of them could have been treated. The local average treatment effect we obtain from our

[19]This exercise was adopted from Bennet et al. (2015), and despite significant differences in the type and time length of handwashing interventions being tested between this paper and Bennet et al., the change in per day caloric intake due to the intervention is remarkably similar (12 v. 14 calories per day).

instrumental variables analysis is therefore equivalent to the treatment on the treated.[20]

In particular, we run the following two-stage regression for child c in household h, village v, and time t:

$$Wash_{hvt} = \alpha_{hvt} + \beta_1 Incentives_{hvt} + \beta_2 Monitoring_{hvt} + \beta_3 Dispenser_{hvt} + \epsilon_{hvt}$$
$$Health_{chvt} = \alpha_{chvt} + \beta_2 W\hat{ash}_{hvt} + \delta_{chv} + \theta_v + \gamma_t + \epsilon_{chvt}$$

in which δ_{chv} is a vector of child health baseline controls and θ_v and γ_t represent village and time fixed effects respectively (the latter only for the day-level incidence analysis). This instrumental variables estimate thus examines the health effects of dispenser *use* rather than ownership only.

Table 10 presents the TOT estimates. A child in a household that uses the dispenser before one hundred percent of evening meals experiences a 65% decrease in the likelihood of having loose stool, an 83% decrease in the number of days she experiences loose stool, a 28% decrease in the likelihood of experiencing any ARI symptoms, a 45% decrease in having ARI as defined by a cough and runny nose, and 31.5% fewer days of ARI. She also sees a 19% rise in her HAZ score and a 21% rise in her WAZ score.

The preceding analysis yields two key takeaways. First, the results provide the first causal evidence in the literature that handwashing alone generates significant positive health internalities in the developing world. Second, the analysis highlights that the mere provision of the dispenser and liquid soap has a significant impact on child health. In fact, the marginal impacts of each treatment arm are for the most part statistically indistinguishable from the impact of the dispenser arm alone. This large treatment effect to dispenser provision cannot be due to a newfound availability of soap in treated households, as baseline estimates point to 99% of households having [and using] soap in the home. Rather, this must be due to some combination of the household's valuation of the dispenser and liquid soap and thereby the act of handwashing ("if we receive something so nice, handwashing must be important and we should use it") along with the convenience of the dispenser location, being stationed right next to the place of eating. Novelty is a less likely explanation, since our results are estimated seven to eight months after the distribution of the machines.

The mechanisms underlying these results are a fruitful avenue of future research: at the fixed cost of $4.00 USD per dispenser and variable cost of $0.10 USD per liter of soap every 1.5 to 2 months (the average household consumption rate), child respiratory infection incidence can be reduced by at least one fifth and loose stool incidence can be reduced by at least one third of baseline rates in our study context.

[20] If we instead consider "treated" as a child who washes her hands at dinnertime (with or without the dispenser), pure control households will have some positive fraction of treated children ("noncompliers"). Consider if children in pure control households wash their hands before dinnertime one day per week. This will inflate all presented TOT estimates by 1.4. If it is two times a week, the inflation will be by 2.8, and so forth. We can therefore regard the TOT estimates presented in Table 10 as a lower bound on the effect of handwashing on child health, barring significant selection concerns.

7 Conclusion

This study seeks to analyze the process of habit formation in the high-impact preventative health behavior of handwashing with soap and to better understand how individuals interact with this habit-forming behavior. Our results suggest that monetary incentives and third party monitoring are effective means of increasing handwashing rates in the short run, and the former more than the latter; that these temporary interventions have persistent effects, thereby establishing that handwashing is indeed a habitual behavior; and that those who initially received incentives exhibit a more rapid decay in handwashing persistence than those who had received monitoring after the removal of the interventions. We have also presented strong evidence that agents are rationally habit forming when anticipating a monitoring intervention but not when anticipating a boost in incentives, despite reacting to both when the interventions arrive; this suggests that there is a feature of each intervention - *beyond* the anticipated change in the value of handwashing that it will generate - which motivates agents to consider the intertemporal complementarities in consumption in one setting and not the other. This exercise offers the first well identified estimates of the presence of rational habit formation, and additionally for *good* habits, in the literature.

The experiment also sheds light on the production function for child health as it relates to the input of hand hygiene. We establish the strong link between handwashing with soap and child incidence of respiratory infection and diarrhea, and can uniquely offer treatment on the treated estimates which suggest that a child who washed her hands at dinnertime every day saw an 83% decrease in the number of days she experienced loose stool motion and a 31.5% decrease in the number of days she experienced acute respiratory illness.

8 Appendix

8.1 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 et al. 2015, 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 the Hawthorne effect. 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 4 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.[21] 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

[21] 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

in hand hygiene may be complemented be changes in other sanitation practices, if for example the act of having handwashing top of mind makes remembering to maintain other preventative 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 5 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 handwashing had no complementary effect on other dimensions of household sanitation.

8.2 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.[22] At each level, the respondent was asked whether she would prefer to take the prize or take a month's worth of soap.[23] 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 and received the drawn prize.

Results are presented in Appendix Table 6. 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 18% lower than that of control households. Valuations among monitoring and dispenser control 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. This reinforces the potential failure of temporary extrinsic rewards to generate sustained changes in behavior.

However, formerly incentivized households are also significantly more likely than their pure control

[22] 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.

[23] 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 a 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.

counterparts to have non-project liquid soap in the household (Appendix Table 4, 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 14.5% lower valuation of soap than control households.

Figure 10 plots the distribution of WTP across each treatment arm for this restricted sample. The bimodality of the incentives distribution highlights the immediate lack of interest in liquid soap by a disproportionate number of incentive households. The figure also plots the household's monthly cost of soap as reported at baseline as a benchmark.

This valuation exercise underscores a problem at the heart of behavioral change in preventative health: health outcomes of preventative 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, 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.[24] 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 to shift households' monetary valuations of soap.

8.3 Behavioral and health spillovers

Despite no obvious changes imposed on dispenser control households throughout the experiment, these households demonstrate a rise and fall in handwashing rates that closely mimics the pattern of monitored households (see Figure 8, for example). This could be due to parallel time trends, the dispenser control households undergoing their own process of habit formation[25], or to spillovers in behavior from neighboring monitored households.

Because treatment assignment between dispenser control and monitoring was randomized at the household level, we capitalize on the random variation in the concentration of monitoring households nearby dispenser control households to estimate the size of spillovers in handwashing behavior.[26] We choose a radius

[24] 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.

[25] A habit formation model in which there exists some fixed cost to pay in consumption stock before σ kicks in is consistent with the initial shallow decay of handwashing rates in dispenser control households (days -70 to 0, see Table 5b) followed by their steady rise (days 0 to 90, see Table 8b). Given that surveyors switched from twice monthly visits to collect health data to once monthly visits to collect data (across all sample households) around day 110, which can be regarded as a negative shock to y_t, the subsequent decay in handwashing rates can also be consistent with the habit formation model.

[26] We define concentration of treated households in levels (number of households) rather than percentages because our sample

of one kilometer around each dispenser control 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 control households, we should only observe the effects of spillovers in the middle regression, and potentially remnants in the third specification.[27] Results are presented in Appendix Table 7. 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 a kilometer of a dispenser household is associated with a 1.6 percentage point (5.7%) increase in dispenser household handwashing rates. This suggests that, while spillovers may have played some role in the handwashing behavior of dispenser households, the parallel pattern we observe is likely also due to parallel time trends or the natural process of habit formation.

Even absent 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 8 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 0.7 percentage points (2.4%) in monitoring villages (coefficients significant at the ten percent level only prior to adjusted for multiple hypothesis testing). Therefore despite substantial positive health internalities, the habit of handwashing at dinnertime produces modest to zero 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

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.

[27] 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).

hand hygiene interventions linked to schools or a child's mid-day meal. This is an important direction for future research.

8.4 Learning

The staggered roll-out of dispensers at the beginning of our study implied that some households were told about their future reward two weeks after receiving the dispenser while others were told two days after receiving the dispenser. We embed this variation in timing of the announcement in an effort to disentangle the effects of learning and experimentation from those of habit formation.[28]

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 addiction, identical current returns to the behavior. In a world of experimentation with risky technology, an increase in the return to future use of a technology should only affect current experimentation with the technology if the agent believes that current experimentation affects her ability to use the technology in the future and thereby reap returns to future use. This is distinct from a world where subsidizing experimentation with a risky technology increases use, where the subsidy is an adjustment to *current* (or constant) returns to the technology, *not* a time-varying adjustment to the return. In former case, returns to experimentation will be realized in the future; in the latter case, returns to future *behavior* are higher than returns to current 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.

[28]This variation also speaks to the malleability of habit formation: households who have not yet determined a routine in dispenser use may be more able to respond to a future change in the value of the behavior than those who have already completed their experimentation and learning regarding the new dispenser.

Figure 1: Simulation of incentives and monitoring arms

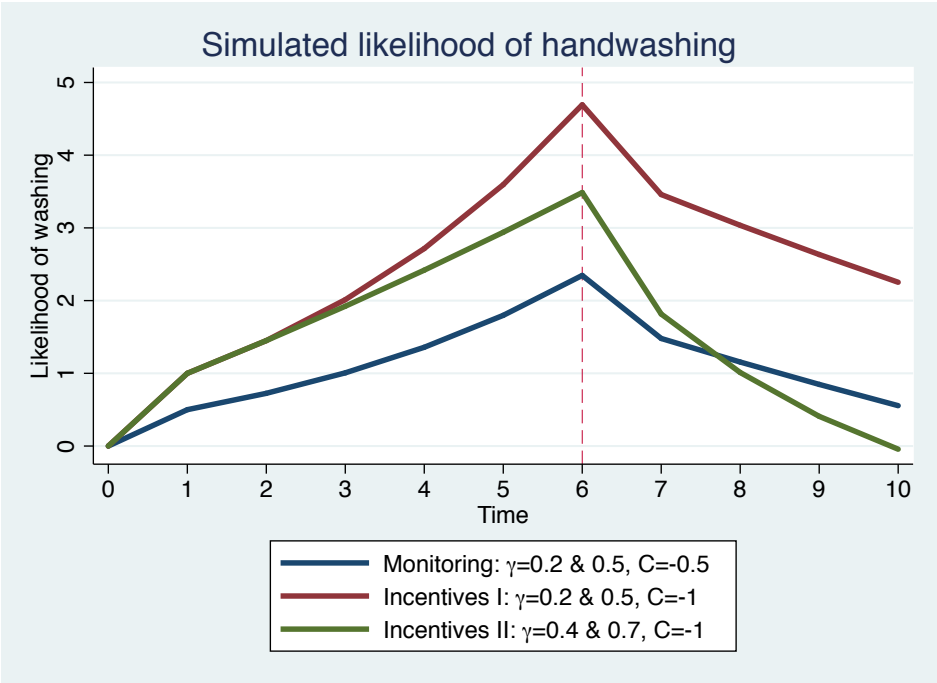


Figure 2: Soap dispenser anatomy

Figure 3: Typical dispenser location

Children using dispenser

Figure 4: Randomization map

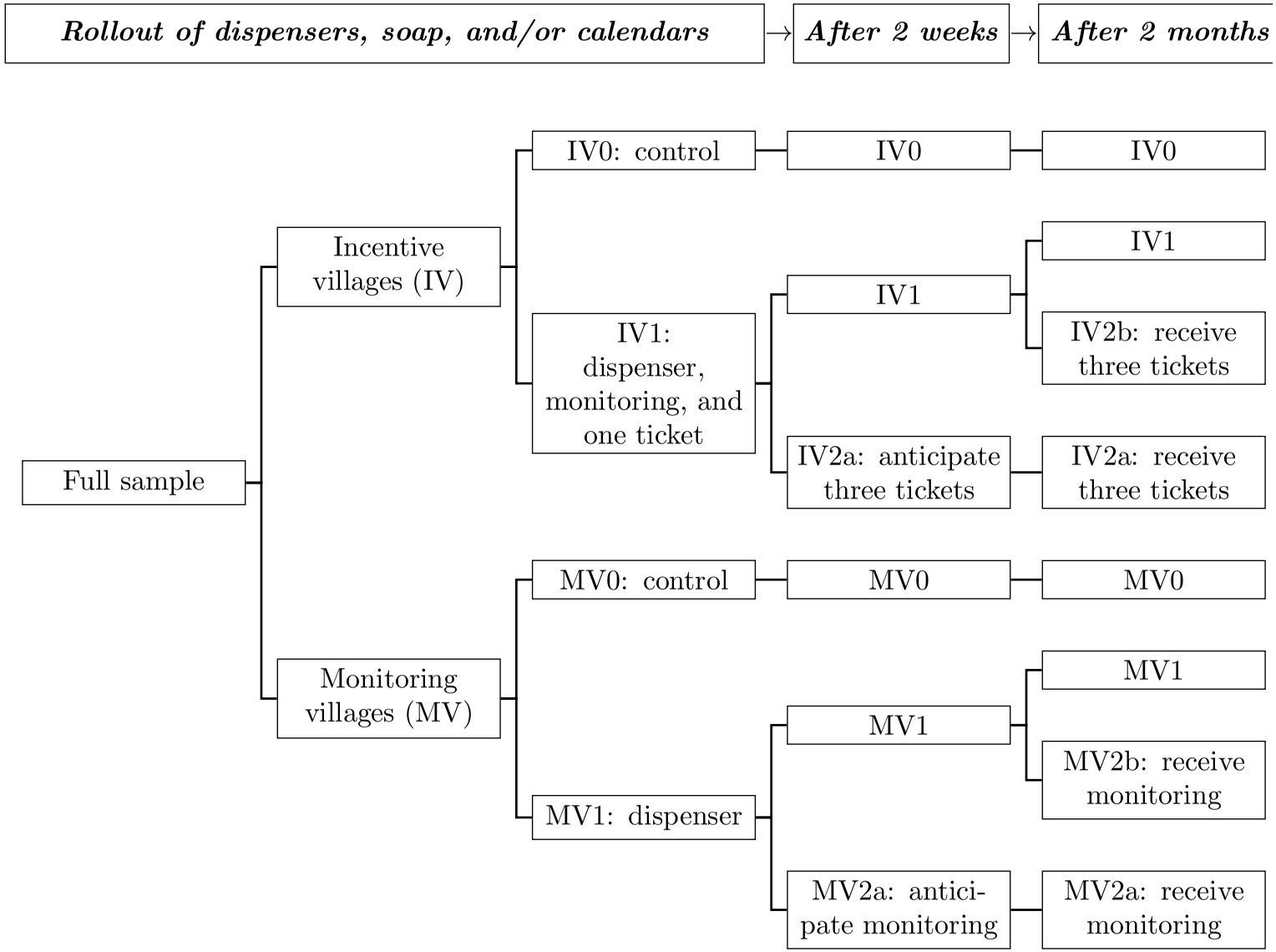
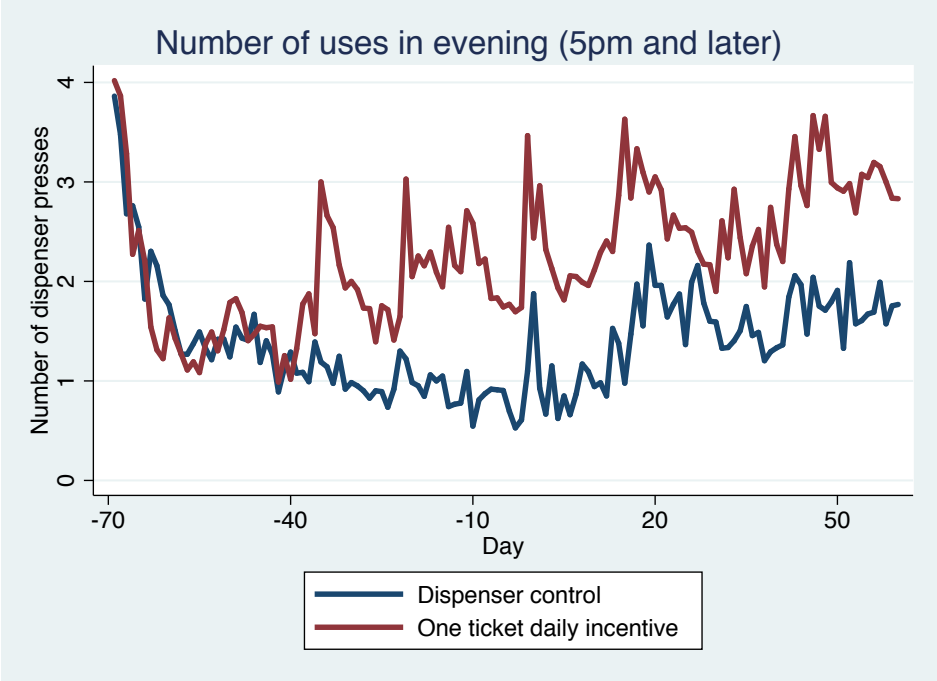


Figure 5a: Dispenser use



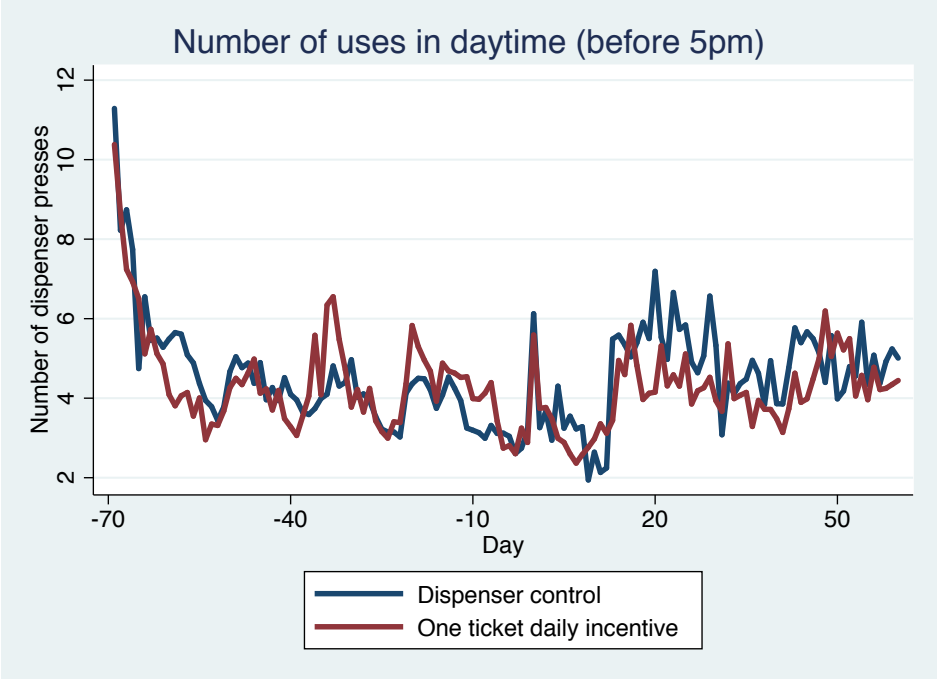


Figure 5b: Binary use at dinnertime

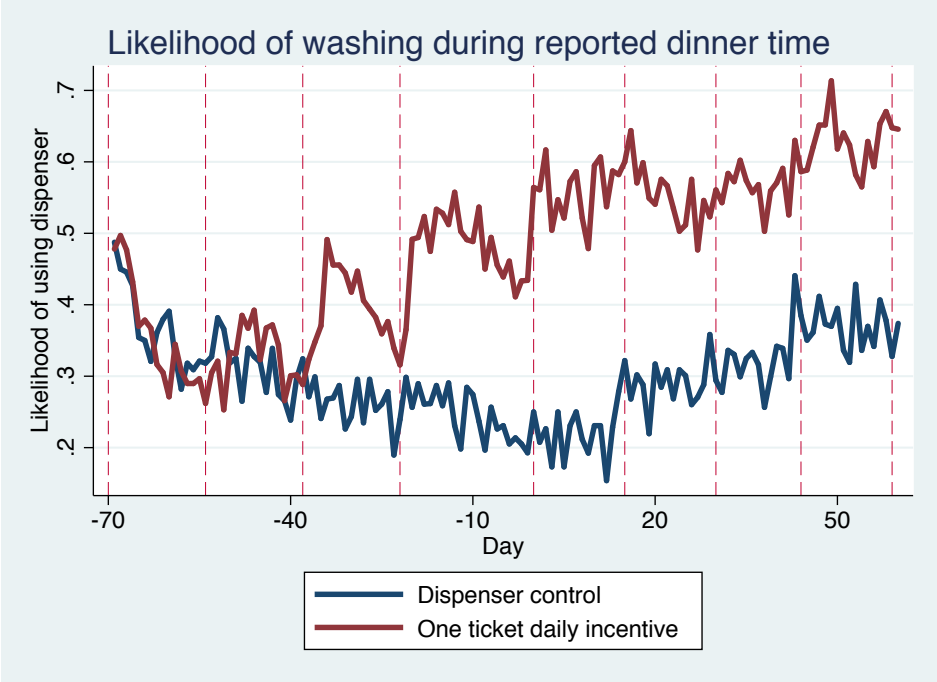


Figure 6a: Incentive effect during intervention regime

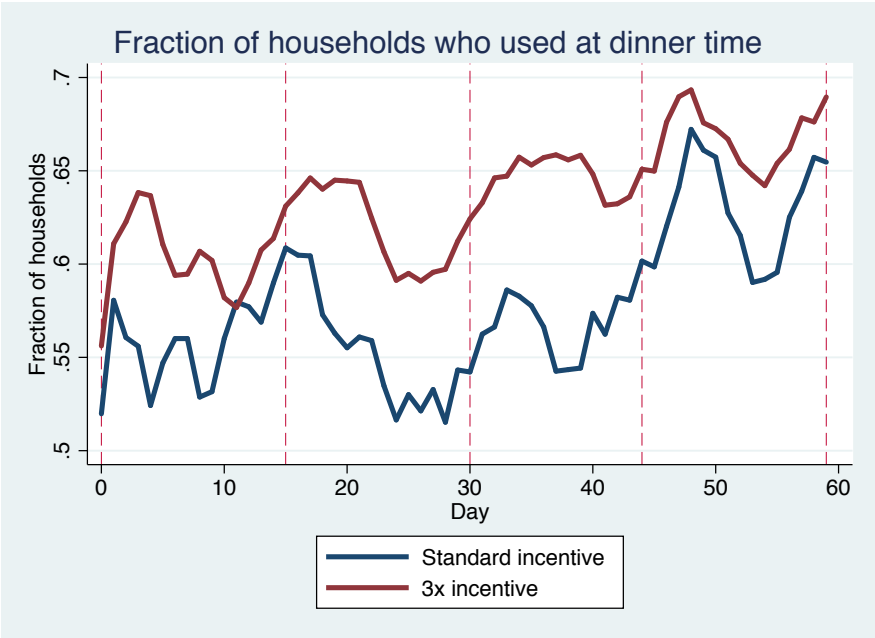


Figure 6b: Monitoring effect during intervention regime

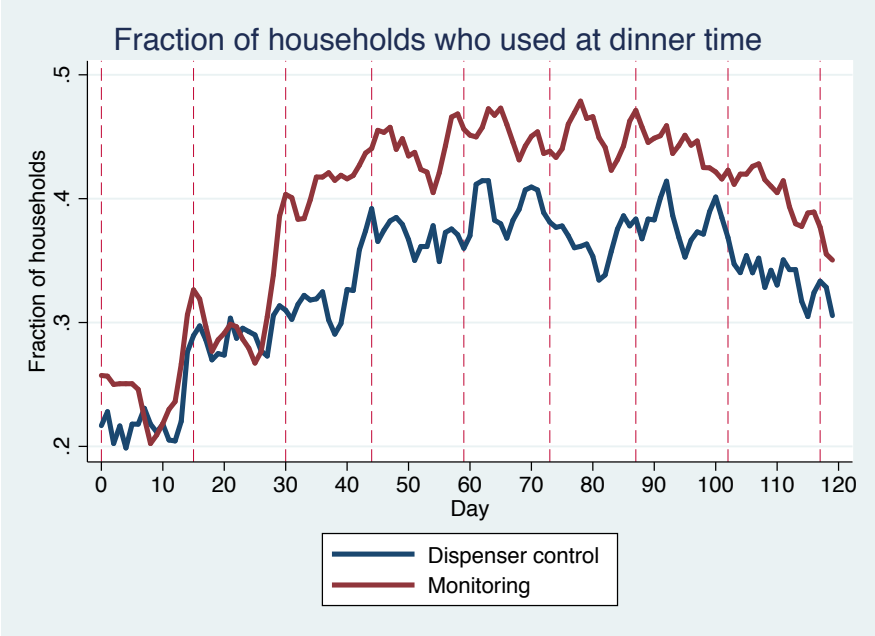


Figure 7a: Persistence of incentives

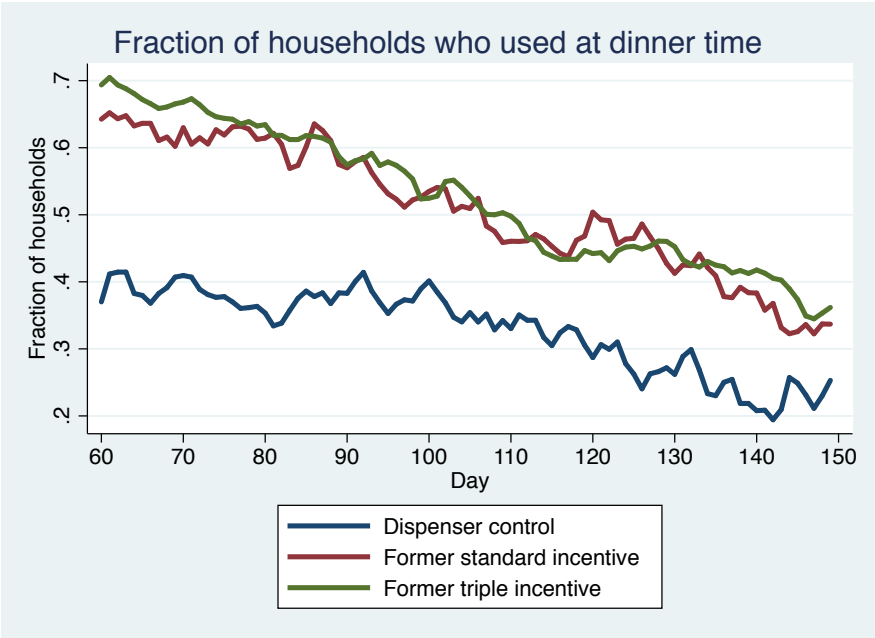


Figure 7b: Persistence of monitoring

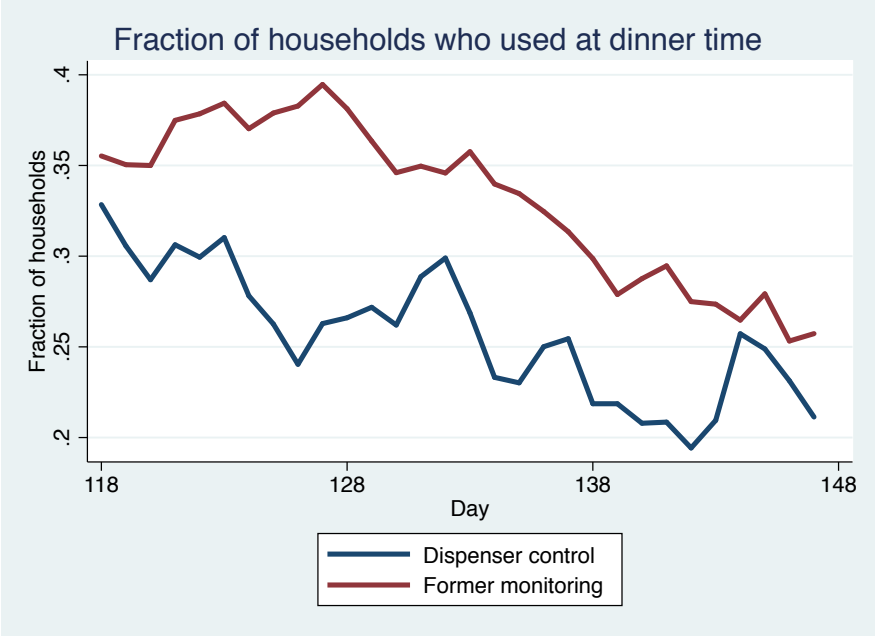


Figure 8: Time trends across treatment arms

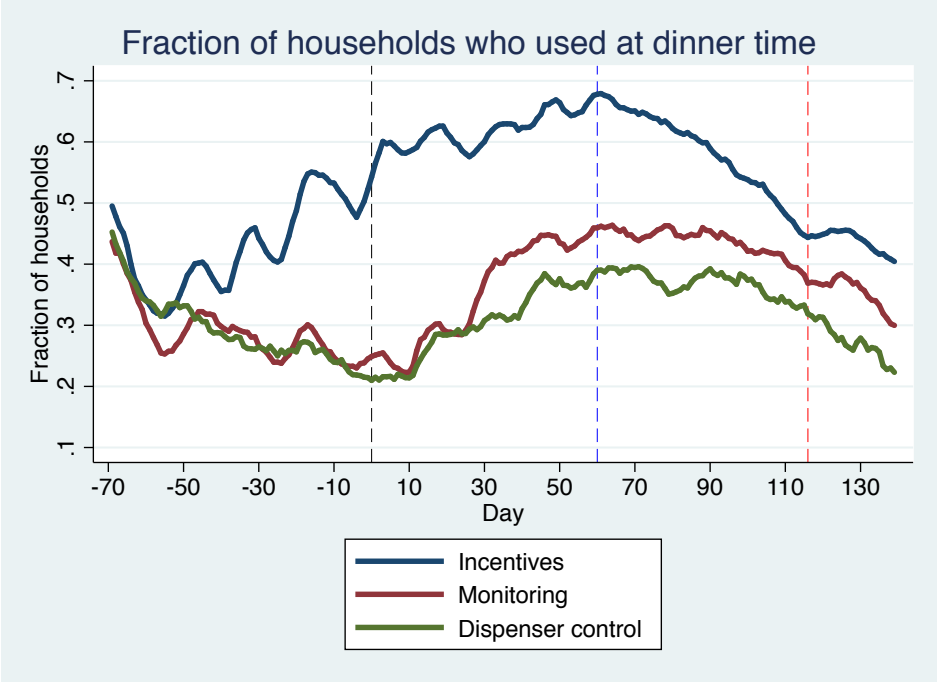


Figure 9a: Rational addiction in incentives

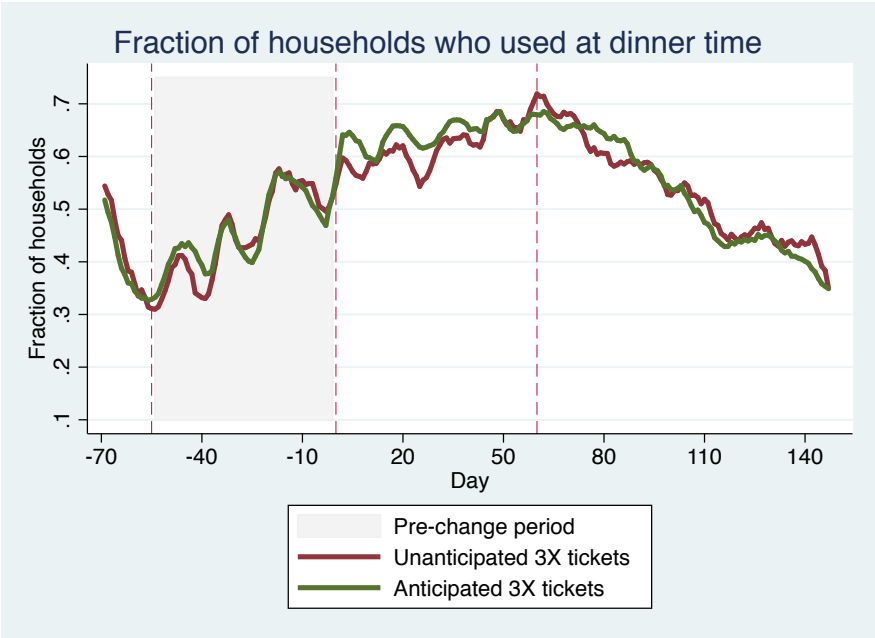


Figure 9b: Rational addiction in monitoring

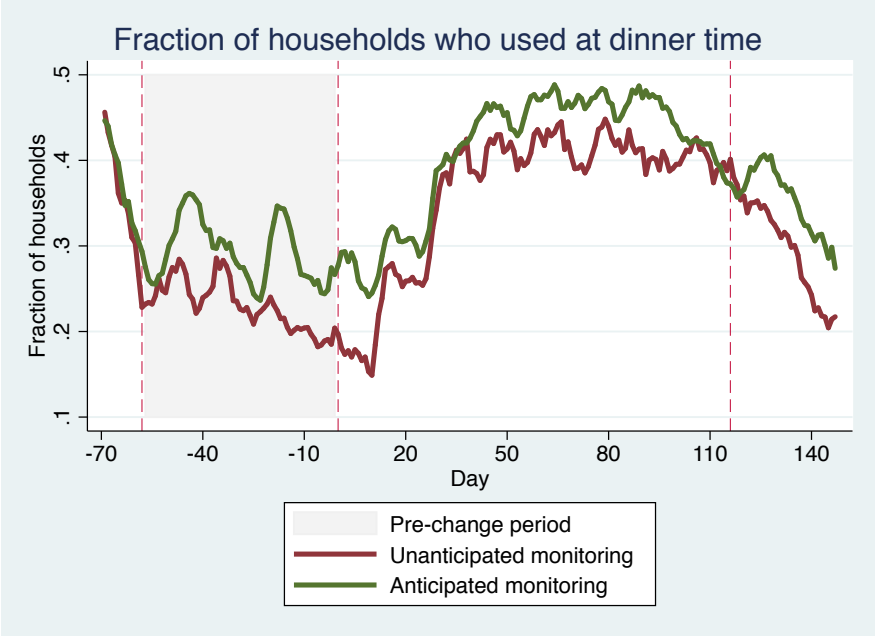


Figure 10: Willingness to pay for soap

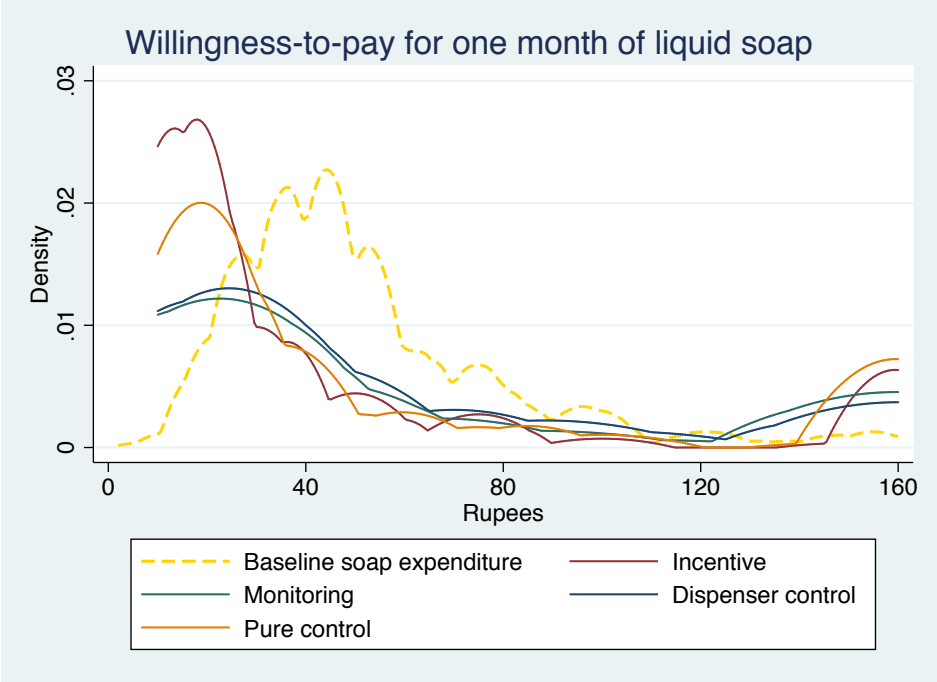


Table 1. Baseline sample means

		Mean	N
Mother and household	Age	31.73	2945
	Education level	6.03	2945
	Hindu	0.72	2945
	Age at marriage	16.72	2945
	Daily labor work	0.55	2945
	Agriculture work	0.21	2945
	Number of rooms in house	2.07	2945
	Deep tubewell used for drinking	55.99	2945
	Distance to drinking source	9.44	2945
	Latrine	0.38	2945
Child	Age (months)	69.43	4829
	Male	0.50	4829
	Height (cm)	104.98	4829
	Weight (kg)	15.22	4829
	# of doctor visits in last two weeks	0.73	4829
	Had cold in last two weeks	0.37	4829
	Had cough in last two weeks	0.08	4829
	Had diarrhea in last two weeks	0.05	4829
Hygiene knowledge	Soap makes hands cleaner than water	94.59	2904
	Soap prevents sickness	80.33	2903
	Soap cleans germs	78.99	2904
	Cold can spread across people	60.70	2903
Hygiene practice	Eat with hands	100.00	2903
	Rinse hands before cooking	96.38	2897
	Wash with soap before cooking	8.60	2897
	Rinse hands before eating	98.83	2900
	Wash with soap before eating	13.95	2875
	Kids wash with soap before eating	30.72	2894
	Reason not wash: no habit	57.09	2454
	Reason not wash: forget	16.87	2454
	Wash with soap after defecation	84.84	2857
	Use soap for bathing	90.41	2898
	Open defecation	67.96	2903
	Has soap in house	99.76	2903
	Monthly soap cost (Rs.)	54.12	2903

Table 2. Balance across treated and control

		Pure control mean	Treated mean	t-statistic	N
Household	Access to electricity	0.954	0.95605	0.261	2,903
	Daily labor occupation	0.543	0.54975	0.358	2,904
	Agriculture occupation	0.217	0.20813	-0.572	2,904
	Number of rooms	2.066	2.0766	0.208	2,900
	Deep tubewell drinking source	0.559	0.56624	0.385	2,904
	Distance to drinking source (min)	9.268	9.664	1.360	2,901
	Latrine	0.379	0.37308	-0.322	2,903
	Mobile	0.770	0.7573	-0.786	2,904
	Breakfast start hour	8.028	8.072	0.803	2,893
	Lunch start hour	12.92	12.9601	1.193	2,893
	Dinner start hour	20.37	20.3809	0.300	2,901
Hygiene and sanitation	Cold can spread	0.611	0.60178	-0.311	2,903
	Soap cleans germs from hands	0.945	0.94684	0.216	2,904
	Number of times hands washed	2.701	2.6876	-0.718	2,904
	Open defecation practiced	0.683	0.67485	-0.460	2,903
Mother	Age (years)	31.67	31.82	0.448	2,908
	Education (years completed)	6.017	6.0422	0.156	2,906
	Hindu	0.727	0.6978	-1.699	2,903
	General caste	0.336	0.3576	1.196	2,899
	Age at marriage	16.41	16.658	2.567	2,885
	People listen	3.001	3.0491	1.504	2,902
	Mother makes child health decision	3.352	3.193	-2.227	2,898
Children below eleven years	Age of child (months)	69.48	69.336	-0.196	4,829
	Male child	0.500	0.4952	-0.303	4,835
	Height (cm)	104.8	105.261	1.022	4,821
	Weight (kg)	15.22	15.1876	-0.187	4,820
	Preventive check up (no. of times 6 mo.)	0.756	0.7	-0.840	1,748
	Sick doctor visit (no. of times 6 mo.)	1.659	1.798	1.364	1,703
	Had cold in last two weeks	0.355	0.3852	1.814	4,827
	Had cough in last two weeks	0.0757	0.08474	0.992	4,771
	Had diarrhea in last two weeks	0.0478	0.0584	1.414	4,832
	Exclusively breastfed (no. of months)	4.698	4.6078	-0.556	3,211

Table 3. Impact of incentives on the extensive margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Two month mark (Day -10 to 0)				Four month mark (Day 50 to 59)			
	Total daily presses	Total presses before 5pm	Total presses 5pm to midnight	Likelihood of use during reported dinner time	Total daily presses	Total presses before 5pm	Total presses 5pm to midnight	Likelihood of use during reported dinner time
One ticket daily incentive	1.676***	0.357	1.224***	0.239***	1.184**	-0.0856	1.250***	0.260***
	[0.622]	[0.493]	[0.215]	[0.0357]	[0.677]	[0.512]	[0.280]	[0.0443]
Mean of dispenser control	2.927	2.605	1.83	0.2	6.544	4.851	1.525	0.369
	[0.592]	[0.531]	[0.340]	[0.0312]	[0.619]	[0.524]	[0.243]	[0.0422]
Observations	3,265	3,265	3,265	3,118	3,046	3,046	3,046	2,957

Notes: Observations are at the household-day level. Robust standard errors in brackets and clustered at village level. All regressions include fixed effects for day. p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1. Households in the one ticket daily incentive group are compared to households in the dispenser control group.

Table 4. Impact of incentives on the intensive margin and monitoring service

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Pooled		Surprised		Pooled		Surprised	
Day	0-59	30-59	0-59	30-59	0-116	30-116	0-116	30-116
	Likelihood of using during reported dinnertime							
Tripled incentive	0.0526**	0.0629**	0.0441**	0.0675**				
	[0.0216]	[0.0237]	[0.0281]	[0.0304]				
Monitoring					0.0705***	0.0837***	0.0581**	0.0734**
					[0.0230]	[0.0247]	[0.0284]	[0.0311]
Mean of control group	0.579	0.598	0.579	0.598	0.341	0.36	0.341	0.36
	[0.0212]	[0.0217]	[0.0212]	[0.0217]	[0.0213]	[0.0219]	[0.0213]	[0.0219]
Observations	31,580	16,781	31,580	16,781	50,308	41,507	50,308	41,507

Notes: Observations are at the household-day level. All regressions include village and day fixed effects. Robust standard errors in brackets are clustered at the household level. Regressions control for average dinnertime handwashing rates at "baseline", which in this case is defined as handwashing rates prior to the announcement (or lack thereof) of the future price boost, which occured on Day -54. Control group for columns 1-4 is the standard (1 ticket) incentive treatment arm. Control group for columns 5-8 is the dispenser control arm. "Pooled" includes data for all households (both those who anticipated and those who were surprised) who were exposed to the intervention; "Suprised" includes only those households who were surprised with the intervention on Day 0. p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Persistence in handwashing after withdrawal of interventions

	(1)	(2)	(3)	(4)	(5)
Day	Day 60-89	Day 90-170	Day 60-89	Day 90-170	Day 116-146
	Likelihood of using during reported dinnertime				
Former standard incentive	0.225***	0.120***			
	[0.0385]	[0.0366]			
Former tripled incentive			0.0324	-0.00137	
			[0.0250]	[0.0242]	
Former monitoring					0.0959***
					[0.0274]
Mean of control group	0.379	0.301	0.619	0.439	0.267
	[0.0269]	[0.0254]	[0.0208]	[0.0212]	[0.0234]
Observations	7866	15,322	16,886	32,289	9,634

Notes: Observations are at the household-day level. Robust standard errors in brackets and clustered at the village level for Columns 1-2 and at the household level for Columns 3-5. All regressions include day level fixed effects; Columns 3-5 also include village level fixed effects. Control group for "Former standard incentive" and "Former monitoring" is the dispenser control; control group for "Former tripled incentive" is the former standard incentive. Columns 1, 3, and 5 estimate effects during the first month after the withdrawal of the relevant intervention; Columns 2 and 4 estimate effects from the second month onwards after withdrawal (we have not yet received data for monitoring household performance in the second month onwards). p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Decay of handwashing behavior after withdrawal of interventions

	(1)	(2)	(3)	(4)
	Incentives		Monitoring	
Day	Day 60-90	Day 90-170	Day 120-150	Day 150-170
	Likelihood of using during dinnertime			
Intervention	0.317***	0.305***	-0.000754	0.0603
	[0.102]	[0.0697]	[0.214]	[0.692]
Slope (Intervention X 10 days)	-0.00956	-0.0141**	0.00785	-0.00229
	[0.0140]	[0.00598]	[0.0165]	[0.0453]
Constant	0.296***	0.131***	-0.227***	-0.0997**
	[0.0369]	[0.0484]	[0.0310]	[0.0467]
Observations	2089	4533	1040	400

Notes: Observations are at the household by ten-day level for ease of exposition. Robust standard errors in brackets; errors clustered at the village level for Columns 1-2 and at the household level for Columns 3-4. [Formerly] incentivized and monitored household behavior is compared to that of the dispenser control households. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Rational habit formation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Likelihood of using during reported dinnertime											
	Prior to intervention				During intervention				After withdrawal of intervention			
	Day -54 to -1	Day -21 to -1	Day -54 to -1	Day -21 to -1	Day 0 to 59	Day 30 to 59	Day 0 to 116	Day 30 to 116	Day 60 to 170	Day 90 to 170	Day 117 to 170	Day 147 to 170
Anticipated triple incentive	-0.00593	-0.0381			0.00999	-8.78e-05			0.0105	-0.0174		
	[0.0235]	[0.0305]			[0.0218]	[0.0223]			[0.0257]	[0.0254]		
Anticipated monitoring			0.052*	0.08**			0.0306	0.0261			0.0204	0.00549
			[0.0243]	[0.0284]			[0.0240]	[0.0257]			[0.0295]	[0.0379]
Mean of control group	0.454	0.536	0.231	0.205	0.609	0.616	0.375	0.409	0.445	0.331	0.311	0.206
	[0.0197]	[0.0257]	[0.0191]	[0.0218]	[0.0192]	[0.0193]	[0.0226]	[0.0249]	[0.0256]	[0.0295]	[0.0278]	[0.0366]
Observations	23,273	9,176	16,268	6,297	48,886	37,270	33,401	26,243	13,239	25,325	7,720	2,902

Notes: Observations are at the household-day level. Robust standard errors in brackets and clustered at the household level for all regressions. All regressions include day and village level fixed effects. Control group for "Anticipated triple incentive" is the group that was surprised with the triple incentive on Day 0; control group for "Anticipated monitoring" is the group that was surprised with the monitoring service on Day 0. p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1.

Table 8a. Daily child diarrhea and ARI outcomes, ITT estimates

	(1)	(2)	(5)	(6)
VARIABLES	Whether child had diarrhea		Whether child had ARI	
Treated household	-0.000866		-0.0222***	
	[0.000718]		[0.00575]	
Incentive		-0.000268		-0.0183**
		[0.000897]		[0.00710]
Monitoring		-0.00214*		-0.0292***
		[0.00127]		[0.0107]
Dispenser control		-0.00141		-0.0288**
		[0.00156]		[0.0146]
Mean of pure control	0.00448	0.00448	0.144	0.144
	[0.000455]	[0.000455]	[0.00361]	[0.00361]
Observations	129,410	129,410	129,410	129,410

Notes: Observations at the child-day level. All regressions include day and village 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. 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. Robust standard errors in brackets and are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1.

Table 8b. Alternative diarrhea and ARI measures at eight months, ITT estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Whether child had any loose stool in last two weeks		Total days of loose stool in last two weeks		Whether child showed any ARI symptoms in last two weeks		Whether child had ARI in last two weeks		Total days of ARI in last two weeks	
Treated household	-0.0230**	-0.0315***	-0.0575**	-0.0817***	-0.0281**	-0.0393**	-0.0148**	-0.0185**	-0.163**	-0.204**
	[0.00849]	[0.00975]	[0.0208]	[0.0236]	[0.0138]	[0.0154]	[0.00877]	[0.0100]	[0.0770]	[0.0884]
Mean of pure control	0.100	0.100	0.209	0.209	0.270	0.270	0.0865	0.0865	1.247	1.247
	[0.00572]	[0.00572]	[0.0151]	[0.0151]	[0.00886]	[0.00886]	[0.00585]	[0.00585]	[0.0504]	[0.0504]
With baseline controls		X		X		X		X		X
Observations	4,940	3,820	4,955	3,830	4,955	3,830	4,955	3,830	4,955	3,830

Notes: Observations are at the child level. Sample includes children younger than fourteen years. Data was collected six to seven months after rollout. "Treated household" is any household that received a dispenser (the pooled sample of incentive, monitoring, and dispenser control households). "Whether child showed any ARI symptoms" equals one if the child experienced any of the following in the two weeks prior: 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. "Whether child had ARIs" is defined as whether the child reported cough *and* nasal congestion or a runny nose in the last two weeks; this is the standard definition of ARI utilized by the partner organization and their surveyors in practice in the field. 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. Robust standard errors are in brackets and are clustered at the household level. p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Child anthropometric outcomes after eight months, ITT estimates

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	WAZ		HAZ		MAZ	
Treated household	0.145*	0.135*	0.217*	0.227*	0.0991*	0.0752*
	[0.0766]	[0.0640]	[0.101]	[0.0902]	[0.0603]	[0.0518]
Mean of pure control	-2.167	-2.167	-1.866	-1.866	-1.365	-1.365
	[0.0459]	[0.0459]	[0.0666]	[0.0666]	[0.0432]	[0.0432]
With baseline controls		X		X		X
Observations	945	863	944	862	940	858

Notes: Observations are at the child level. HAZ represents the height-for-age z-score, WAZ the weight-for-age z-score, and MAZ the midarm circumference-for-age z-score as calculated using WHO anthropometric methodology. Sample in columns 8-10 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 six to seven months after rollout. "Treated household" is any household that received a dispenser (the pooled sample of incentive, monitoring, and dispenser control households). 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. Robust standard errors are in brackets and are clustered at the household level. p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Child health outcomes, TOT estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Whether child had diarrhea on day	Whether child had any loose stool in last two weeks	Total days of loose stool in last two weeks	Whether child had ARI on day	Whether child showed any ARI symptoms in last two weeks	Whether child had ARI in last two weeks	Total days of ARI in last two weeks	WAZ	HAZ	MAZ
Dinnertime dispenser use	-0.000179	-0.0637***	-0.164***	-0.0249***	-0.0668**	-0.0379**	-0.358**	0.254**	0.417**	0.104
	[0.000923]	[0.0200]	[0.0485]	[0.00853]	[0.0317]	[0.0208]	[0.179]	[0.124]	[0.167]	[0.0981]
Mean of pure control	0.00448	0.100	0.209	0.144	0.270	0.0865	1.247	-2.167	-1.866	-1.365
	[0.000455]	[0.00572]	[0.0151]	[0.00361]	[0.00886]	[0.00585]	[0.0504]	[0.0458]	[0.0665]	[0.0432]
Observations	126,268	3,814	3,824	126,268	3,824	3,824	3,824	861	860	856

Notes: Observations are at the child-day level in Columns 1 and 5 and at the child level for remaining columns. Outcome data was collected during February and March (four to five months after rollout) for Columns 1 and 5 and seven to eight months after rollout for remaining columns. For columns 8-10, HAZ represents the height-for-age z-score, WAZ the weight-for-age z-score, and MAZ the midarm circumference-for-age z-score as calculating using WHO anthropometric methodology. Sample in columns 8-10 is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Sample in columns 1-7 include children younger than fourteen. Dispenser use data is the average dinnertime dispenser use from February to July. Regression shows the treatment on the treated estimates where "treated" is a household who uses the dispenser at dinnertime, which is instrumented for by each of the three treatment groups (incentives, monitoring, and dispenser). "Whether child showed any ARI symptoms" equals one if the child experienced any of the following in the two weeks prior: 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. "Whether child had ARIs" is defined as whether the child reported cough and nasal congestion or a runny nose in the last two weeks; this is the standard definition of ARI utilized by the partner organization and their surveyors in practice in the field. All regressions include the following baseline 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. Robust standard errors are in brackets and are clustered at the household level. p-values adjusted for multiple hypothesis testing using Anderson (2008). *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 1. Balance comparisons for disaggregated treatments

	Variable	Control	Incentives	t-stat	N	Control	Monitoring	t-stat	N	Control	Dispenser	t-stat	N
House-hold	Electricity	0.953	0.95	-0.111	1,933	0.955	0.95967	0.373	854	0.955	0.97	0.881	634
	Daily labor	0.534	0.55	0.747	1,934	0.567	0.56285	-0.128	854	0.567	0.4991	-1.383	634
	Agriculture	0.224	0.22	-0.488	1,934	0.200	0.2056	0.223	854	0.200	0.20763	0.220	634
	No. of rooms	2.151	2.19	0.636	1,931	1.848	1.9157	0.909	853	1.848	1.965	1.059	633
	DTW drinking	0.549	0.54	-0.443	1,934	0.585	0.6133	1.468	854	0.585	0.5567	-0.981	634
	Distance to source	8.990	9.36	1.297	1,933	9.976	9.97389	-0	852	9.976	10.384	0.541	633
	Latrine	0.361	0.32	-2.006	1,933	0.425	0.4484	0.771	854	0.425	0.4774	1.192	634
	Mobile	0.773	0.75	-1.302	1,933	0.762	0.7742	0.421	855	0.762	0.8033	1.034	635
	Breakfast time	8.049	8.07	0.516	1,924	7.977	8.09	1.970	851	7.977	8.151	2.077	631
	Lunch time	12.90	12.94	1.094	1,924	12.97	12.9638	-0.171	851	12.97	12.9843	0.231	631
	Dinner time	20.38	20.37	-0.252	1,931	20.34	20.4161	1.221	852	20.34	20.34256	0.0381	632
Hygiene and sanitation	Cold can spread	0.628	0.62	-0.206	1,933	0.568	0.5785	0.233	855	0.568	0.5821	0.212	635
	Soap cleans germs	0.945	0.94	-0.0461	1,934	0.945	0.94249	-0.191	855	0.945	0.94986	0.282	635
	Times handwash	2.743	2.74	-0.0608	1,934	2.594	2.6249	1.025	855	2.594	2.6312	0.944	635
	Open defecation	0.697	0.72	0.930	1,934	0.648	0.623	-0.800	854	0.648	0.585	-1.392	635
Mother	Age (years)	31.71	31.70	-0.0320	1,934	31.58	32.346	1.548	855	31.58	31.46	-0.184	635
	Education (years)	5.974	5.67	-1.551	1,933	6.128	6.462	1.179	854	6.128	6.819	1.663	634
	Hindu	0.772	0.77	-0.0939	1,933	0.613	0.6242	0.636	854	0.613	0.545	-2.595	634
	General caste	0.291	0.26	-1.432	1,930	0.450	0.4671	0.625	853	0.450	0.5474	2.471	633
	Age at marriage	16.33	16.42	0.824	1,922	16.61	17.044	2.519	847	16.61	16.846	0.920	630
	People listen	2.987	3.02	0.853	1,933	3.034	3.0867	0.905	854	3.034	3.0221	-0.162	635
	Child health dec.	3.477	3.35	-1.580	1,930	3.036	2.9738	-0.509	853	3.036	3.0113	-0.144	634
Children below eleven years	Child age (mo.)	69.88	69.09	-0.858	3,225	68.50	69.9	1.057	1,406	68.50	69.469	0.589	1,070
	Male child	0.504	0.50	-0.0748	3,231	0.491	0.506	0.505	1,406	0.491	0.4553	-0.895	1,070
	Height (cm)	105.2	105.14	-0.108	3,221	103.8	105.385	1.824	1,403	103.8	105.33	1.372	1,067
	Weight (kg)	15.35	15.23	-0.491	3,221	14.92	15.091	0.713	1,402	14.92	15.308	1.172	1,068
	Check up visit (#)	0.768	0.80	0.415	1,167	0.726	0.6616	-0.740	512	0.726	0.8136	0.577	390
	Sick visit (#)	1.678	1.78	0.880	1,139	1.616	1.737	0.796	498	1.616	1.719	0.533	382
	Cold	0.337	0.36	1.419	3,228	0.401	0.427	0.870	1,404	0.401	0.4272	0.646	1,065
	Cough	0.0816	0.09	0.434	3,171	0.0614	0.0799	1.205	1,404	0.0614	0.0784	0.856	1,068
	Diarrhea	0.0442	0.05	1.192	3,228	0.0567	0.056384	-0.0224	1,406	0.0567	0.0686	0.566	1,070
	Breastfed (mo.)	4.525	4.71	0.991	2,170	5.137	4.542	-2.257	906	5.137	4.529	-2.369	687

Appendix Table 2. Child diarrhea and ARI outcomes after six months disaggregated by treatment arm

VARIABLES	(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) Whether child had ARI in last two weeks	(5) Total days of ARI in last two weeks
Incentives	-0.0280**	-0.0756**	-0.0291	-0.0255*	-0.179
	[0.0135]	[0.0311]	[0.0204]	[0.0136]	[0.112]
Monitoring	-0.0293	-0.0933**	-0.0457	-0.00807	-0.214
	[0.0182]	[0.0439]	[0.0288]	[0.0200]	[0.179]
Dispenser control	-0.0622***	-0.120**	-0.0809**	-0.0280	-0.296
	[0.0240]	[0.0592]	[0.0396]	[0.0224]	[0.253]
Mean of pure control	0.104	0.215	0.273	0.0903	1.267
	[0.00597]	[0.0153]	[0.00914]	[0.00615]	[0.0525]
Observations	3,491	3,501	3,501	3,501	3,501

Notes: Observations are at the child level. 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 prior: 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. "Whether child had ARIs" is defined as whether the child reported cough and nasal congestion or a runny nose in the last two weeks; this is the standard definition of ARI utilized by the partner organization and their surveyors in practice in the field. All regressions include the following baseline 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. Robust standard errors are in brackets and are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 3. Child anthropometric outcomes after six months disaggregated by treatment arm

VARIABLES	(1) HAZ	(2) WAZ	(3) MAZ
Incentives	0.136	0.190*	0.0356
	[0.0847]	[0.107]	[0.0632]
Monitoring	0.00408	0.156	0.173
	[0.182]	[0.183]	[0.108]
Dispenser control	1.258	0.358	0.155
	[1.017]	[0.353]	[0.149]
Mean of pure control	-2.168	-1.848	-1.365
	[0.0478]	[0.0693]	[0.0450]
Observations	791	789	785

Notes: Observations are at the child level. HAZ represents the height-for-age z-score, WAZ the weight-for-age z-score, and MAZ the midarm circumference-for-age z-score as calculating using WHO anthropometric methodology. Sample in columns 8-10 is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodolog. Data was collected six to seven months after rollout. Baseline controls are included in all regressions and consist of: 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. Robust standard errors are in brackets and are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 4. Alternative hygiene measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Observed hand cleanliness		Observed nail cleanliness		Whether handwashing habit was achieved		Whether household has non-project liquid soap		Total dispenser presses in 24 hours	
VARIABLES										
Treated	0.0674***		0.122***		1.478***		0.0456***			
	[0.0214]		[0.0263]		[0.0392]		[0.00932]			
Incentives		0.0562**		0.132***		1.587***		0.0623***	0.583**	
		[0.0273]		[0.0344]		[0.0476]		[0.0116]	[0.239]	
Monitoring		0.105***		0.108**		1.269***		0.0158		-0.0164
		[0.0381]		[0.0436]		[0.0729]		[0.0165]		[0.358]
Dispenser		0.0410		0.0974		1.318***		0.0165		
		[0.0495]		[0.0625]		[0.0979]		[0.0257]		
Mean of control group	1.552	1.552	1.179	1.179	1.615	1.615	0.0548	0.0548	5.354	3.291
	[0.0167]	[0.0168]	[0.0206]	[0.0206]	[0.0306]	[0.0306]	[0.00565]	[0.00565]	[0.188]	[0.440]
Observations	2,672	2,672	2,671	2,671	2,669	2,669	2,670	2,670	951	458

Notes: Observations are at the householdlevel in Columns 1-8 and at the child-day level in Columns 9-10. "Treated" households are the pooled sample of incentive, monitoring, and dispenser control households. All regressions use the pure control group as the comparison group. All regressions include village level fixed effects except column 9, which compares treatment arms across villages. Columns 9 and 10 have a restricted sample because the outcome variable is only observed amongst treated households. Therefore for these two columns, the relevant control group is the dispenser control. In all other regressions, the relevant control group is the pure control. Observed hand and nail clenliness 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 acheived is rated by the respondent on a 5 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." Robust standard errors are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 5. Sanitation outcomes

VARIABLES	(1)	(2)	(3)	(4)
	Whether household defecates in open		Whether household treats drinking water	
Treated	-0.000599		0.00642	
	[0.0175]		[0.0102]	
Incentives		0.0232		0.00993
		[0.0214]		[0.0128]
Monitoring		-0.0434		0.000136
		[0.0325]		[0.0180]
Dispenser control		-0.0420		0.000222
		[0.0467]		[0.0298]
Mean of pure control	0.647	0.647	0.0861	0.0861
	[0.0119]	[0.0119]	[0.00696]	[0.00697]
Observations	2,672	2,672	2,669	2,669

Notes: Observations are at the household level. All regressions include village level fixed effects. Robust standard errors are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 6. Willingness to pay for soap at six months

	(1)	(2)	(3)
VARIABLES	Willingness to pay (Rs.)		
Treated household	-4.738**		
	[1.935]		
Incentive		-9.060***	-7.755***
		[2.303]	[2.393]
Monitoring		1.415	-0.681
		[3.705]	[3.706]
Dispenser control		6.011	4.500
		[5.243]	[5.434]
Mean of pure control	55.74	55.74	53.64
	[1.476]	[1.477]	[1.439]
Observations	2,750	2,750	2,478

Notes: Observations are at the household level. All regressions include village level fixed effects. Robust standard errors are clustered at the household level. 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<0.01, ** p<0.05, * p<0.1.

Appendix Table 7. Spillovers in handwashing rates

	(1)	(2)	(3)
VARIABLES	Likelihood of use during reported dinner time		
	Days -40 to -30	Days 40 to 50	Days 120 to 130
No. of monitored households	-0.00794	0.0162*	0.00830
	[0.00659]	[0.00838]	[0.00863]
Constant	0.252	0.279	0.230
	[0.0355]	[0.0399]	[0.0412]
Observations	1,106	1,165	1,019

Notes: Observation at the household level. Sample is all dispenser control households. Independent variable is the number of monitored households within 1 km of the dispenser control household. Robust standard errors in brackets and clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 8. Health spillovers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Whether child had any loose stool in last two weeks		Total days of loose stool in last two weeks		Whether child showed any ARI symptoms in last two weeks		Whether child had ARI in last two weeks		Total days of ARI in last two weeks	
No. of treated households, MV	-0.00220		-0.00576		-0.00234		0.000304		-0.0286	
	[0.00203]		[0.00487]		[0.00377]		[0.00210]		[0.0215]	
No. of treated households, IV		-0.000767		-0.000983		-0.00672*		0.00109		-0.0337*
		[0.00249]		[0.00575]		[0.00359]		[0.00253]		[0.0183]
Mean of pure control	0.108	0.0867	0.216	0.181	0.275	0.277	0.0937	0.0664	1.452	1.342
	[0.0162]	[0.00955]	[0.0385]	[0.0248]	[0.0270]	[0.0147]	[0.0170]	[0.00947]	[0.175]	[0.0873]
Observations	624	1,601	629	1,602	629	1,602	629	1,602	629	1,602

Notes: Observations at child level. Sample is the number of children in pure control households in each type of village (monitoring village (MV) or incentive village (IV)). Independent variable is the number of treated households (pooled incentive, monitoring, and dispenser) within 1 km of the pure control household. Robust standard errors in brackets and clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1

9 Bibliography

Alcott, H. Kessler, J. (2015). The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. Working Paper.

Alcott, H. Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. American Economic Review, 104(10): 3003-37.

Auld, M.C. Grootendorst, P. (2004). An Empirical Analysis of Milk Addiction. Journal of Health Economics, 23(6): 1117-33.

Banerjee, A., Duflo, E., Glennerster, R., Kothari, D. (2010). Improving immunization coverage in rural India: clustered randomized controlled evaluation of immunization campaigns with and without incentives. BMJ. 340: c2220.

Becker, G.S., Murphy, K.M. (1988). A Theory of Rational Addiction. The Journal of Political Economy. 96(4): 675-700.

Benabou R, Tirole J. (2003). Intrinsic and extrinsic motivation. Review of Economic Studies. 70:489-520.

Bennett, D., Naqvi, SAA, Schmidt, WP. (2015) Learning, Hygiene, and Traditional Medicine. Working Paper.

Condry, J. Chambers, J. (1978), *Intrinsic Motivation and the Process of Learning*, in M. Lepper and D. Greene (eds.) The Hidden Cost of Reward: New Perspectives on the Psychology of Human Motivation (New York: John Wiley).

Clasen, T., Boisson, S., Routray, P., Torondel, B., Bell, M., Cumming, O., Ensink, J., Freeman, M., Jenkins, M., Odagiri, M., Ray, S., Sinha, A., Suar, M., Schmidt, W. (2014). Effectiveness of a rural sanitation programme on diarrhea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. The Lancet Global Health. 2(11): e645-e653.

Duhigg, C. (2012). The Power of Habit. Random House Trade Paperbacks, New York.

Dupas, P. (2011). Health Behavior in Developing Countries. Annual Review of Economics, Vol. 3.

Dupas, P. Miguel, E. (2016). Impacts and Determinants of Health Levels in Low-Income Countries. NBER Working Paper No. 22235.

Gneezy U, Meier S, Rey-Biel P. (2011). When and why incentives (don't) work to modify behavior. Journal of Economic Perspectives. 25:191-210.

Gruber, J. Koszegi, B. (2001). Is Addiction 'Rational'? Theory and Evidence. The Quarterly Journal of Economics 116(4): 1261-1303.

Haggerty, P. A., Muladi, K., Kirkwood, B. R., Ashworth, A. Manunebo, M. (1994). Community-based hygiene education to reduce diarrhoeal disease in rural Zaire: impact of the intervention on diarrhoeal morbidity. Int J Epidemiol 23, 1050-1059.

Halder, A., Tronchet C., Akhter S., Bhuiya A., Johnston R., and Luby S. (2010). Observed Hand Cleanliness and Other Measures of Handwashing Behavior in Rural Bangladesh. BMC Public Health, September 2010, 10.

Han, A. Hliang, T. (1989). Prevention of diarrhea and dysentery by hand washing. Trans R Soc Trop Med Hyg. 83(1):128-31.

Heyman J, Ariely D. (2004). Effort for payment: a tale of two markets. Psychol. Sci. 15:787-93.

Ito, K., Ida, T. Tanaka, M. (2015). The Persistence of Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand. Working Paper.

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling. Quarterly Journal of Economics, 125(2): 515-548.

Rapson, J. Jessoe, K. (2014). Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use. American Economic Review. 104(4): 1417-1438.

Kremer, M. Zwane, A. (2007). Cost-Effective Prevention of Diarrheal Diseases: A Critical Review. Center

for Global Development Working Paper Number 117.

Luby, S. P., Agboatwalla, M., Painter, J., Altaf, A., Billhimer, W. L. Hoekstra, R. M. (2004). Effect of intensive handwashing promotion on childhood diarrhea in high-risk communities in Pakistan. JAMA 291, 2547-2554.

Luby, S., Halder A., Huda T., Unicomb L., and Johnston R. (2011). Using Child Health Outcomes to Identify Effective Measures of Handwashing. American Journal of Tropical Medicine and Hygiene, 85(5): 882-892.

Ludwig, J., Kling J.R., Mullainathan S. (2011). Mechanism experiments and policy evaluations. Journal of Economic Perspectives, 25 (2011): 17-38

McKay, S., Gaudier E, Campbell D., Prentice A., and Albers R. (2010). Environmental Enteropathy: New Targets for Nutritional Interventions. International Health, 2: 172-180.

Mobarak, A.M., Dwivedi P., Bailis R., Hildemann L., and Miller G. (2012). Low Demand for Nontraditional Cookstove Technologies. Proceedings of the National Academy of Sciences, 109(27):10815-10820.

O?Donoghue, T., M. Rabin, (2001). Addiction and Present-Biased Preferences. Working Paper.

Ruel, M. Arimond M. (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.

Taubinsky, D. (2014). From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice. Working Paper.

WHO Multicentre Growth Reference Study Group (2006). WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl. Apr(450): 76-85.

Wishnofsky, M. (1958). Caloric Equivalents of Gained or Lost Weight. American Journal of Clinical Nutrition, 6(5): 542-546.