HARVARD | BUSINESS | SCHOOL

Habit Formation and Rational Addiction: A Field Experiment in Handwashing

Reshmaan Hussam
Giovanni Reggiani
Atonu Rabbani
Natalia Rigol

Working Paper 18-030

Habit Formation and Rational Addiction: A Field Experiment in Handwashing

Reshmaan Hussam
Harvard Business School

Atonu Rabbani
Dhaka University

Giovanni Reggiani
Massachusetts Institute of Technology

Natalia Rigol
Harvard University

Working Paper 18-030

Habit Formation and Rational Addiction: A Field Experiment in Handwashing

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

September 12, 2017

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 scholarship that suggests 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 develop and produce a novel soap dispenser, within which a time-stamped sensor is embedded. We randomize distribution of these soap dispensers as well as provision of monitoring (feedback reports) or monitoring and incentives for daily handwashing. Relative to a control arm in which households receive no dispenser, we find that all treatments generate substantial improvements in child health as measured by child weight and height. Our key test of rational addiction is implemented by informing a subset of households about a future boost in monitoring or incentives. We find that (1) both monitoring and incentives increase handwashing relative to receiving only a dispenser; (2) these effects persist after monitoring or incentives are removed; and (3) the anticipation of monitoring increases handwashing rates significantly, implying that individuals internalize the habitual nature of handwashing and accumulate habit stock accordingly. Our results are consistent with the key predictions of the rational addiction model, expanding its relevance to settings beyond what are usually considered 'addictive' behaviors.

*We are indebted to research assistant Sami Safiullah for his excellent field work and continual 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: in particular, the 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 the sensors, as well as AQS, who manufactured them. We received guidance from Esther Duflo, Abhijit Banerjee, Frank Schilbach, Ben Olken, Chris Udry, Dean Karlan, Mushfiq Mubarak, Lasse Brune, Hannah Trachtman, and David Levine, as well as helpful comments from conference and seminar participants at Berkeley, Chicago, MIT, and Yale. 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.

†Harvard Business School; corresponding author: rhussam@hbs.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 (World Bank 2005), 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 no impact on behavior or health (WSP 2012, WSP 2013, Galiani et al. 2015). The few that do are intensive 'omnibus' interventions (including information, resources, community involvement, monitoring, and other hygiene and sanitation recommendations), which are difficult to replicate in practice and do not generate clear evidence on the key mechanisms 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 the difficulty of sustained change is that, in order to be repeated as often as needed, the new 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." The need for repetition is not unique to handwashing: preventive 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 preventive health behaviors suffer low rates of takeup in the developing world despite their affordability, and neither information provision nor materials and/or infrastructure provision appear to generate sustained improvements in such practices (Dupas and Miguel 2016, Clasen et al. 2014, Kremer and Zwane 2007, Banerjee et al. 2010, WSP 2012, WSP 2013). Given their repetitive, reflexive nature and ties to contextual cues, the psychology literature highlights such behaviors as ideal candidates for habit formation (Wood and Neal 2007).

In this study, we examine whether handwashing is indeed a habit-forming activity and, in addition, whether individuals internalize and respond to its habitual nature, and explore 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, overcoming the identification concerns typical to the literature on rational addiction.

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 per dispenser. This technology addresses the key problems of standard handwashing outcomes: desirability bias (hand hygiene is self-reported or 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 households, yielding more objective data; it is precise, measuring use at the second level and allowing us to connect observed use with critical times of use (such as prior to eating); and it tracks the use of liquid soap, which is uniquely associated in our study context with handwashing rather than with bathing or laundry.

In our conceptual framework, habit formation is generated by intertemporally linked preferences in consumption: the more one consumes in the past, the easier or more likely is consumption in the present. Intertemporal complementarities imply that front-loaded (i.e. temporary) interventions, which maximize initial takeup of a behavior, can generate a larger stock of consumption - and thereby persistence in behavior - than interventions that are spread over time. While there could be several reasons for why temporary subsidies have persistent effects (such as the adoption of new technology or learning about the returns), persistence due to habit formation is derived purely from changes in consumption stock. This mechanism in turn yields additional predictions given by theory.

In particular, Becker and Murphy (1988), who popularized the theoretical framing of habits (or equivalently, addictive behaviors), 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 a behavior is indeed habit-forming but agents fail to internalize this feature in their consumption decisions, they will underinvest, justifying short-term subsidies to boost usage. Alternatively, if agents are rational habit formers, an intervention that increases the future value of an activity will generate a larger habit stock - and thereby more persistence - than one that increases the present value of the activity. In its starkest form, a large, one-time incentive to engage in the future will motivate a rational habit former to engage today (and increasingly so as the future nears), but will have zero effect on the non-rational habit former. Understanding the nature of the behavior (habitual or not) and how an agent conceptualizes the behavior (rationally or otherwise) is thus important for the optimal design of interventions.

In addition to their potentially habitual nature, a second feature of preventive health behaviors is that the returns to the activity, by virtue of being preventive, are not salient. Agents' perceived returns may therefore be lower than the true returns to the activity. In this setting, incentives that offer agents tangible and immediate returns to good behavior may be an effective way to increase takeup.

A third common feature of such behaviors is the absence of social norms around preventive health: given ubiquitously low takeup rates, individuals have no expectations to engage and thus face minimal repeated social costs to shirking. In this setting, an intervention that monitors activity and thereby invokes social pressure to engage may effectively change behavior.

Our experiment is designed to test for the presence of habit formation and rational habit formation in the context of one such preventive health behavior: handwashing with soap. We draw from the psychology literature on habit formation and the features of our measurement device 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 (monetary or social incentives) (Duhigg 2011, Aunger 2010, Neal et. al 2015).

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 the first, we inform households that we are monitoring their activity with the sensor technology and provide reports on daily handwashing performance (a form of social incentives). In the second, we additionally offer daily financial incentives for handwashing in the form of tickets that can be redeemed for household goods. In both cases, social and financial incentives are removed after four months, and we continue to track behavior. Persistence in handwashing after the withdrawal of incentives is consistent with handwashing being a habit-forming activity. Our key test of rational habit formation enters when we experimentally vary whether agents *anticipate* these interventions: to a subset of households, we announce two months in advance that they should look forward to receiving a monitoring service or extra daily incentives at a future specified date.[1] A present reaction to anticipated changes in future handwashing behavior confirms that individuals are aware of the intertemporal complementarities between performance today and performance tomorrow.

We find that, relative to those who receive only a dispenser, monitoring succeeds in raising short run handwashing rates by 23%. These higher handwashing rates persist strongly after the withdrawal of the service, suggesting that handwashing is indeed a habit-forming activity. 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 non-anticipating counterparts, with the difference increasing as

[1] Note that *all* households are notified that the future monitoring service or incentive boost is a possibility at the beginning of the experiment. They are told that resources are limited, so whether or not they should anticipate (receive) these future boosts will be determined by lottery. If households use the partner organization's incentive scheme as a signal of the true returns to handwashing, this lottery method equalizes expected value of handwashing *as judged by the partner organization* across anticipating and non-anticipating households.

the date of the monitoring service approaches. It appears, therefore, that households do indeed recognize the habit-forming nature of handwashing, and they additionally act upon this knowledge by accumulating consumption stock in preparation of a future rise in the consumption value of handwashing.

Adding financial incentives to monitoring likewise increases handwashing rates substantially. Relative to those who received monitoring only, those who additionally receive financial incentives wash 25% more frequently. Relative to those who receive a dispenser only, financial incentives raise short run handwashing rates by more than 70%. After incentive withdrawal, higher handwashing rates persist for several months, substantiating evidence on the habitual nature of the activity. However, these effects are not mirrored on the intensive margin of financial incentives: those who experience an *increase* (in particular, a tripling) in financial incentives wash only 8% more than their standard incentive counterparts, suggesting rapidly diminishing marginal returns to financial incentives. This slightly higher rate of washing decays to the standard incentive level soon after incentives are withdrawn. In line with the small contemporaneous and persistence effects, we also find no anticipated reaction to the tripling of incentives: households anticipating this change wash no more than their non-anticipating counterparts.

Our results are consistent with the key predictions of the rational addiction model. When faced with a future increase in incentives, households choose not to invest in accumulating handwashing stock to 'prepare' given the low contemporaneous benefit or prospects of habit formation associated with the change. On the other hand, households invest considerably in accumulating stock for an intervention with significant contemporaneous and long run bite, as evidenced by the monitoring setting.

Lastly, 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 social or monetary 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 74% fewer days of loose stool and 28% fewer days of ARI when we examine the impact of the treatment on the treated, where 'treated' is defined as those who self-report regularly washing their hands at the eight month mark.[3] These reductions in morbidity translate to significant improvements in child weight-for-age and height-for-age: treated children experience a 0.14-0.17 standard deviation increase in their weight-for-age and a 0.23-0.26 standard deviation increase in their height-for-age (ITT-TOT estimates) eight months after dispenser distribution.

This study makes five contributions. First, to our knowledge, this is the first field experiment designed to test for rational habit formation. Existing literature in rational addiction employ non-experimental time

[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.

[3] This measure is correlated with actual dispenser use (correlation of 0.15), the latter of which we cannot employ in an IV regression since our comparison group of pure control households do not have dispenser use data.

series data vulnerable to several identification concerns: price instruments are endogenous, consumption data is self-reported, knowledge of a future change in price is implausible, and serial correlation in prices yields false positives in favor of rationally addictive behavior (Auld and Grootendorst 2004). The experimental design of our study systematically addresses each issue that has previously challenged causality. This is also the first study to examine rational habit formation in the context of *good* habits, an important feature of preventive health behaviors.

Second, this study advances the measurement of habit formation, even apart from the test of rational addiction. Existing literature typically equates habit formation with long run persistence of temporary interventions. However, persistence can be due to multiple mechanisms: the purchase of a technology that changes the production function; the process of learning more about an activity such that one updates her desire or ability to engage; or the accumulation of consumption stock. Habit formation is driven only by the latter. Importantly, our evidence of rational habit formation must be due to changes in future consumption stock because we experimentally vary only the future value of handwashing behavior, not that of health returns. Paired with an experimental design that includes a dispenser experimentation period across all arms and evidence that handwashing behavior does not vary with the size of child health returns, we can rule out the most relevant alternative mechanisms of persistence and highlight habit formation as the driving force behind persistence in handwashing.

Third, by identifying the marginal impacts of financial incentives, social incentives, 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 aspect of the intervention led to the health improvements. More generally, interventions that employ financial incentives often conflate two mechanisms in their estimated treatment effect of incentives: in any setting in which a conscious principal is rewarding an agent for her behavior, a financial incentive is a sum of (1) the financial reward and (2) social monitoring and feedback on performance. Particularly in contexts where returns are not salient (i.e. preventive health behavior) and there are no social costs to shirking (i.e. behaviors that are not social norms), it is important to estimate the impact of each mechanism alone.

Fourth, 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 preventive technologies. It complements the recent collection of energy conservation studies in developed countries that utilize household-level meter data from energy utility companies to examine how various informational interventions affect household energy consumption (Allcott and Rogers 2014, Ito et al. 2014, Jessoe and Rapson 2015, Allcott 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 preventive 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 experiment; Section 4 specifies our outcomes of interest and the empirical strategy; Section 5 presents results on handwashing behavior; Section 6 presents results on child health; 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 2.1. Throughout our discussion, we use 'addiction' and 'habit formation' interchangeably, as their underlying mechanisms are identical.

Intertemporal complementarities in the utility from consumption are an intrinsic property 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: a rational habit former is one who internalizes the habit forming nature of the activity, or the craving and tolerance developed through continual engagement, and chooses to engage conditional on this knowledge. The key tradeoff that a rational habit former faces when choosing whether to engage in a good habit is therefore 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 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

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

where $x_t \geq 0$ is the net cost associated with handwashing before dinnertime.

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

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

The model has two key features. First, a good habit generates positive internalities: the more one has washed her hands in the past, the greater her current wellbeing 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 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 $\sigma \geq 0$). 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 generated by habituation (σ). The desire to engage $d_t(k)$ is positive iff

$$x_t - \sigma(\gamma k_{t-1} + w_{t-1}) < 0 \tag{4}$$

A myopic agent will only wash her hands today if the marginal cost of washing, reduced by the benefits of habituation, is less than zero. 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 nor σ 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 net cost of handwashing, x_t. If sufficiently lowered, an agent will wash, raising $k_{t-1} > 0$. If the cost is lowered for sufficiently long, the agent will accumulate enough consumption stock such that, even absent the subsidy on cost, the desire to engage will be positive. In a setting of habit formation, subsidies need only be temporary to generate long run behavioral change.

We actualize the reduction in net cost x_t in two ways. Our first intervention subsidizes the cost of daily handwashing by providing daily financial incentives for good behavior. However, in a context where incentives are directly linked to countable units of behavior and a sentient principal is providing the incentives to the agent (a setting not unique to this study), incentives are implicitly a sum of both (1) financial rewards and (2) feedback on behavior and the social pressure of being observed. The latter mechanism can be conceptualized as the imposition of a cost to shirking. In order to disentangle the relative importance of each in reducing the net cost to handwashing, we implement both an incentives intervention and a feedback sans financial reward intervention, which we term 'monitoring.'

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 by providing households with no feedback, monitoring feedback only, or additionally daily incentives for handwashing (described in more detail in Section 3.3).

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:

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

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, such as a drop in the future cost of handwashing, changes her likelihood of engaging in the future, she should update accordingly her likelihood of engaging today.

In summary, the model yields the following testable implications.

1. **Incentives:** $\frac{\partial d_t}{\partial x_t} \leq 0$. Reducing the cost of handwashing (by increasing the value of handwashing) raises handwashing rates.

2. **Monitoring:** $\frac{\partial d_t}{\partial x_t} \leq 0$. Reducing the cost of handwashing (by increasing the cost of *not* handwashing) raises handwashing rates.

3. **Habit formation:** $\frac{\partial d_t}{\partial k_t} \geq 0$. A rise in past handwashing rates increases current handwashing rates.

4. **Rational habit formation:** $\frac{\partial^2 U_t}{\partial k_t \partial k_{t+1}} \geq 0$. An anticipated [and actual] rise in future handwashing rates is associated with an increase in current handwashing rates.

2.2 Empirical evidence on rational habit formation

The vast majority of the literature, all of which explores bad habits such as smoking and alcohol consumption, rests in favor of rational addiction (Becker et al. 1991, Chaloupka 1991, Cameron 1998, Baltagi and Griffin 2002, Gruber and Koszegi 2001). The typical 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 the 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).

However, Auld and Grootendorst (2004) describe the implausible variation in discount rates, unstable demand, and low price elasticities implied by such literature. 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 prices of the commodity of interest and endogeneity in the price instruments can yield a positive coefficient on future consumption that is incorrectly interpreted as evidence of rational addiction. Of significant added concern is the implausibility of consumer knowledge of future price changes in the contexts explored in the literature.[4] These features are directly tied to the nature of the non-experimental, aggregate time-series data employed in existing scholarship.

Our field experiment addresses each concern above. Our design allows us to: (1) impose price changes exogenously, avoiding endogeneity between prices or tax rates and consumption; (2) explicitly announce future prices so consumer knowledge is assured; (3) avoid concerns of differential time trends given randomization; (4) avoid endogenous misreporting using our objective measurement device; and (5) avoid the implications of serial correlation in commodity prices as we impose prices exogenously and randomization permits us to compare outcomes across groups rather than over time.

3 Experimental Design

3.1 Study sample and context

Our sample population is made up of 2,943 peri-urban and rural households containing 3,763 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 old 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 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.[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, the Society for Health and Demographic Surveillance (SHDS), is a public

[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) 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.

[5]While these numbers are low, they are likely to be overestimates given self-reporting.

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. SHDS surveyors had 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, a practice that we continued for the duration of the present study.

3.2 Dispenser and soap features

We employed a standard wall-mounted dispenser as depicted in the top picture of Figure 1, which was outfitted with a time-stamped 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' biweekly visits. The sensor module is fit between the container and the soap spout, as shown in the bottom picture of Figure 1. 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 2 depicts a typical setting for the dispenser: families usually eat on a mat in 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 the dispenser 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 to find that households preferred: (1) unscented or lightly scented soap that would

[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.

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 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. Henceforth, treatments associated with social incentives will be referred to as "monitoring", and those associated with financial incentives will be referred to as "incentives."

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 + incentive. Recall that receiving financial incentives implicitly involves receiving feedback and monitoring. 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 households 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 of appreciation 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.

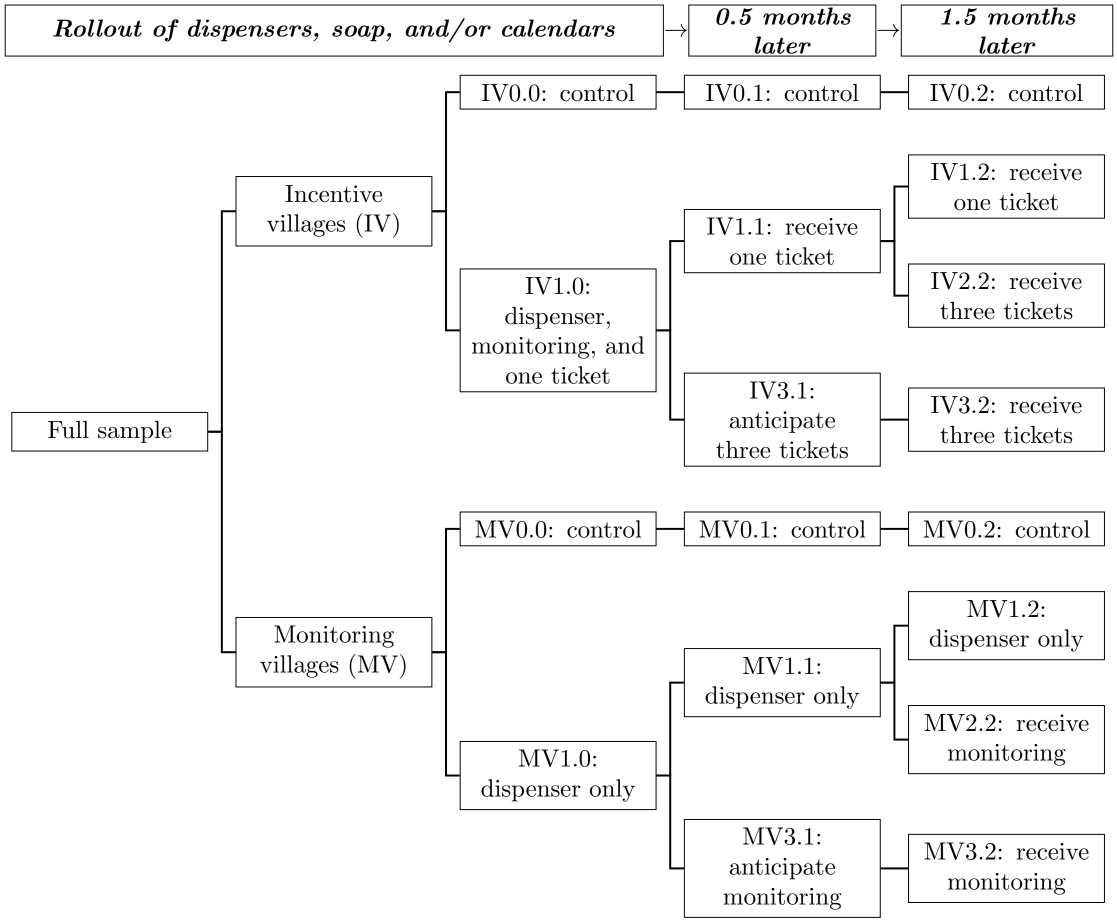
The remainder of the randomizations were conducted at the household level, with households in monitoring villages and those in incentive villages experiencing a parallel evolution in treatments over time. Each treatment arm is described in detail below.

(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 dispenser, which 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 handwashing would be reported biweekly and tracked on their calendar. Because resources were limited, the service would be administered by lottery. If they did not get selected, their switch would not be turned on and their behavior would not be monitored.[7]

[7]These lotteries were publicly announced in order to equalize the expected value of the monitoring 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 service was a signal that they should value the behavior more.

Figure 4: Randomization map



(MV3) Anticipated monitoring: Two weeks after dispenser distribution, these households were informed that they had been selected in the lottery: the internal switch would soon be turned on, and the device would record the time and frequency with which the household washed their hands with soap.[8] The surveyor would be carefully observing this data every two weeks 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 information and feedback, third-party monitoring, and self (or parent-child/intrahousehold) monitoring. The service would begin two months after dispenser distribution on a date circled clearly by the surveyor on

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

the household calendar and written on a sticker attached to the dispenser. This upcoming date was reannounced at each proceeding surveyor visit to ensure comprehension.

(MV2) Unanticipated monitoring: Two months after dispenser distribution, these households were surprised with an identical monitoring service to those in MV3, effective immediately.[9]

(IV1) Incentives: At the point of dispenser distribution, these households were informed that there was a switch in their dispenser which, when on, tracked the frequency and time of use; and that this switch was on and 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. Households 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]

(IV3) Anticipated triple incentives: Two weeks after dispenser distribution, these households were informed that they had been selected in the lottery for the incentive boost and could soon expect to receive 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 and written on a sticker attached to the dispenser. As in the monitoring scenario, this date was reannounced at each proceeding surveyor visit to ensure comprehension.

(IV2) Unanticipated triple incentives: Two months after dispenser distribution, these households were surprised with an identical incentive boost to those in IV3, effective immediately.

[9] As with the households in MV3, 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 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 addition to such logistical concerns, we also designed 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.

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 MV3 relative to MV1 and IV3 relative to IV1 (*before* the price change) demonstrates the presence of rationally habit forming behavior: households who increase 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 MV3 versus MV1 and IV3 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 change in the value of handwashing 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 (i.e. the tripling of tickets or monitoring service provision) when it is enacted. This contemporaneous effect can be identified by comparing households in MV1 to those in MV2 and households in IV1 to those in IV2 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 contemporaneous 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, IV2, and IV3, all incentives [and implicitly, monitoring] services were discontinued approximately two months after the price change. For households in arms MV2 and MV3, 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 surveyors 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.[14]

[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 significant policy relevance of the lessons learned. In particular, this practice allows us to estimate the effects of (1) third party monitoring and feedback, yielding a measure of the extent of bias in typical observational outcome measures used in these studies as well as a measure of the role that monitoring effects may play in the cultivation of social norms; and (2) persistence after the withdrawal of interventions, yielding a measure of the sustainability of the interventions and the habit-forming nature of handwashing.

[14]We equate persistence to habit formation under the assumption that persistence is driven purely by the increase in con-

We track household handwashing behavior for fourteen months after rollout (this collection process remains ongoing in the field).

By maintaing the same incentive stream across both groups, a comparison of MV3 to MV2 and IV3 to IV2 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 3 and 2 groups demonstrates whether forward-looking behavior in fact facilitated the formation of the handwashing habit.

Finally, comparison of MV2 to IV1 offers the first estimate in the literature of the marginal value of monetary rewards on top of monitoring and feedback on daily behavior.[15] [16]

4 Methods

4.1 Outcomes of Interest

Our primary outcomes of interest encompass behavioral change in households and child health. We capture behavioral change through recorded dinner time-specific daily handwashing rates and recorded total daily handwashing rates. Note that 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 therefore supplement these with alternative measures of hand hygiene commonly employed in the literature. We collect child health data in the form of self-reported biweekly incidence of child diarrhea and respiratory illness and anthropometric measures of height, weight, and mid-arm circumference. Each is defined in detail below.

A. Household handwashing behavior

sumption stock accumulated through the interventions, not through the acquisition of a technology that shifts households onto a new hand hygiene path or through learning about the activity or about its returns. These are not trivial assumptions, and we address each in detail in Sections 5.2 and 6.1.

[15] Note that the experimental design precludes perfectly capturing the effect of incentives on top of monitoring, although it is quite close. Monitoring was introduced (MV3 and MV2) 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 power 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 possibility 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.

[16] Indeed, this concern is echoed by early economics literature on behavioral change. In describing the law of diminishing utility in his *Principles of Economics*, Marshall writes: "There is, however, an implicit condition in this Law 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."

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 elucidated 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 the outcome by which we determine calendar markings and tickets earned, and therefore our primary outcome measure of handwashing at dinnertime.[17]

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 at baseline 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. These survey questions were adjusted at the eight month mark to account for many relevant cases

[17]We choose this binary measure as our preferred measure of "proper" handwashing because we wanted to minimize Type II error in our feedback: we preferred that households were overcompensated for washing than undercompensated due to stricter and less verifiable measures of success (such as "all family members must wash", which is both harder to achieve and more difficult to verify), which in turn might diminish treatment effects.

being excluded given the strict initial definitions (described in detail in Section 6); this cross-sectional measure at eight months is our primary incidence outcome measure.

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 ages 60 months and below (WHO, 2006).

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

Pre-change (rational habit formation) period is defined as the time between dispenser distribution and the monitoring service introduction/price change. We also zoom in on the three week period just prior to the date of change. This is because (1) we showed a video three weeks prior to the date of change 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.

Intervention period is defined as the two months following the price change for incentive households (IV2) and four months following the monitoring service for monitoring households (MV2). For IV1 households only, this period is defined as the four months after rollout. These are the times during which pure intervention effects can be measured.

Persistence (habit formation) period is defined as the period of 2-12 months after the price change for incentive households and 4-12 months after the monitoring service change for monitoring households.

4.3 Empirical strategy

Our preferred specification for our primary behavioral outcomes is as follows:

$$Wash_{hvt} = \alpha_{hvt} + \beta Treatment_{hvt} + \delta BaselineWash_{hvt} + \gamma_t + \theta_v + \epsilon_{hvt} \tag{6}$$

in which $Wash_{hvt}$ represents the outcomes specified above, $Treatment_{hvt}$ is the assigned treatment for each subset of comparisons described in Section 3, $BaselineWash_{hvt}$ represents the baseline value of the outcome variable, 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 rollout, we omit day fixed effects.

Our preferred specification for our primary child health outcomes is as follows:

$$Health_{cvt} = \alpha_{cvt} + \beta Treatment_{cvt} + \delta BaselineHealth_{cvt} + \theta_v + \epsilon_{hvt} \tag{7}$$

in which H_{cvt} represents the outcomes specified above, $Treatment_{cvt}$ is the assigned treatment group specified in the analysis, $BaselineHealth_{hvt}$ represents the baseline value of the outcome variable, and θ_v is village fixed effects. Standard errors are clustered at the household level.

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 likewise show no obvious patterns in differences across treatment arms and control; 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 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 only 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, respectively, 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 occurring around Day 0), 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 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 binary 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, which they would begin earning immediately. 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 decay, 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 there exist no intertemporal complementarities in the behavior (measured by 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 examined these two features of handwashing can we consider whether agents are *rational* about their habit formation in our setting. We thus 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 contemporaneous impact of an intensive margin shift in incentives on handwashing. Columns 1 and 2 of Table 4 present the results for the comparison between households who were surprised with 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 exposed to the boost (i.e. earning triple tickets), 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, though modestly, to the tripling of daily tickets: they washed an average of 2 percentage points more than their single ticket counterparts over the duration of the triple ticket regime, increasing to a statistically significant 5 percentage points (8.3%) upon receiving the extra tickets in hand.

Figure 6a plots the three-day moving average of dinnertime handwashing rates for the tripled incentive arm relative to the standard incentive arm before and after the incentive boost. Note that the regression results of Table 4 control for the pre-trends evident in the plot.

5.1.3 Monitoring

Columns 3 and 4 of Table 4 estimate the contemporaneous impact of the monitoring service on household handwashing behavior as compared to dispenser only households.[18] Column 3 presents results for the full tenure of the monitoring service, while Column 4 presents the lagged results of Days 30 to 116. The monitoring service has a statistically significant and substantial impact on behavior, increasing handwashing rates by 7.1-8.4 percentage points (21-23%) over the duration of the service provision.

Figure 6b presents the three-day moving average of dinnertime handwashing rates for monitored households relative to those who received the dispenser only. The graph demonstrates how household behavior to the monitoring arm reacts most strongly on the day of the first calendar receipt (Day 30), highlighting the important role of a public feedback mechanism in the effectiveness of the monitoring service. As with incentives, the regression results in Table 4 control for the pre-trends evident in the plot.

5.2 Persistence

Section 5.1 establishes that the experiment exogenously increased the value and consequently the 'consumption stock' of handwashing in each treatment arm, albeit substantially more under the monitoring regime

[18]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.

than the triple ticket regime. This addresses our first and second testable implications: $\frac{\partial d_t}{\partial x_t} \leq 0$. We now explore whether this exogenous shift in stock had an impact on subsequent handwashing behavior after the interventions ceased.

Many studies have examined the role of temporary interventions on persistence in the long run (Charness and Gneezy 2009, Conley and Udry 2010, Allcott and Rogers 2014, Royer et al. 2015, Dupas 2010, among others). The persistence of temporary interventions does not readily imply habit formation, however. Persistence can be generated by the purchase of a technology that changes the production function; the process of learning more about a technology (whether in how to use it, what the optimal set of inputs is, or what the returns are) such that one updates her desire to engage; or the accumulation of consumption stock. Habit formation is driven only by the latter. Isolating this mechanism is a challenge, and existing studies lack the data or the context to distinguish the effects of consumption stock accumulation from learning or technology acquisition.

In the present study, we can easily rule out the first alternative mechanisms behind persistence: because our outcome measure is the likelihood of dispenser use, it will not capture the effects of any other hygiene-related technology the household may acquire. Additionally, we find no changes in sanitation or water treatment practices (see Appendix Section 8.1), suggesting that households do not invest in alternative technologies that may alter their hand hygiene production function. In contrast, the mechanism of learning is a greater challenge to address, since the process of engaging in an activity repeatedly generates both learning about the activity and a growing stock of consumption.

We identify three dimensions of learning that can occur in our context: (1) learning how to physically wash ones hands; (2) learning how to use the handsoap dispenser; and (3) learning about the health returns to handwashing. We argue that the extent of learning required for the washing process is negligible: 99% of households already rinse their hands with water before mealtime, and 100% of households own and thus are familiar with the use of soap; to combine the two activities should require minimal learning and there is little reason to expect this to be differential across treatment groups. The extent of learning required for using the handsoap dispenser, which is a novel technology, may be greater; to address this, we allow a two week learning period between the rollout of the dispensers and the assignment to treatment during which all households can get acquainted with the dispenser. As is evident in Figure 5b, households do indeed experiment with the dispenser technology for the first ten days, but behavior stabilizes thereafter, suggesting that this learning is largely complete within the first two weeks prior to treatment assignment. Finally, households may persist in their handwashing because, by washing more, they also learn that handwashing leads to improvements in health, and therefore update their beliefs of the returns to the behavior. Upon presenting the child health results in Section 6, we offer evidence that households who experience larger child health returns are no more likely to persist in their handwashing behavior than those who experience small child health returns, suggesting that this dimension of learning plays a minimal, if any, role in handwashing persistence.

We therefore interpret persistence in handwashing behavior after the withdrawal of the interventions 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 the marginal utility of handwashing.

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 and 4). 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 only counterparts during the first month after incentive withdrawal; this drops to 12 percentage points over the following two months, suggesting some decay of the consumption stock. 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, Column 5 demonstrates 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 third testable implication: $\frac{\partial d_t}{\partial k_t} \geq 0$.

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

5.3 Rational habit formation

Having established that handwashing is a habitual activity and that the interventions change, to varying degrees, 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 7a. 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. The coefficient of interest is small and imprecise, offering no evidence that anticipation of a future price change affects current handwashing 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.

Table 7b explores whether these patterns continue throughout the rest of the experiment. Columns 1-4 examine the behavior of (now formerly) anticipating households relative to their surprised counterparts over the course of the triple ticket or monitoring interventions, and Columns 5-6 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 8a, which plots the five-day moving average of handwashing behavior between anticipating and nonanticipating triple ticket households. Household behavior follow essentially identical patterns over the course of the experiment. In contrast, Figure 8b, 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 through the remainder of the experiment. Although the estimates are not statistically significant, formerly anticipating households wash 3.1 percentage points more than their surprised counterparts during the monitoring regime (Columns 3-4), decreasing to 1.8 percentage points one month after the monitoring service has stopped (Column 7).

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 several 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. Second, 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. Third, the impending date was circled in red on the household calendar and written on a sticker attached to the handsoap dispenser. Finally, if households did indeed believe that the monitoring service started on the day of announcement rather than Day 0, then we should expect their patterns of response to be similar to those of households who were surprised with the monitoring (had no anticipation) on Day 0. However, if we compare the pattern of behavior between MV3.1 and MV2.2 (see Figure 8b), we see little in common: anticipating households respond sharply on visit days and decay nearly as sharply afterwards, while households actually being monitored respond sharply on visit days and remain responsive, steadily increasing their handwashing rates over time. Although not definitive given the different time periods being compared, this offers strongly suggestive evidence that anticipating households were not confused regarding the timing of the monitoring service.

An alternative argument for why we see a rational addiction effect among households anticipating monitoring but not those anticipating triple tickets is one of salience. Perhaps the term "monitoring" makes the act of handwashing more salient today than the term "tripling of tickets." (Note that this is distinct from a salience argument in which being monitored in the future makes handwashing more salient today than receiving triple tickets in the future: if salience from future activity were the mechanism, responding to such salience would fall within the realm of rational habit formation.) Although there is little one can do to ensure equal salience of handwashing across related terminology, we offer the following test: on the days of surveyor visits, it is likely that the salience of handwashing is maximized regardless of what treatment arm one is in. Surveyors ask households about the dispenser, how it is functioning, whether there are any problems; they ask the household to forecast how many days mothers and children anticipate washing their hands before mealtime in the coming week; and they replenish the soap and perform "maintenance" on (i.e. collect data from) the dispenser. Indeed, handwashing rates spike even in the dispenser only arm (Figure 8b) on the visit days. If we interpret these spikes as the effect of maximizing the salience of handwashing, then it is evident that anticipation of future monitoring has an *additional* impact on handwashing above and beyond that of salience alone.

Our set of behavioral results instead appears to be entirely consistent with the predictions of the rational addiction model. While the contemporaneous effects of monitoring were substantial (up to 20% more handwashing than the relevant control mean), those of the tripling of tickets were smaller and less precise (up to 8% more handwashing than the relevant control mean). Similarly, while the persistence of monitoring was substantial (essentially zero decay in the first month), that of triple tickets decayed rapidly, suggesting that the contemporaneous impact was too small to facilitate habit formation. Consistent with the utility function of a rational habit former, households chose not to invest in accumulating handwashing stock to 'prepare' for an intervention with little contemporaneous benefit ($\frac{\partial d_t}{\partial x_t} \approx 0$) or prospects of habit formation ($\frac{\partial d_t}{\partial k_t} \approx 0$). On the other hand, they invested considerably in accumulating stock for an intervention with significant contemporaneous and long run bite, confirming our fourth testable prediction: $\frac{\partial^2 U_t}{\partial k_t \partial k_{t+1}} \geq 0$.

6 Health results

Thus far, we have established that handwashing is a habit forming behavior and that households in our sample are sophisticated about its habitual nature. It remains an open question, however, whether the habit of handwashing is worth acquiring in the high-disease environment of West Bengal, where the marginal value of this simple activity may be small given the high exposure to disease from other sources. We now ask: 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.[19]

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.[20]

Our final set of health outcomes comes from the midline survey as well: we recollect anthropometric measures of child height, weight, and mid-arm circumference. We combine these measures with child age and gender data to create the child's height-for-age, weight-for-age, and mid-arm circumference-for-age Z-scores based on guidelines by the World Health Organization (WHO). This analysis is restricted to children 0 to 60 months as specified by the WHO guidelines. [21]

[19] 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.

[20] The wider net was cast as follows: mothers and children together were asked whether the child had experienced *any* loose stool motion in the last two weeks. If so, the days they experienced loose stool 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. Any amount less than three was not recorded as an episode. 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 by 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 together 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 these symptoms, the surveyor then asked how many days the child 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.

[21] The WHO provides the distribution of each age and gender-specific anthropometric measure for a reference population of well-nourished children from Brazil, Ghana, India, Norway, Oman and the United States, such that a Z-score of 0 is the median of the reference population. We place a special focus on height-for-age as a metric of child health improvement, as linear growth is regarded as the most relevant indicator of overall nutrition (Hoddinott et al. 2013).

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 3 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 and 4 disaggregate this sample into each treatment group: incentives, monitoring, and dispenser control households.

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 2.9 percentage points, or 20.1%.

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. 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 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. Weight-for-age increases by 0.14 standard deviations, height-for-age increases by 0.23 standard deviations, and mid-arm circumference-for-age increases by 0.08 standard deviations. 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.[22] Appendix Table 3a disaggregates these results into each treatment arm, and Appendix Table 3b disaggregates

[22]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).

by age of child. Unsurprisingly, younger children (one to two years of age) benefit most in weight, height, and mid-arm circumference.

Since the average rate of handwashing at dinnertime among treated households is 47%, we now consider estimates of the treatment on the treated (TOT). However, because control households were not given a dispenser, we cannot use dispenser use as a proxy for handwashing in this instrumental variables exercise. Instead, we employ two alternative hand hygiene measures we collected across all sample households: self reports on whether the mother and child wash regularly (whether they have achieved a handwashing habit) and enumerator obsevations of hand cleanliness. Both measures, and especially the self reports, are correlated with dispenser use and thus seem like reasonable proxies for handwashing (Appendix Table 4, Column 3). For ease of interpretation, we transform both measures into binary variables: the self report is equal to one when the respondent articulates that a habit has been achieved and zero otherwise; the enumerator observation is equal to one when she records clean hands and zero otherwise. We instrument each with treatment assignment, employing the three treatment groups of incentives, monitoring, and dispenser households as instruments.

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

$$Wash_{hv} = \alpha_{hv} + \beta_1 Incentives_{hv} + \beta_2 Monitoring_{hv} + \beta_3 Dispenser_{hv} + \epsilon_{hv}$$
$$Health_{chv} = \alpha_{chv} + \beta_2 \hat{Wash}_{hv} + \delta_{chv} + \theta_v + \epsilon_{chv}$$

in which $Wash_{hv}$ is either the self report or the enumerator observation, δ_{chv} is a vector of child health baseline controls and θ_v represents village fixed effects.

Table 10 presents the TOT estimates. A child in a household that reports regularly washing at dinnertime experiences a 59% decrease in the likelihood of having loose stool, a 74% decrease in the number of days she experiences loose stool, a 24% decrease in the likelihood of experiencing any ARI symptoms, and 27% fewer days of ARI. She also sees a 0.17 standard deviation rise in her WAZ score (noisy) and a 0.26 standard deviation rise in her HAZ score (significant at the ten percent level).

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.

6.1 Learning and child health

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

$$Persistence_{cv} = \alpha_{cv} + \beta_1 Health_{cv} + \beta_2 HandwashStock_{cv} + \beta_3 BaselineHealth_{cv} + \delta + \gamma_v + \epsilon_{cv} \tag{8}$$

in which $Persistence_{cv}$ is the average handwashing performance during the month following the withdrawal of incentives or monitoring for child c in village v, *Health* is a health index constructed using Anderson (2008), separately for self-reported disease incidence and anthropometric outcomes[23], *HandwashStock* is the average likelihood of washing during dinnertime over the course of the intervention, *BaselineHealth* is the identical incidence or anthropometric index constructed using baseline health variables, δ is a vector of child and household level characteristics (sex and age of child, whether child was breastfed exclusively, household occupation, number of rooms, mother's age at marriage, and mother's education) and γ is village fixed effects. Standard errors are clustered at the household level. A significant and positive β_1 coefficient implies that, conditional on having accumulated the same amount of consumption stock of handwashing, households that experience larger improvements in health are more likely to persist in their handwashing behavior.[24]

Appendix Table 9a presents the results separately for each treatment arm and health index type. All estimates of the coefficient on the health index are statistically insignificant and close to zero. It does not appear that households are internalizing health gains and updating their handwashing performance accordingly.

Despite the host of controls for child health and household characteristics, it is possible that learning effects are washed out by endogeneity in handwashing behavior to household type: households who experience larger health returns may also be the types of households who handwash little (for example, the sick children who experience the largest health improvements may reside in poor households - who are on average less likely to wash than their affluent counterparts - in a manner that is not sufficiently controlled for in our

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

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

vector of child and household characteristics). Therefore, we also exploit our panel data on illness collected during months three through five of the experiment and consider the following exercise: conditional on households having built the same amount of handwashing stock and experiencing equal levels of sickness, does a household that experiences an illness the week before a handwashing observation behave differently from a household that experiences an illness in the week after the observation? Any difference can plausibly be attributed to the reaction to the health event rather than changes in consumption stock, since the latter is equivalent across comparison households. To evaluate this, we run the following regression for households who report an ARI episode in either the week before or after the week of handwashing observation, run separately for each week of child health panel data[25]:

$$Handwashing_{cv}^{t} = \alpha_{cv} + \beta_1 Sick_{cv}^{t-1} + \beta_2 Sick_{cv}^{t} + \beta_3 SickStock_{cv}^{t-1} + \beta_4 HandwashStock_{cv}^{t-1} + \gamma_v + \epsilon_{cv} \quad (9)$$

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

Appendix Table 9b presents the results. Panel A presents results for households in either the dispenser only or the monitoring arms, and Panel B presents results for households in either the dispenser only or the incentivized arms. These samples thus correspond to those of the persistence analysis in Table 5. Over the course of the weeks in which we can observe before, during, and after ARI incidences, no consistent pattern emerges. Estimates are noisy, with an equal distribution of negative and positive coefficients. It does not appear that households are - at least coherently or consistently - internalizing the health returns of their children and updating their valuation and performance of handwashing accordingly.

Finally and most decisively, note that the rational habit formation effect can only be driven by intertemporal complementarities in the stock of consumption, *not* by learning about child health effects. This

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

is because the experiment exogenously increased only the value of handwashing in the future, not that of the health returns to handwashing in the future.[26] Evidence of rational addiction by households anticipating the monitoring of handwashing behavior therefore offers further evidence that learning about the health returns cannot be the primary driver of intertemporal complementarities in handwashing. Rather, the persistence we observe is driven by the accumulation of consumption stock, or the building of a habit.

7 Conclusion

This study analyzes the process of habit formation in the high-impact preventive health behavior of handwashing with soap, examining how individuals internalize and interact with this habit-forming behavior. Our results suggest that monetary incentives and third party monitoring and feedback are effective means of increasing handwashing rates in the short run. While the impact of incentives on the extensive margin is substantial, intensive margin changes in incentives have diminishing returns. Both extensive margin incentives and monitoring have persistent effects, establishing that handwashing is indeed a habitual behavior. We also present evidence that agents are rationally habit forming: they internalize the intertemporal complementarities in the marginal utility of handwashing. Specifically, households respond strongly in anticipation of a future monitoring intervention, but show no response in anticipation of a future intensive margin change in incentives. This is consistent with the theory of rational addiction, in which agents should only respond in anticipation to interventions that alter the consumption value of future behavior.

This exercise offers the first well identified estimate of the presence of rational habit formation, and additionally for *good* habits, in the literature. These findings inform the optimal incentive design of programs that seek to increase the takeup of good habits; namely, if a behavior is habit-forming, then an intervention may do better to front-load incentives, and thereby maximize habit stock, rather than spread incentives over time. If individuals are rational regarding the habitual nature of the behavior, incentives that are offered at a future date will generate a larger stock of consumption in the long run than those administered immediately. The optimal type, size, and length of such incentives remain important areas of future exploration.

This paper also sheds light on the production function for child health as it relates to the input of hand

[26]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 the 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.

hygiene. We establish the strong link between handwashing with soap and child incidence of respiratory infection and diarrhea and the experiment can uniquely offer treatment on the treated estimates which suggest that a child who achieved a regular dinnertime handwashing practice saw a 74% decrease in the number of days she experienced loose stool motion and a 27% decrease in the number of days she experienced acute respiratory infection. These translate into substantial improvements in anthropometric measures that have long run implications for the health of the child: receiving a handsoap dispenser and liquid soap generates, within eight months, improvements in child mid-arm circumference of 0.08 standard deviations, child weight of 0.15 standard deviations, and child height of 0.23 standard deviations. These findings point to the importance of human-centric design: dispenser provision was not effective because it provided the households with soap; rather, the location, ease of use, and attractiveness of the dispenser and soap must have motivated the practice of handwashing. While this study was not designed to identify these effects, at the fixed cost of $4.00 USD per dispenser and variable cost of $1.00 USD per 15 liters of foaming soap per year (the average household consumption rate), such product-design mechanisms are additionally a fruitful avenue of future research.

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 observation bias. We also collected respondent ratings on handwashing habit formation. Respondents were asked "Has handwashing with soap before eating become habitual for you?" and were rated on a five point scale using the following metric: 0 = "How? You did not give us soap"; 1 = "No, not at all"; 2 = "No, not yet, but it is growing"; 3 = "Yes, mostly, but still needs time"; 4 = "Yes, definitely, the habit has been established." Third, surveyors asked the respondent whether they had any liquid soap in the household; for treated households, the question specified that we were interested in non-project liquid soap. If households mentally assign barsoap to purposes like bathing and laundry, the presence of liquid soap may be a signal that handwashing is a household priority. These three hygiene measures were collected at midline, seven to eight months after rollout. Finally, we proxy for the amount of soap consumed by a household using the total number of dispenser presses per day.

Results are presented in Appendix Table 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.[27] 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

[27] 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 by changes in other sanitation practices, if for example the act of having handwashing top of mind makes remembering to maintain other preventive health practices easier. It is also important to examine effects of the interventions on other sanitation outcomes as they affect our interpretation of the results on child health: improvements in sanitation may be the real cause of improvements in child health and handwashing merely a correlate. Appendix Table 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 the interventions 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.[28] At each level, the respondent was asked whether she would prefer to take the prize or take a month's worth of soap.[29] To ensure incentive compatibility, each choice was made in the form of a token and dropped into a bag; after the completion of all choices, the respondent chose one token at random and received the drawn prize.

Results are presented in Appendix Table 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 only arms are statistically indistinguishable from those of pure control. One interpretation of this result is that the prizes from the incentives intervention gave the mothers (and/or children) a taste for such rewards which crowded out, rather than complementing, the value of soap. Households may have anchored their valuation of soap to a negative price as they became accustomed to being paid to use it.

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

[28]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.

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

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. Appendix Figure 1 plots the average WTP across each treatment arm for this restricted sample.

Echoing the results on child health and the absence of learning, this valuation exercise underscores a problem at the heart of behavioral change in preventive health: health benefits of preventive behaviors are often too small, too delayed, or too difficult to observe relative to what is required for households to internalize the causal relationship between behavior and health. Even in a setting where behavioral change generates health effect sizes that are twenty percent at the lower bound, 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.[30] Importantly, the same argument applies to habit formation: despite the considerable handwashing stock accumulated over eight months and evidence of persistence in handwashing, households do not increase their willingness to pay for soap. At the point of playing the willingness-to-pay game, neither the return from habit nor the return to health was sufficiently internalized (or sufficiently high) to shift households' monetary valuations of soap.[31]

8.3 Behavioral spillovers

Despite no obvious changes imposed on dispenser-only households throughout the experiment, these households demonstrate a rise and fall in handwashing rates that closely mimics the pattern of monitored households (Appendix Figure 2). This pattern could be due to parallel time trends, the dispenser control households undergoing their own process of habit formation, or to spillovers in behavior from neighboring monitored households.

Because treatment assignment between dispenser only and monitoring households was randomized at the household level, we capitalize on the random variation in the concentration of monitoring households nearby dispenser only households to estimate the size of spillovers in handwashing behavior.[32] We choose a radius of one kilometer around each dispenser only household, as this is a typical distance within which

[30] 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.

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

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

children play with one another and attend the same government nursery school, mothers walk to the local pond or road-side shop, and most conversations are likely to occur. We examine spillovers at three points in time: Day -40 to -30, when there is little that dispenser households can learn from monitoring households; Day 40 to 50, ten days after monitoring households have received their first calendar (which gives them time to share their experiences with neighbors), and Day 120 to 130, after monitoring is officially over. If spillovers drive the rise in rates among dispenser only households, we should only observe the effects of spillovers in the middle specification, and potentially remnants in the third specification.[33] 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 one kilometer of a dispenser household is associated with a 1.6 percentage point (5.7%) increase in dispenser household handwashing rates. Thus while spillovers from monitored neighbors may have played some role in the handwashing behavior of dispenser households, they can only explain a fraction of the observed rise (nearly a doubling) in handwashing among dispenser households in the first three months of the experiment.

The pattern we observe may alternatively be due to parallel time trends or the natural process of habit formation. While we cannot rule out the former, habit formation is not unlikely. Consider a habit formation model in which there exists some fixed amount of consumption stock which must be accumulated before σ kicks in. This permutation of the model is consistent with the initial shallow decay of handwashing rates in dispenser control households (Appendix Figure 2, Day -70 to 0) followed by their steady rise (Day 0 to 90). Given that surveyors switched from twice monthly visits to collect health data to once monthly visits to collect data (across all sample households) around Day 110, which can be regarded as a positive shock to x_t, the subsequent decay in handwashing rates is likewise consistent with the habit formation model. Therefore the pattern of a secular rise in handwashing rates amongst dispenser households suggests the role of habit formation in handwashing over time even absent monitoring or incentive interventions.

8.4 Health spillovers

Despite the lack of significant behavioral spillovers, we may expect to see spillovers in health given that viral and bacterial contamination are the primary sources of diarrhea and ARI morbidity. To measure these

[33] 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).

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, unadjusted for multiple hypothesis testing). Therefore despite substantial positive health internalities, the habit of handwashing at dinnertime produces modest health externalities for neighboring children. This is not especially surprising given the timing of the behavioral change we focus on: while children are most prone to spreading germs during the daytime at school and as they play, our intervention improves hand hygiene only at night. To maximize positive spillovers, we may want to focus on hand hygiene interventions linked to schools or a child's midday meal. This is an important direction for future research.

Figure 1: Soap dispenser anatomy

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

Figure 2: Typical dispenser location

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

Figure 3: Child using dispenser

Notes: A child uses the dispenser by pushing the black button once or twice. The foaming soap can be rubbed on the hands within minimal water. He then goes to the nearby water pail or tubewell in the courtyard and rinses the soap off with the help of the mother, who pours the water.

Figure 4: Randomization map

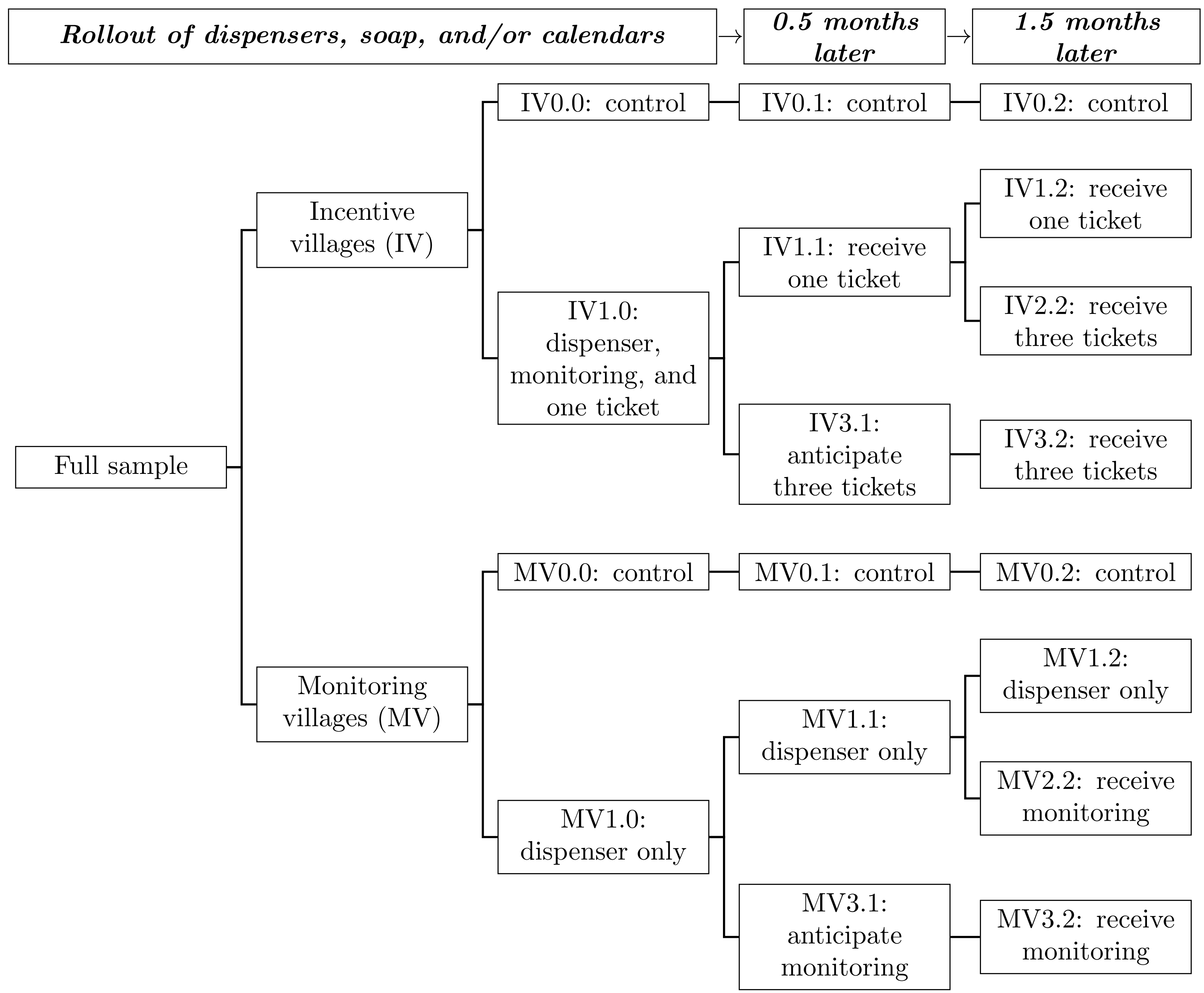
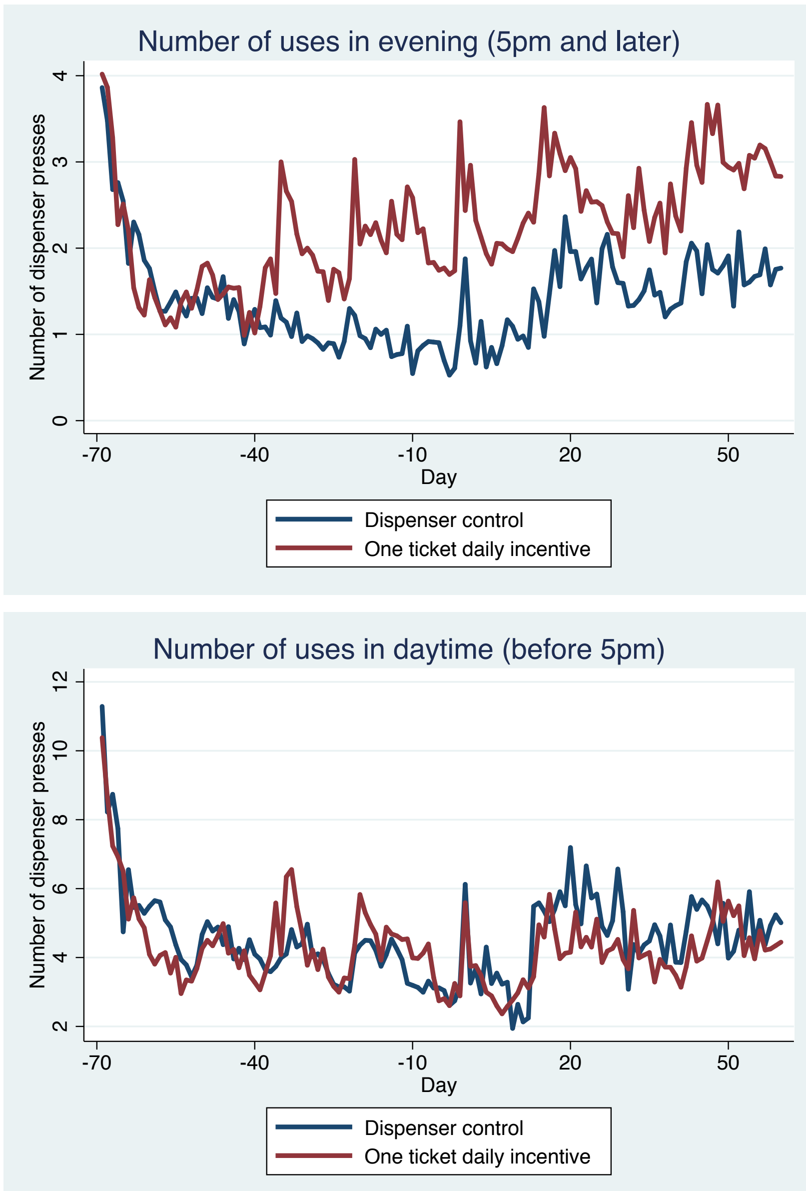
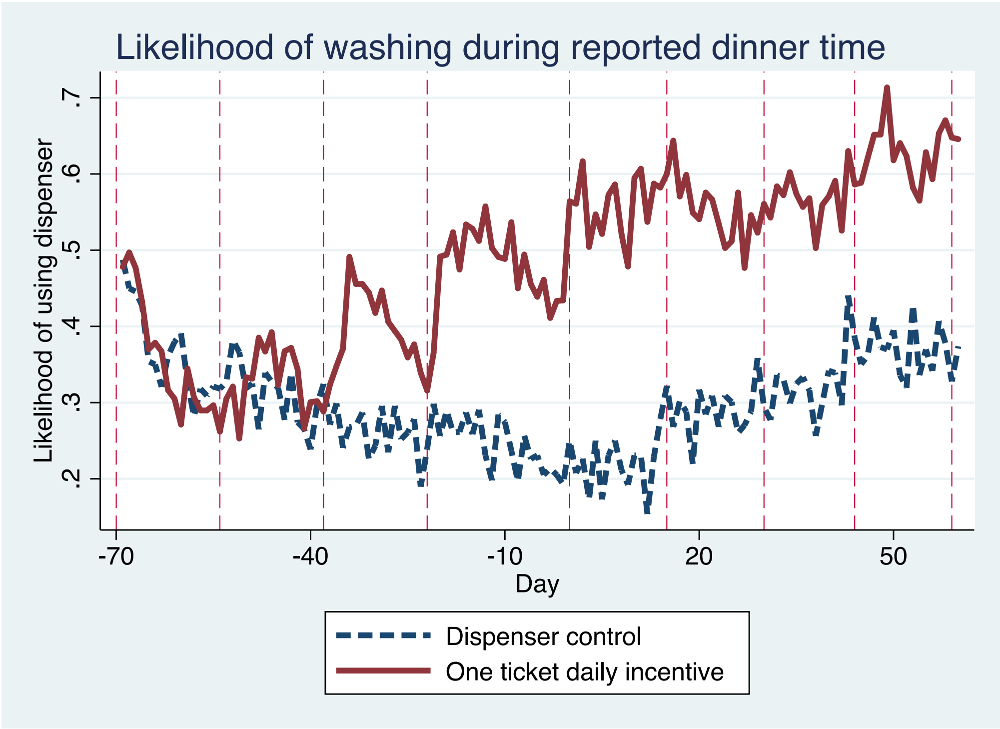


Figure 5a: Dispenser use over 24 hours



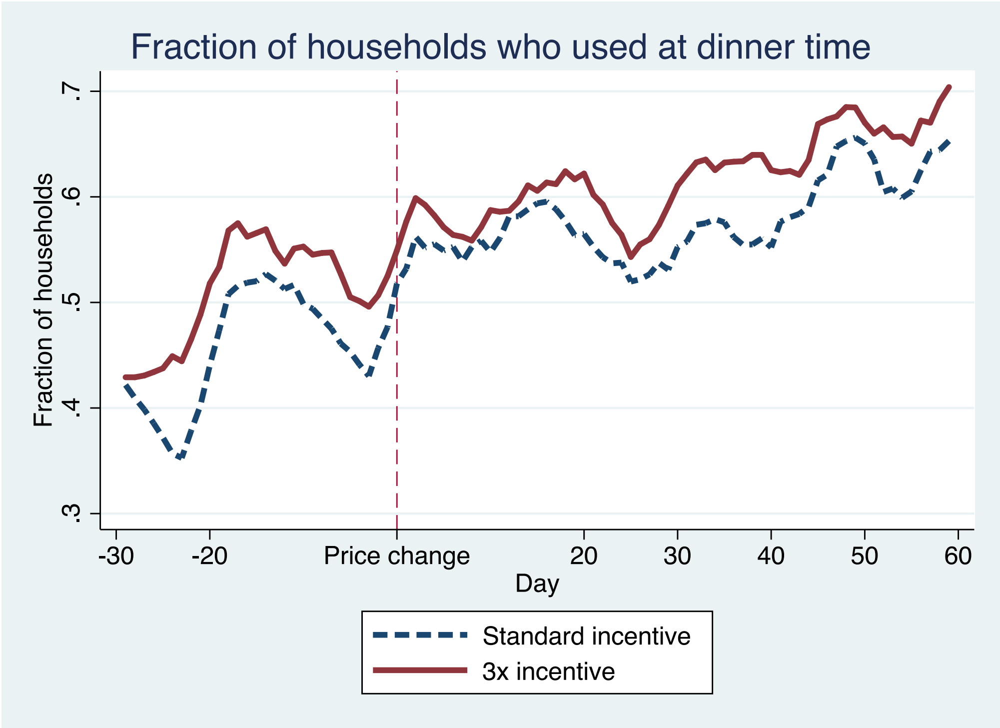
Notes: Figures show the average number of individual presses per day after 5pm and before 5pm, respectively. Blue line represents households who received only the dispenser; red line represents households who received the dispenser, feedback, and one ticket for every night the dispenser was active around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

Figure 5b: Binary use at dinnertime



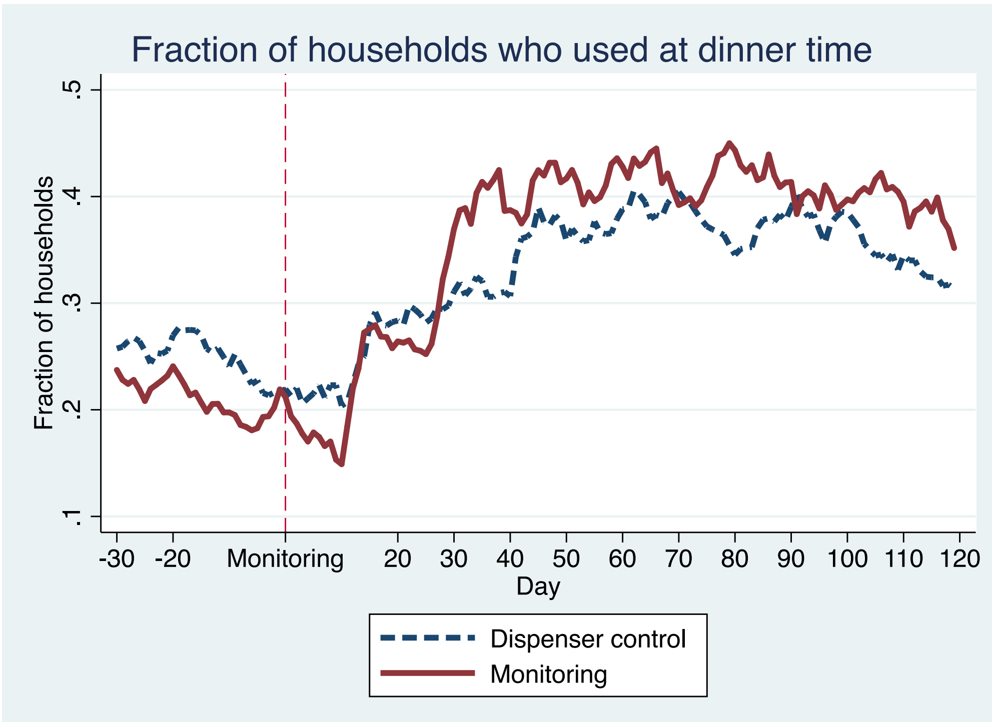
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Blue dashed line represents households who received only the dispenser; red line represents households who received the dispenser, feedback, and one ticket for every night the dispenser was active around their self-reported dinnertime. Vertical red lines represent the approximate surveyor visit day. Day -70 is the day of rollout. Tickets were distributed for the full length of the graph shown (until Day 60).

Figure 6a: Incentive effect during intervention regime



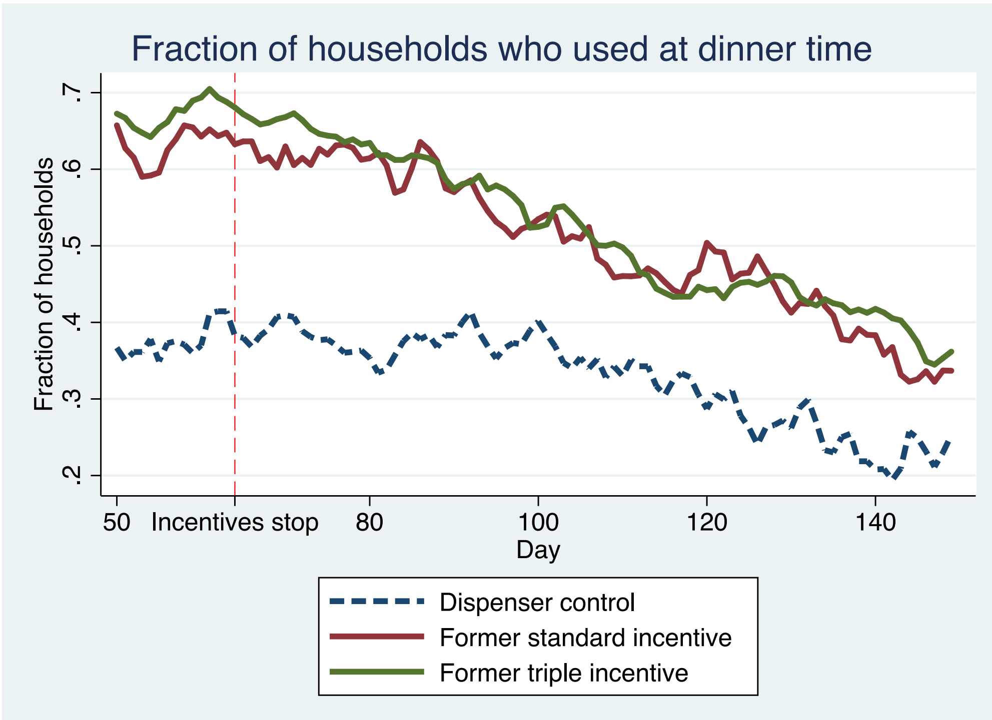
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Blue dashed line represents households who received the dispenser, feedback, and one ticket for every night the dispenser was active around their self-reported dinnertime; red line represents households who received one ticket until the point of the "Price change" (Day 0) and received three tickets for every night the dispenser was active during dinnertime for the remainder of the days displayed in the figure.

Figure 6b: Monitoring effect during intervention regime



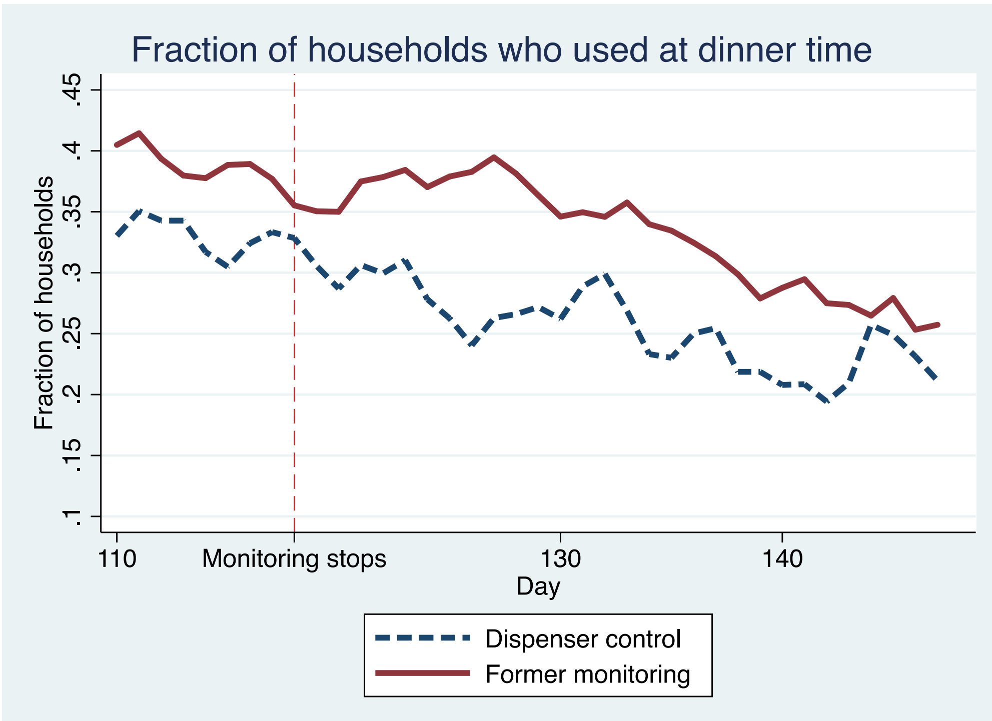
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Blue dashed line represents households who received the dispenser only; red line represents households who received the dispenser only until the point of the "Monitoring" (Day 0) and received feedback/monitoring on behavior thereafter for the duration displayed in the figure.

Figure 7a: Persistence of incentive effect



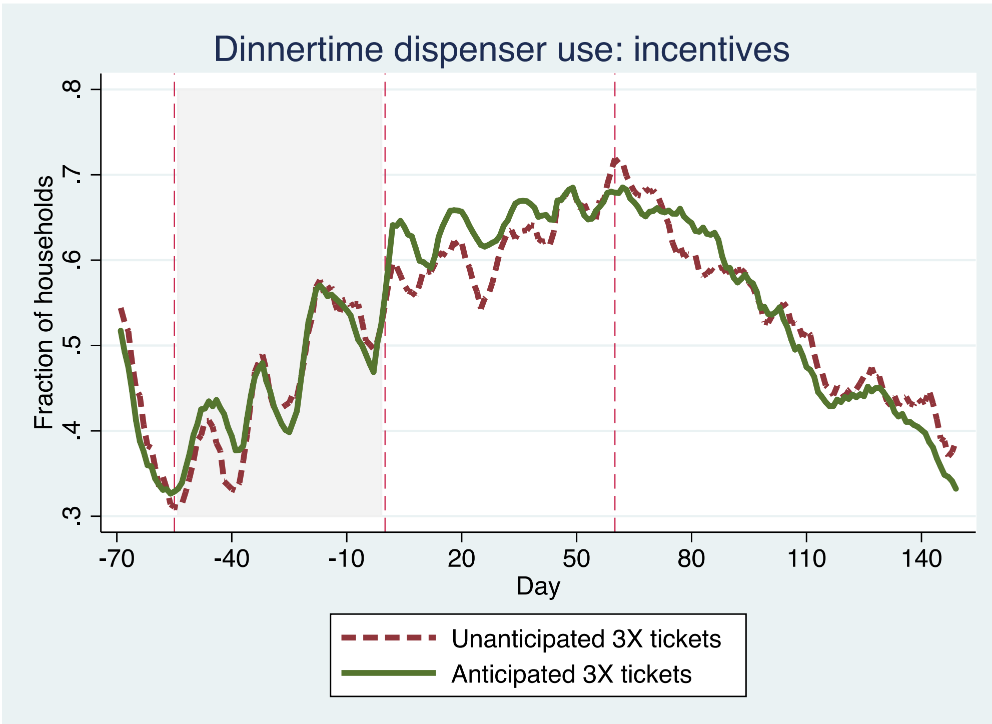
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Blue dashed line represents households who received the dispenser only; red line represents households who received the dispenser, feedback,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

Figure 7b: Persistence of monitoring effect



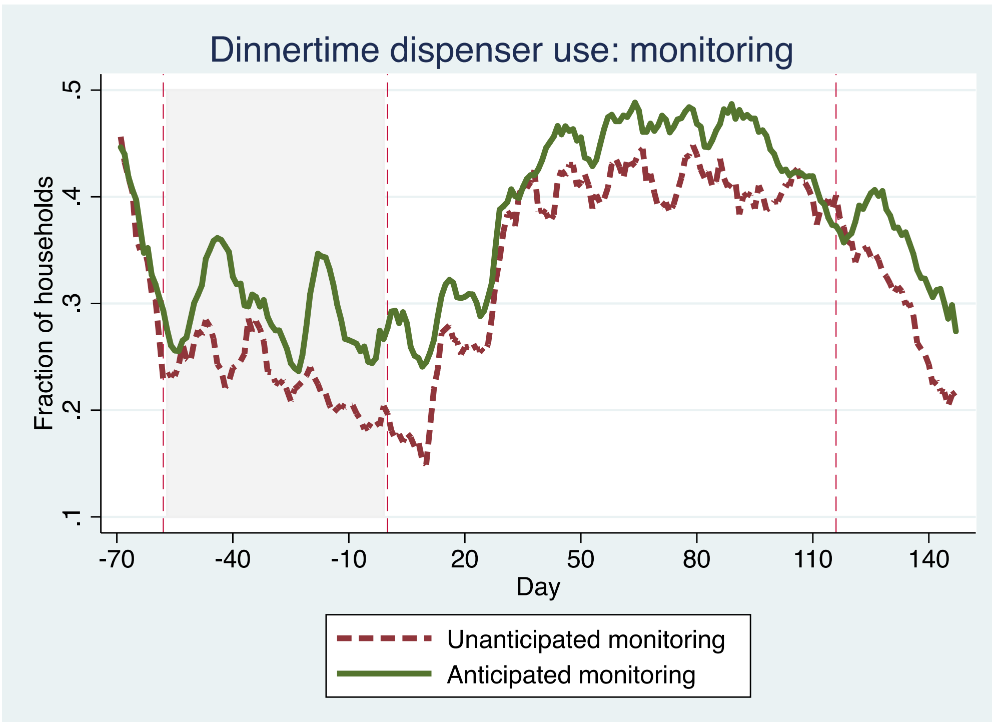
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Blue dashed line represents households who received the dispenser only; red line represents households who received the dispenser and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

Figure 8a: Rational addiction in incentives



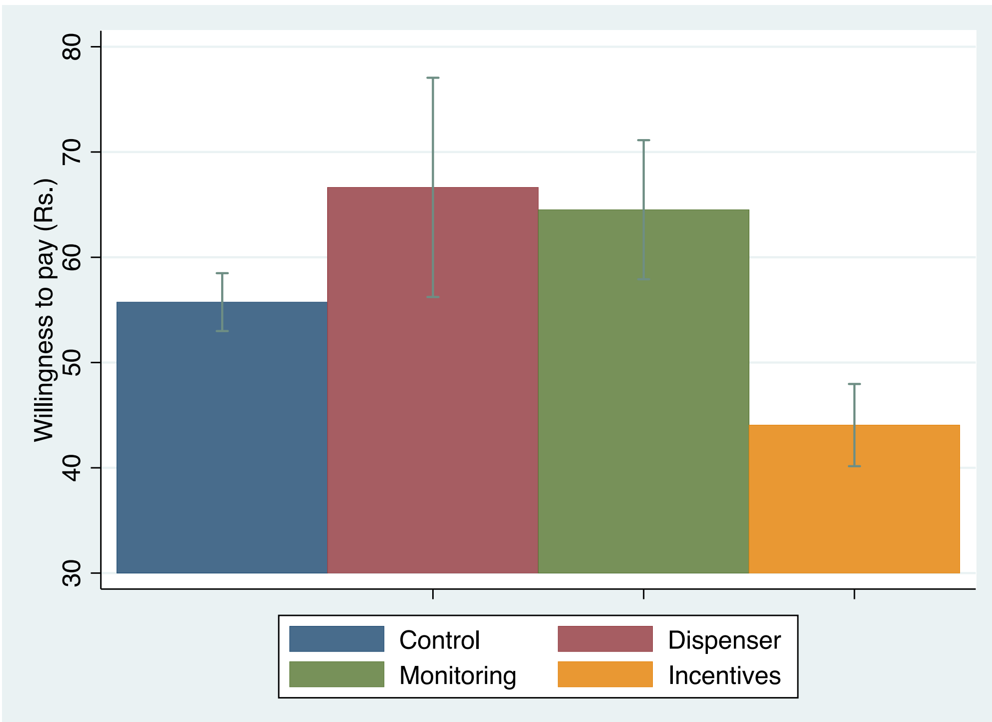
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Both red and green lines represent households who received the dispenser, feedback, and one ticket until Day 0, after which they received three tickets per day the dispenser was active during the evening mealtime; however, green households were anticipating the tripling of the tickets while red households were not. The gray box represents the time during which green households were anticipating. Triple tickets then commenced on Day 0 and lasted until Day 60 (third vertical red line).

Figure 8b: Rational addiction in monitoring



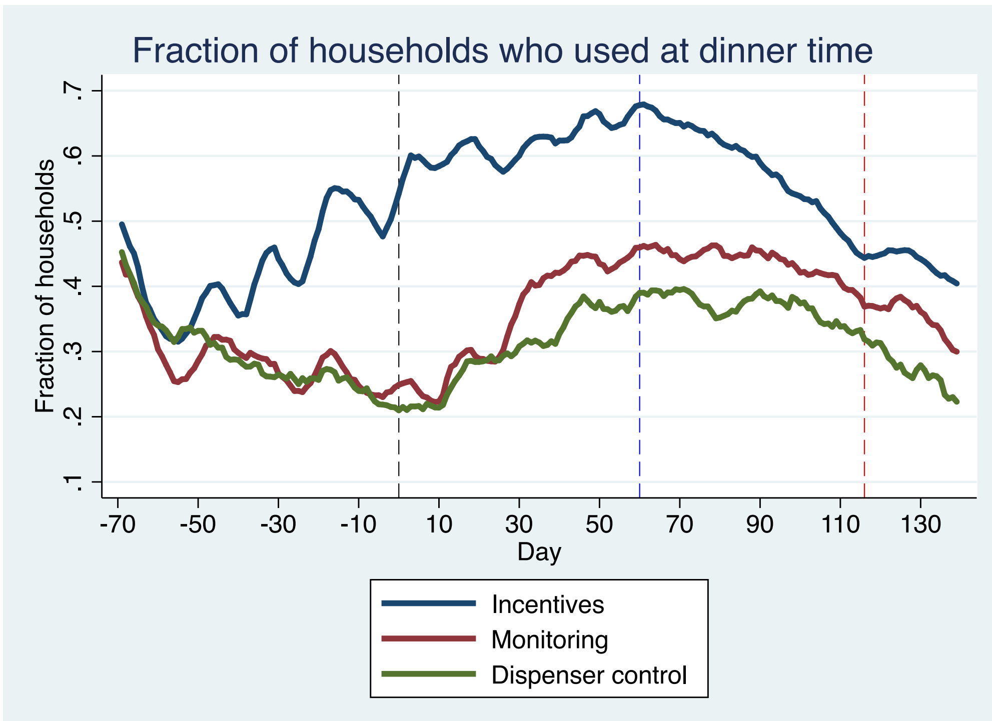
Notes: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Both red and green lines represent households who received the dispenser only until Day 0, after which they additionally received feedback/monitoring; however, green households were anticipating the start of monitoring/feedback while red households were not. The gray box represents the time during which green households were anticipating. Feedback then commenced on Day 0 and lasted until Day 117 (third vertical red line).

Appendix Figure 1: Willingness to pay for soap



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

Appendix Figure 2: Time trends across treatment arms



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

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

Notes: "Treated" pools all households that received a dispenser. "Pure control" are households who did not receive a dispenser. t-statistics computed in a regression of the variable on treatment assigment with village level fixed effects.

Table 3. Impact of incentives on the extensive margin

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

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

	(1)	(2)	(3)	(4)
Day	0-59	30-59	0-59	30-59
	Likelihood of using during reported dinnertime			
Contemporaneous tripled incentive	0.0268	0.0503**		
	(0.0227)	(0.0261)		
Contemporaneous monitoring			0.0711***	0.0842***
			(0.0224)	(0.0254)
Mean of comparison group	0.579	0.599	0.342	0.361
	[0.0212]	[0.0217]	[0.0213]	[0.0219]
Observations	18,487	9,905	24,063	19,098

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. All regressions control for average dinnertime handwashing rates prior to price boost or commencement of the service, which occured on Day 0. Control group for columns 1-2 is the standard (1 ticket) incentive treatment arm. Control group for columns 3-4 is the dispenser only arm. 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 single ticket 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 comparison 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. Comparison group for "Former single ticket incentive" and "Former monitoring" is the dispenser only group; comparison group for "Former tripled incentive" is the former single ticket 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 are currently collecting 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 7a. Rational habit formation

	(1)	(2)	(3)	(4)
	Likelihood of using during reported dinnertime			
	Prior to intervention			
	Day -54 to -1	Day -21 to -1	Day -54 to -1	Day -21 to -1
Anticipated triple incentive	-0.00593	-0.0381		
	[0.0235]	[0.0305]		
Anticipated monitoring			0.052*	0.08**
			[0.0243]	[0.0284]
Mean of comparison group	0.454	0.536	0.231	0.205
	[0.0197]	[0.0257]	[0.0191]	[0.0218]
Observations	23,273	9,176	16,268	6,297

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. Comparison group for "Anticipated triple incentive" is the group that was surprised with the triple incentive on Day 0; comparison 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 7b. Persistence in rational habit stock

Likelihood of using during reported dinnertime:	During intervention				After withdrawal of intervention	
	(1)	(2)	(3)	(4)	(5)	(6)
	Day 0 to 59	Day 30 to 59	Day 0 to 116	Day 30 to 116	Day 60 to 90	Day 117 to 147
Anticipated triple incentive	0.00999	-8.78e-05			0.011	
	[0.0218]	[0.0223]			[0.0256]	
Anticipated monitoring			0.0306	0.0261		0.0175
			[0.0240]	[0.0257]		[0.0294]
Mean of comparison group	0.609	0.616	0.375	0.409	0.446	0.31
	[0.0192]	[0.0193]	[0.0226]	[0.0249]	[0.0255]	[0.0279]
Observations	48,886	37,270	33,401	26,243	13,239	7,720

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. Comparison group for "Anticipated triple incentive" is the group that was surprised with the triple incentive on Day 0; comparison 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)	(3)	(4)
	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 only		-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. Biweekly child health data spans February and March of 2016 (4-5 months after rollout). All treatment effects are estimated relative to the pure control group. Robust standard errors in brackets and are clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1.

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

	(1)	(2)	(3)	(4)
	Whether child had any loose stool in last two weeks		Total days of loose stool in last two weeks	
Treated household	-0.0230**	-0.0315***	-0.0575**	-0.0817***
	[0.00849]	[0.00975]	[0.0208]	[0.0236]
Mean of pure control	0.100	0.100	0.209	0.209
	[0.00572]	[0.00572]	[0.0151]	[0.0151]
With baseline controls		X		X
Observations	4,940	3,820	4,955	3,830

	(5)	(6)	(7)	(8)
	Whether child showed any ARI symptoms in last two weeks		Total days of ARI in last two weeks	
Treated household	-0.0281**	-0.0393**	-0.163**	-0.204**
	[0.0138]	[0.0154]	[0.0770]	[0.0884]
Mean of pure control	0.270	0.270	1.247	1.247
	[0.00886]	[0.00886]	[0.0504]	[0.0504]
With baseline controls		X		X
Observations	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 seven to eight months after rollout. "Treated household" is any household that received a dispenser (the pooled sample of incentive, monitoring, and dispenser only 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. 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)
	Weight-for-age z-score		Height for age z-score		Mid-arm circ. for age z-score	
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. Height-for-age z-score (HAZ), weight-for-age z-score (WAZ), and midarm circumference-for-age z-score (MAZ) are calculated using WHO anthropometric methodology. Sample is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Data was collected seven to eight months after rollout. "Treated household" is any household that received a dispenser (the pooled sample of incentive, monitoring, and dispenser only 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)
	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	Total days of ARI in last two weeks	Weight for age z-score	Height for age z-score	Mid-arm circ. for age z-score
Regularly washes (self report)	-0.0591***	-0.154***	-0.0652**	-0.335*	0.172	0.262*	0.0991
	[0.0194]	[0.0460]	[0.0295]	[0.172]	[0.109]	[0.148]	[0.0855]
Hands observed clean	-0.484*	-1.451**	-0.519	-2.884	0.650	0.675	0.287
	[0.263]	[0.694]	[0.347]	[2.052]	[0.904]	[1.336]	[0.630]
Mean of pure control	0.100	0.209	0.270	1.247	-2.167	-1.866	-1.365
	[0.00572]	[0.0151]	[0.00886]	[0.0504]	[0.0458]	[0.0665]	[0.0432]
Observations	3,814	3,824	3,824	3,824	861	860	856

Notes: Outcome data was collected seven to eight months after rollout. Z-scores in columns 5-7 are calculated using WHO anthropometric methodology. Sample in columns 5-7 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 1-4 include children younger than fourteen. Regression shows the treatment on the treated estimates where "treated" is either (1) a household who reports washing hands regularly during dinnertime, or (2) a household whose respondent has clean hands as judged by the enumerator, both of which are 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. 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 health after eight months disaggregated by treatment arm

	(1)	(2)	(3)	(4)
	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	Total days of ARI in last two weeks
Incentives	-0.0267**	-0.0726**	-0.0225	-0.138
	[0.0125]	[0.0297]	[0.0195]	[0.108]
Monitoring	-0.0333**	-0.0971**	-0.0572**	-0.287*
	[0.0169]	[0.0407]	[0.0276]	[0.169]
Dispenser only	-0.0586***	-0.100*	-0.102***	-0.421*
	[0.0219]	[0.0567]	[0.0363]	[0.228]
Mean of pure control	0.100	0.209	0.270	1.247
	[0.00572]	[0.0151]	[0.00886]	[0.0504]
Observations	3,820	3,830	3,830	3,830

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. 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 3a. Child anthropometric outcomes after eight months disaggregated by treatment arm

	(1)	(2)	(3)
	Weight for age z-score	Height for age z-score	Midarm circ. for age z-score
Incentives	0.114	0.192*	0.00801
	[0.0788]	[0.102]	[0.0599]
Monitoring	0.181	0.295	0.172*
	[0.122]	[0.181]	[0.103]
Dispenser only	0.143	0.269	0.250*
	[0.180]	[0.325]	[0.137]
Mean of pure control	-2.167	-1.866	-1.365
	[0.0459]	[0.0666]	[0.0432]
Observations	863	862	858

Notes: Observations are at the child level. Dependent variables calculated using WHO anthropometric methodology. Sample is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Data was collected eight months after rollout. Baseline controls 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. "Monitoring" is the pooled sample of all households in the surprised monitoring arm and anticipated monitoring 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 3b. Child anthropometric outcomes after eight months disaggregated by age

Age in months at rollout	1-12	13-24	25-36	37-48	49-60
Panel A			**Weight for age z-score**		
Received dispenser	-0.0966	0.114	0.283*	-0.135	0.236
	(0.254)	(0.252)	(0.157)	(0.123)	(0.251)
Constant	-2.070	-0.790	-1.785	-1.809	-1.108
	(0)	(0.330)	(1.719)	(0.133)	(1.559)
Observations	104	178	206	271	103
Panel B			**Height for age z-score**		
Received dispenser	-0.0869	0.195	0.304*	0.0384	-0.00498
	(0.469)	(0.403)	(0.179)	(0.154)	(0.397)
Constant	-1.280	-0.200	-0.397	-1.542	0.621
	(2.26e-07)	(0.779)	(1.299)	(0.151)	(1.837)
Observations	104	177	207	270	103
Panel C			**Mid-arm circumference for age z-score**		
Received dispenser	0.212	0.452**	0.0567	-0.0336	0.240
	(0.284)	(0.212)	(0.120)	(0.107)	(0.208)
Constant	-1.340	-0.689	-1.163	-0.847	-0.764
	(2.66e-08)	(0.275)	(1.000)	(0.131)	(0.661)
Observations	103	178	207	270	99

Notes: Observations are at the child level. Dependent variables calculated using WHO anthropometric methodology. Sample is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Data was collected eight months after rollout. Baseline controls 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. "Received dispenser" is any household that received a dispenser, pooled over treatment arms. Unadjusted p-values presented. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 4. Alternative hygiene measures

	(1)	(2)	(3)	(4)	(5)
	Observed hand cleanliness	Observed nail cleanliness	Whether handwashing habit was achieved	Whether household has non-project liquid soap	Total dispenser presses in 24 hours
Received dispenser	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 only	0.0410	0.0974	1.318***	0.0165	--
	[0.0495]	[0.0625]	[0.0979]	[0.0257]	--
Mean of comparison group	1.552	1.179	1.615	0.0548	5.354
	[0.0167]	[0.0206]	[0.0306]	[0.00565]	[0.188]
Observations	2,672	2,671	2,669	2,670	951

Notes: Observations are at the household level in Columns 1-4 and at the child-day level in Column 5. "Received dispenser" is the pooled sample of incentive, monitoring, and dispenser control households. Coefficients are reported from two separate regressions: the first pools all dispenser households ("Received dispenser" row); the second includes covariates for each treatment arm (Incentives, Monitoring, and Dispenser only). All regressions include village level fixed effects except Column 5, which compares treatment arms across villages. Column 5 has a restricted sample because the outcome variable is only observed amongst households who received a dispenser. Therefore for this column, the relevant comparison group is the dispenser only group. In all other regressions, the relevant comparison 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

	(1)	(2)	(3)	(4)
	Whether household defecates in open		Whether household treats drinking water	
Received dispenser	-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 only		-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. "Received dispenser" is the pooled sample of incentive, monitoring, and dispenser control households. All regressions include village level fixed effects. 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)
	Willingness to pay (Rs.)		
Received dispenser	-4.738**		
	[1.935]		
Incentive		-9.060***	-7.755***
		[2.303]	[2.393]
Monitoring		1.415	-0.681
		[3.705]	[3.706]
Dispenser only		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. "Received dispenser" is the pooled sample of incentive, monitoring, and dispenser control households. 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)
	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]
Mean of comparison group	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. All regressions include village and day level fixed effects. Robust standard errors in brackets and clustered at the household level. Comparison group is dispenser only households who have zero monitored households within a one kilometer radius. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 8. Health spillovers

	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	Total days of ARI in last two weeks
Panel A: Monitoring villages				
No. of dispensered households	-0.00220	-0.00576	-0.00234	-0.0286
	[0.00203]	[0.00487]	[0.00377]	[0.0215]
Mean of comparison group	0.108	0.216	0.275	1.452
	[0.0162]	[0.0385]	[0.0270]	[0.175]
Observations	624	629	629	629
Panel B: Incentive villages				
No. of dispensered households	-0.000767	-0.000983	-0.00672*	-0.0337*
	[0.00249]	[0.00575]	[0.00359]	[0.0183]
Mean of comparison group	0.0867	0.181	0.277	1.342
	[0.00955]	[0.0248]	[0.0147]	[0.0873]
Observations	1,601	1,602	1,602	1,602

Notes: Observations at child level. Sample is composed of the children in pure control households in each type of village (monitoring village or incentive village). Independent variable is the number of households who received a dispenser (monitoring and dispenser only households for monitoring villages; incentivized households for incentive villages) within 1 km of the pure control household. Comparison group is made up of pure control households who have no dispenser receiving households within a one km radius of itself. Robust standard errors in brackets and clustered at the household level. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 9a: Learning about health (midline data)

	(1)	(2)	(3)	(4)	(5)	(6)
	Average likelihood of handwashing at dinnertime one month after withdrawal of interventions					
	Dispenser only		Monitoring		Incentives	
Health index type:	Incidence	Anthro	Incidence	Anthro	Incidence	Anthro
Health index	-0.00205	-0.0407	-0.00183	0.0277	0.00206	-0.00771
	[0.0261]	[0.0624]	[0.0108]	[0.0293]	[0.00825]	[0.0260]
Constant	0.183	-0.212	-0.0485	-0.279	0.242***	0.0539
	[0.239]	[0.626]	[0.0894]	[0.226]	[0.0782]	[0.215]
Observations	154	32	408	100	889	201

Notes: Observations at the child level; standard errors clustered at the household level. All regressions include village level fixed effects and controls for the average likelihood of washing during dinnertime during the course of the intervention, baseline health index, child sex, child age, number of months the child was breastfed, household occupation, number of rooms, mother's age at marriage, and mother's education. Health index is constructed using Anderson (2008); the "Incidence" index is constructed as a weighted average of the child being free of loose stool or ARI in the two weeks prior to surveying and the number of days she was free of these illnesses; the "Anthro" index is constructed using child height, weight, and mid-arm circumference z-scores. Therefore a higher health index implies better health. The dependent variable is the average likelihood of the dispenser being active during dinnertime over the course of the one month after the withdrawal of monitoring or incentives (the time frame for monitoring is also applied to the dispenser only group). Columns 1, 3, and 5 include all children below the age of 14 years; columsn 2, 4, and 6 include only children 60 months and below.

Appendix Table 9b: Learning about health (panel data)

	(1)	(2)	(3)	(4)	(5)	(6)
	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Panel A: Dispenser only and monitored households						
Sick in previous week	-0.148	0.217	-0.281	0.0617	0.764	-0.0349
	[0.217]	[0.300]	[0.397]	[0.361]	[0.476]	[1.077]
Observations	358	341	337	305	236	259
Panel B: Dispenser only and incentivized households						
Sick in previous week	-0.278	-0.384	0.499	0.870**	-0.100	0.371
	[0.332]	[0.446]	[0.403]	[0.374]	[0.389]	[1.048]
Observations	578	562	575	497	455	427

Notes: Observations are at the child level and sample is restricted in each specification to those children who experienced a sickness either in the week prior to handwashing observation or the week after handwashing observation (but not both). Standard errors clustered at the household level. All regressions include village level fixed effects and controls for whether or not the child experienced ARI in the week that the handwashing outcome is observed, the total number of ARI incidences up to the week before observation, and the total number of days the dispenser was used up to the week before observation. The dependent variable is the total number of days the dispenser was active during dinnertime during the week of observation.

9 Bibliography

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

Allcott, 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.

Aunger, R., Schmidta, W.P., Ranpurab, A., Coombesc, Y., Mainad, P.M., Matikod, C.N., Curtis, V. (2010) Three kinds of psychological determinants for hand-washing behavior in Kenya. Social Science Medicine. 70(3): 383-391.

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.

Charness, G. Gneezy, U. (2009). Incentives to exercise. Econometrica. 77(3): 909-931.

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).

Conley, T. Udry, C. (2010). Learning about a New Technology: Pineapple in Ghana. American Economic Review. 100(1): 35-69.

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.* New York: Random House Trade Paperbacks.

Dupas, P. (2010). Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment. NBER Working Paper 16298.

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. International Journal of Epidemiology. 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. 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.

Hoddinott J., Maluccio J., Behrman J., R Martorell, Melgar P., Quisumbing A., Ramirez-Zea M., Stein A. Yount K. (2013). Adult consequences of growth failure in early childhood. American Journal of Clinical Nutrition. 98: 1170-1178.

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

James, W. (1914). *Habit*. New York: Henry Holt and Co.

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

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

Marshall, A. (1920). *Principles of Economics*. London: Macmillan and Co. 8th ed.

McKay, S., Gaudier, E., Campbell, D., Prentice, A., 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.

Neal, D., Vujcic, J., Hernandez, O., and Wood, W. (2015). The Science of Habit: Creating Disruptive and Sticky Behavior Change in Handwashing Behavior. Washington D.C., USA. USAID/WASHplus Project.

O'Donoghue, T., M. Rabin, (2001). Addiction and Present-Biased Preferences. Working Paper.

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

Royer, H., Stehr, M., Sydnor, J. (2015). Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. American Economic Journal: Applied Economics. 7(3): 51-84.

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.

Wood, W. Neal, D. T. (2007). A new look at habits and the habit-goal interface. Psychological review. 114(4): 843.