Experiential and Social Learning in Health*

Agha Ali Akram, Gabriella Fleischman, Reshmaan Hussam, and Akib Khan †

January 31, 2025

Abstract

We study how experiential and social learning interact in health technology adoption, focusing on water chlorination among 1,800 caregivers of children in Pakistan. Our experiment has four arms: control households; households receiving free chlorine tablets; households receiving tablets plus financial incentives for chlorine use; and households receiving tablets plus an experiential learning intervention where community health workers help caregivers visually track child diarrhea rates relative to control households – before and after chlorine distribution (Info-Tool). Although incentives increase short-run chlorination, their effects fade quickly. By contrast, experiential learning effects persist—but *only* when these households also have a neighbor in the learning arm. Welfare implications are significant: ITT (TOT) estimates indicate that learning households with learning neighbors see a 0.12 SD (0.51 SD) increase in child anthropometrics. Households not in the learning arm show no behavioral change even when near a learning neighbor, revealing that experiential and social learning not only complement, but *rely*, on each other. Learning participants do not hold stronger beliefs in chlorine's efficacy; rather, they report greater motivation to use chlorine for their families' health and place more trust in other learning participants' child-health knowledge. These results suggest a departure from standard Bayesian models, where a signal, independent of its effects on belief, can motivate action if it comes from sources caregivers trust through firsthand experience. The Info-Tool fosters such trust, leading individuals to prioritize signals recorded through it, and to value signals from others who also use it. From this complementarity between experiential and social learning emerges an insight into experimental design: standard cluster-randomized trials risk overestimating individual treatment effects by conflating within-treatment spillovers with purely individual effects. Comparing our results with a cluster-randomized trial of the experiential learning arm shows that interpreting those findings as individual treatment effects would yield a six-fold overestimation. Thus, when experiential and social learning are complements, achieving individual-level efficacy requires sufficient treatment density, with cost per DALY averted decreasing in neighborhood saturation.

*We are thankful for useful comments and suggestions from seminar participants at Y-RISE 2024, Stockholm School of Economics, NEUDC 2024, and AFE 2023. Fatimeh Munawar, Syed Ali Rehan, and Ibadullah Channa provided invaluable research assistance. We gratefully acknowledge financial support from the United States Agency for International Development (USAID) and the World Bank. This project would not have been possible without Interactive Research and Development (IRD) Pakistan. The contents are the authors' sole responsibility and do not necessarily reflect the views of USAID or the United States Government.

†Akram: Mathematica, Fleischman: Harvard Kennedy School, gfleischman@g.harvard.edu, Hussam: Harvard Business School, Khan: MISUM, Stockholm School of Economics, akib.khan@hhs.se.

1 Introduction

Randomized controlled trials (RCTs) are a cornerstone of empirical research. Their validity often rests on the assumption that individual treatment effects are independent of other individuals' treatment status, commonly referred to as the Stable Unit Treatment Value Assumption (SUTVA). Economists have made substantial progress in analyzing cross-treatment spillovers, where treated units influence control units, to understand general equilibrium effects. However, much less attention has been given to *within*-treatment spillovers, where one treated unit affects *another* treated unit, despite their relevance for economic theory and policy.

Why might an intervention generate a complementarity between receiving treatment and being exposed to other people who receive treatment? Treatments that are dose-responsive and that can be dispensed through social channels are prime candidates for within-treatment spillovers.[1] A rapidly growing body of literature identifies social learning, which meets these criteria under a model of complex contagion, as an important driver of technology adoption and behavioral change (Arrow, 1962; Conley and Udry, 2010; Banerjee et al., 2024; Beaman et al., 2021; Bikhchandani et al., 2024; Breza and Chandrasekhar, 2019; de Janvry et al., 2016; Foster and Rosenzweig, 1994; Khandelwal, 2024; Kondylis et al., 2023).[2] Furthermore, if social learning is predicated on personal experience – for example, if social learning can effectively confirm beliefs established from personal observation, but is ineffective at conveying novel information – then social learning from an information intervention will *only* arise as a within-treatment spillover. But the questions that this literature raises are often geared toward understanding how to disseminate information broadly and cost-effectively to untreated units, which, by the nature of the questions themselves, only identifies cross-treatment spillovers. We underscore that social learning may play an important role in augmenting learning from oneself, but the treatment designs that we most often use to study social learning do not facilitate this learning process.

We test for complementarity between individual experiential learning and social learning in the adoption of health technology among caregivers of children in Pakistan. Our setting allows us to assess the potential of within-treatment spillovers to confound how we interpret estimates from cluster-randomized controlled trials. We draw from an initial experiment (Akram and Mendelsohn, 2021) which finds large treatment effects of a fully-saturated cluster-randomized intervention that encourages health technology adoption through experiential learning. We conduct an *individually*-randomized controlled trial testing the same learning intervention, with random exposure to treated neighbors, and find that the individual treatment effects we estimate are almost entirely attributable to households who have treated neighbors (i.e. within-treatment spillovers). Comparing our results with the cluster-randomized experiment,

[1] By dose-responsive, we mean that receiving a treatment $k+1$ times affects outcomes differently than receiving a treatment k times.

[2] In complex (simple) contagion, nodes need more than one (exactly one) "contact" to become "infected". In learning about new technologies, complex contagion implies that individuals need exposure to information about the usefulness of the technology more than once to adopt the technology.

which is otherwise *ceteris paribus*, a back-of-the-envelope calculation suggests that interpreting the estimates from the cluster-randomized experiment as the treatment effect of receiving the intervention alone would have led us to overestimate the individual treatment effect by a factor of six.[3]

This finding has significant implications for scaling policy interventions. When an intervention's success relies on within-treatment spillovers rather than individual treatment effects, achieving efficacy – even at the individual level – requires a sufficient degree of treatment saturation. In other words, the size of the within-treatment spillovers of an intervention has direct implications for how a social planner administers and scales an intervention in policy design.

For individually randomized experiments where there are cross-treatment spillovers, comparing average outcomes between treated and control units may conceal individual treatment effects in the treatment group. A common way to address this is through clustered randomization. In cluster-randomized experiments, researchers randomize clusters of individuals – groups that ostensibly do not interact with each other – to treatment or control, reducing the possibility that treated units' treatment status will influence outcomes among control units. However, cluster-randomized designs with full treatment saturation, where entire clusters are assigned to treatment or control, can *not* address the possibility of complementarity between receiving treatment and being exposed to other people who receive treatment. Without random variation in exposure to other treated units among *treated units* themselves, within-treatment spillovers are not identified and the threat of a SUTVA violation is not eliminated. Thus, such designs cannot distinguish between individual treatment effects and within-treatment spillover effects.

Two-stage cluster-randomized designs with varying treatment saturation *can* identify within-treatment spillovers by randomizing clusters to different treated-to-control ratios and then assigning individuals within clusters accordingly. This creates random variation in exposure to treated units for both treated and control units, making it possible to measure and understand spillovers comprehensively. However, these designs are rarely used in practice due to logistical challenges and reduced statistical power for detecting individual treatment effects. Moreover, when implemented, they often focus on forces generating cross-treatment spillovers, such as general equilibrium effects in markets, rather than on the forces that we argue are most likely to generate within-treatment spillovers: the behavioral and social forces behind learning and technology adoption. Technology adoption and communication are fields of economic research whose experimental designs are *more* often cluster-randomized with full treatment saturation *than any other experimental design*, thereby concealing within-treatment spillovers.[4]

To understand the role of complementarities between experiential and social learning in chang-

[3]In both trials, we can estimate the combined individual and within-treatment spillover treatment effect. In the individually-randomized trial, we can separately estimate the individual treatment effects and within-treatment spillover effects, and find that the individual treatment effects account for 15% (or less than one-sixth) of the combined individual and within-spillover treatment effect. If we assume that the ratio of individual treatment effects to within-treatment spillover effects is consistent across trials, then only 15% of our estimate of treatment efficacy in the cluster-randomized trial is explained by individual treatment effects.

[4]See Section H for an analysis of trials registered to the American Economics Association RCT Registry.

ing health behaviors, we conduct a field experiment with 1800 households in the peri-urban slums of Karachi, Pakistan. Our experiment has four arms: control households, who receive no intervention; chlorine households, who receive free chlorine tablets; incentivized households, who receive tablets and small daily financial incentives for chlorine use; and learning households, who receive tablets and a visual tool that helps them track their household-specific health returns to chlorine use. Incentive and learning tool provisions are temporary, but we continue providing chlorine tablets (to all treatment households) and tracking chlorine presence in drinking water for over 12 months following intervention withdrawal. A comparison of the incentive and learning arms allows us to disentangle habit formation from learning, as both may raise the consumption stock of chlorine use in the short run, but only the latter can alter the perceived health returns to use. A comparison of both of these with the chlorine arm allows us to disentangle the degree to which access to chlorine is a barrier to use. Finally, a comparison of chlorine with pure control offers us a useful policy benchmark against which to measure the relative health impacts of incentives and learning.

Our learning arm builds on the intervention of Akram and Mendelsohn (2021) (hereafter, AM). They developed a visual tool that plots the health returns to water treatment by (1) utilizing household-specific data on the frequency of diarrhea among treated children, (2) providing a community-specific benchmark against which to estimate health returns to chlorination, and (3) generating multiple observations over time per household before and after chlorine tablet provision during the peak diarrhea months of the year, facilitating learning from repeated data and trends in health. The information is presented through a simple visual tool (henceforth referred to as the Info-tool) in the form of a colorful bar chart that plots days of diarrhea for one's own treated household relative to the expected rate of diarrhea among households that do not use chlorine (Luby et al., 2006). Community health workers (CHWs) present households with the comparison rate twice each month for six months: three months prior to chlorine tablet provision and three months following provision.[5] AM, who also work in peri-urban Karachi, cluster-randomize the Info-tool treatment at the neighborhood-block level. Fifteen months after ending the Info-tool treatment, they find a 197% increase in the rate of chlorine detection in Info-tool households' water relative to free chlorine alone (from 31% to 86%, with no change among the chlorine-only group).

The present study utilizes the same Info-tool intervention in our experiential learning arm and implements the study in peri-urban slums nearby, but not overlapping with, those of AM eight years after their intervention was completed. One key difference is that, while they randomize the Info-tool at the community level, we do so at the individual (household) level. Testing the same intervention at a new level of randomization not only allows us to see if our results replicate under a new methodology but also helps us identify within-treatment spillovers as the underlying mechanism generating large treatment effects, which cluster-randomization conceals.

[5]Designed through extensive piloting both in AM and for the present study, the tool is easily usable and comprehensible to the low-literacy and low-numeracy households in our sample.

We find that all three interventions (chlorine, incentives, and learning) significantly increase chlorine presence in drinking water for as long as we provide chlorine tablets, approximately fourteen months. While incentive households perform better than learning households during the period of incentive and Info-tool provision (the "short-run"), learning households outperform incentive households thereafter, with child diarrhea rates dropping the most for learning households following the withdrawal of interventions. Notably, however, differently from AM, we cannot rule out equality in average chlorination rates between the learning and chlorine-only arms in the long run. Recall that, while AM saturate the Info-tool treatment at the neighborhood level, we randomize the learning treatment at the individual/household level. This motivates an exercise to test for learning spillovers through social learning.

We consider two possible ways in which there might be learning spillovers. Experiential learning may be infectious in the social network, wherein non-learning-arm households are more likely to use chlorine when they have learning-arm neighbors, diluting our ability to detect treatment effects (cross-treatment spillovers). However, information diffusion by social learning *does not* always fit epidemiological models of information transmission (Chandrasekhar et al., 2022). An alternative explanation of learning spillovers that fits our results is a model where experiential learning and social learning are complements, and learning-arm households are more likely to use chlorine when they have other learning-arm neighbors (within-treatment spillovers). In this case, individual randomization reduced the number of learning-arm neighbors a learning household has, thereby weakening the potential impact of the treatment. This distinction is important for policy. If experiential learning is infectious, the optimal policy would be to seed experiential learning to network-central individuals who can then "infect" others with this information. If experiential and social learning are complements, meaning that participants are most likely to change their behavior when they receive *both* information from their own experiences and from those of their social connections, the optimal policy may be to geographically saturate experiential learning (if the within-treatment spillover effect is strong enough, and the intervention cheap enough).

We find that experiential and social learning are complements. Participants in the learning arm who also have a neighbor in the learning arm use chlorine at a significantly higher rate than any other group. We find that this complementarity in learning is in fact *essential* to sustained use: the higher rates of utilization that we observe among participants in the learning arm fade out after one quarter *unless* they have a neighbor also in the learning arm. Participants outside of the learning arm behave non-differentially across access to a neighbor in the learning arm. Taken together, these results suggest that, while individuals do alter their behavior based on their own experience, experiential and social learning *rely* on one another – neither channel in isolation is sufficient to sustain long-term use in our context. The consequences of learning are substantial: one year after the withdrawal of interventions, children of caregivers in the learning arm exhibit the largest improvement in child health. The treatment leads to a 0.24 SD ($p = 0.027$) increase in an index of child anthropometrics among any participants who use chlorine (as instrumented by randomization into treatment), a 0.33 SD ($p = 0.005$) increase

among *learning arm* participants who use chlorine, and a 0.51 SD ($p = 0.002$) increase among learning-arm participants who use chlorine *and* have a learning arm neighbor.

To discipline our exploration of the learning process that Info-tool households may be engaging in, we outline a model of experiential learning. We propose a model where actors are sensitive to the source of the information they receive when deciding to act on that information, are more responsive to signals acquired through a source that they trust, and build trust in information sources through personal experience. While we label increased responsiveness to signals acquired through a source that an individual has personal experience with as "trust", we define trust broadly and allow for related channels such as credibility, persuasion, or ownership effects to operate through the trust mechanism. We argue that our intervention generates within-treatment spillovers by providing Info-Tool participants with firsthand experience in acquiring information through the Info-Tool, which subsequently increases their trust in information acquired by *anyone* through the Info-Tool.

Participants follow Bayesian learning when forming their beliefs about the efficacy of a health technology. Different from standard models, our model also allows individuals to weigh signals with action weights when choosing to adopt the technology, which are independent of signal sign, size, or uncertainty. In our model, signals observed through more-trusted methods are given higher weights than signals observed through less-trusted methods. We interpret the Info-Tool as a technology that teaches participants to place more trust in information collected through this tool. This leads individuals to place more weight on the health signals they receive when recording their health status than those that they observe but do not record, and to place more weight on health signals from somebody else whose information was also acquired through the Info-tool.

Whether trust is an input into signal precision or action weights is an empirical question. If trust, and thereby the Info-Tool, affect signal precision, then stated beliefs should update in concert with chlorine adoption in the Info-Tool group. Furthermore, the Info-Tool participants should then also *convey* a different set of beliefs to others. Conversely, if trust is an input into how actionable a signal is, then stated beliefs and conversational content need not update along with chlorine adoption.

We do not find evidence that participants in the learning arm are more likely to believe that chlorine is an effective technology, to have an accurate memory about its effectiveness, or to share different conversational content about chlorine. While this may be because our questions are an imperfect estimation of beliefs, we think it is more likely that the Info-Tool affects the probability of acting on a belief without affecting the beliefs themselves. Our questions about beliefs in chlorine efficacy were simple to answer, accurate answers were incentivized through gifts, and we asked the beliefs questions during the period when the Info-Tool treatment effect was strongest. Furthermore, we find that participants with any education were more likely to answer accurately, suggesting that the questions *are* sensitive to comprehension skills. Instead, we find that learning arm participants are more likely to say that they are motivated to use

chlorine to achieve a standard of health for their families, which we interpret as a stated action-weighted belief. Moreover, learning-arm participants are more likely to believe that other learning-arm participants would know more about child health than participants in the chlorine or incentives groups, which we interpret as a measure of trust.[6] While nearly all households express that chlorine is a useful and effective technology, the Info-tool appears to make that information trustworthy and thereby motivates action. These results are consistent with Hussam et al. (2023) and Fafchamps et al. (2024), whose interventions induce changes in behavior *without* changes in beliefs or knowledge. From this, we conclude that trust in the Info-Tool is an input into the action weights on signals, rather than signal precision.

Our novel model of learning and technology adoption is micro-founded in an extensive body of evidence documenting that people are hesitant to trust health information relayed by experts (Alsan and Wanamaker, 2018; Alsan and Eichmeyer, 2024; Darden and Macis, 2024; Lowes and Montero, 2021; Martinez-Bravo and Stegmann, 2022; Banerjee et al., 2023), yet update their beliefs *or actions* strongly in response to personal experiences (Bennett et al., 2018; Conlon et al., 2022; Corno, 2014; Malmendier and Nagel, 2015; D'Acunto et al., 2021; Simonsohn et al., 2008). While there are numerous reasons to be skeptical of learning through one's personal experience – particularly in the health domain, which is precisely why the medical community relies on clinical trial evidence to inform medical and behavioral recommendations – reliance on experts or networks to disseminate information may inadvertently deepen social and economic inequities or undermine intervention efficacy. This is due to disparities in trust of experts, access to experts (Dussault and Franceschini, 2006), and access to well-informed networks (Banerjee et al., 2019; Calónico et al., 2023; Chen et al., 2022).[7]

Our model argues the mechanism driving changes in chlorination behavior is a change in the weights participants place on health signals when choosing to act on their beliefs. To test this mechanism, we first check whether individuals are responding to health signals at all. Using a machine learning algorithm on a host of baseline variables, we identify households that are *predicted* to experience health improvements during the intervention period. We find that, during the three months that incentives and the Info-tool interventions are ongoing, the effects of the learning intervention on chlorination are 29% ($p = 0.049$) greater among these households who are predicted to improve relative to their non-predicted-improved counterparts within the Info-tool arm, suggesting that participants indeed respond to the information that they observe in their environment.[8] This separation does not exist for the same predicted-improved households in the Chlorine-Only or Incentives treatment arms, suggesting that the

[6]This analysis takes into account the propensity to choose someone from their own treatment group (see Table 6).

[7]Health is a high-dimensional problem that is subject to random shocks and reliant on many inputs. These inputs are sometimes within an individual's control (for example, compliance with medical treatment, nutrition, and exercise), but oftentimes outside an individual's control (for example, pollution or traffic accidents) and potentially unobservable (for example, pathogens). Furthermore, health outcomes are noisy and rare occurrences that are difficult to observe precisely. Misattributing health inputs to health outcomes could bias beliefs and lead to suboptimal health behavior.

[8]This effect continues but diminishes over the course of the fourteen months, likely because the other Info-tool households are also learning over time.

pattern we identify is unlikely to be due to selection.[9]

Next, we check if learning-arm participants are responding to their learning-arm neighbors' health signals. We find that the spillover is entirely explained by learning-arm neighbors whose diarrhea rate was ex-ante predicted to improve. Learning-arm participants do not respond differentially when they have access to *any* neighbor who was predicted to improve, bolstering the assertion that the spillover is driven by differential weights placed on learning-arm neighbors' health signals, rather than proximity to people with positive attributes that are correlated with health improvements.

We are able to rule out habit formation, a common confounder to learning-through-adoption in the literature, as a competing mechanism by comparing the learning arm to the incentives arm. While we model the learning arm as a program that facilitates participants in trusting health signals gathered from the Info-Tool, it may instead work by leading to higher rates of early adoption which creates habit formation via building a greater 'consumption stock' of chlorination. Incentives for chlorine use should lead to high rates of early adoption *without* helping participants observe more salient health signals, so habit formation (or any technology adoption mechanism that relies on early adoption) should be realized in this treatment arm.[10] Indeed, participants in the incentives arm use chlorine tablets at higher rates in the short-term, but they immediately revert to a lower rate of chlorine use after the withdrawal of incentives.

As detailed in the paper, we are able to rule out a variety of other competing mechanisms as explanations for the social learning we observe, including differential communication, mimicry, social network formation, and changing social norms.

We highlight four central contributions of this study. First, we show that there can be significant implications of interpreting estimates of treatment effects from cluster-randomized designs as individual treatment effects. Estimates from cluster-randomized designs with full treatment saturation conflate individual treatment effects with within-treatment spillovers, and extrapolating these estimates to other settings (for example, settings with sparse treatment implementation that may not generate the same within-treatment spillovers) can bias analyses. We can distinguish within-treatment spillover effects from individual treatment effects, and compare these estimates with the combined estimate in a cluster-randomized test of the same intervention.

Second, we add to the literature on technology adoption and behavioral change by proposing a

[9]Reassuringly, participants who are predicted *ex-ante* to observe null or negative health signals during the facilitated learning experience are *no less likely* to chlorinate their water than other participants outside this treatment arm, indicating that the potential for misattribution to lead to negative belief formation about the health technology is not realized.

[10]We can also use the incentives arm to rule out an alternative learning mechanism, which is that early adoption leads to the accumulation of *more* signals. Salient signals may work more effectively if participants have limited attention (Hanna et al., 2014). However, if participants can attend to and learn from many signals, these signals will be more likely to converge to the truth, which will decrease the potential for misattribution. Our evidence is consistent with Hanna et al. (2014), and suggests that participants are not able to attend to the many signals that they generate without the Info-Tool making these signals salient.

model of learning in which individual experience and social learning are complementary. In our model, this complementarity is a key underlying mechanism by which repeated exposure to, or experience with, a good or behavior translates into long-term adoption. We generate evidence for this model using a randomized experiment that suggests that actors are sensitive to *how* information is acquired and more readily trust information whose source they are familiar with. This may contribute to why people more readily trust in-group members with health information (Alsan et al., 2019; Alsan and Eichmeyer, 2024). Furthermore, we show that the learning treatment changes behavior without changing underlying beliefs about the efficacy of chlorine, adding to a small but growing literature that shows that information that individuals possess is not the only, and potentially not the primary, process through which behavior changes. Lab-in-the-field evidence suggests that an ownership effect over information that an individual herself generates might be important in determining if this information stimulates behavioral change (Conlon et al., 2022). We document the importance of taking action to generate signals for learning outside the lab in a setting with high-stakes decision-making. However, we depart from Conlon et al. (2022) by showing that, under the right conditions, people can also learn from information that comes from other people.

Third, we operate in a setting with significant policy implications for scaling education and information campaigns in preventive health. The returns to preventive health behaviors, such as handwashing with soap, using clean cookstoves or bednets, or treating water, are particularly challenging to perceive the returns to. Existing literature on using social networks for information dissemination has focused on seeding information. This literature often assumes that everyone can learn from social connections if the information reaches them, but seeding omits the value of the experiences of the person *receiving* the information. Our evidence suggests that, for some types of learning, these assumptions could seriously undermine the potential for individuals to learn from disseminated information. In our setting, personal experience is essential to acting on information disseminated by others. *Own experience cannot be seeded*, suggesting that there might be large efficacy costs in seeding certain types of information, rather than saturating it.[11]

Finally, we are one of few papers in the learning-through-adoption literature that can definitively rule out habit formation as a confounding mechanism to learning, two behavioral processes that are often unknowingly conflated in existing work (see Section I.2 for a detailed review of the existing literature).

Our paper proceeds as follows: Section 2 provides the details of the experimental design, sample, and data; Section 3 presents a model of complementarities in experiential and social learning; Section 4 presents overall results on chlorine use, and spillover results on chlorine use; Section

[11]Note that subsidizing learning costs through experiential tools, such as the Info-tool, can generate positive health externalities not only through individual learning but also by enhancing social learning, as households receiving support indirectly reinforce their neighbors' adoption. While the first mechanism justifies, for instance, subsidization of chlorine tablets, the second mechanism makes the case for subsidizing learning costs via offering technologies such as Info-tool that facilitate both individual and social learning.

5 discusses mechanisms; Section 6 presents results on child health; and Section 7 concludes.

2 Experimental Design

We conducted a randomized controlled trial in Ibrahim Hyderi, a peri-urban neighborhood of Karachi, Pakistan. We identified 1,800 eligible households through a census. Households were eligible if there was at least one child between the ages of 6 months and 5 years old at baseline, and if the caregiver was generally home during the day. In all but one case the caregiver was a female adult, typically the mother of the children. Furthermore, we screened households on several dimensions of water usage to ensure that chlorine would be useful and effective for improving children's health.[12] Although 79% of households primarily drank centrally-delivered water that was piped into their own household or plot at baseline, we did not detect chlorine in the drinking water of a single household. By endline, 70% of households had changed their water delivery method (non-differential by treatment group), with the majority of households (56%) now receiving water from a public tap or standpipe (32% still primarily drank piped water into the household).

2.1 Baseline Survey, Randomization, and Balance

Our baseline survey of these 1,800 eligible households ran from mid-May to early June of 2022. We collected data on chlorine exposure, chlorine knowledge, diarrhea prevalence,[13] and child anthropometrics.[14] As expected, there was not a single household whose water vessel tested

[12]Households were excluded if they did not store drinking water in a separate vessel from the water for other uses, if children drank from a separate vessel from adults, if the vessel's capacity was less than 10 liters (to ensure there would be enough water to avoid over-chlorination), or if children frequently drank bottled water (see Figure F.3 for a sample vessel). Of the households we approached, 93.6% were eligible based on these criteria, and 80% of the ineligible households were ineligible on the basis that nobody was home, nobody could regularly be home in the future, or there was no child in the correct age range in the household. Only nineteen households (less than 1% of all households approached) were ineligible based on water infrastructure.

[13]We asked respondents to describe their perception of "motions", or diarrhea. Then, we described diarrhea to respondents as loose or watery stools. For respondents who were not sure what classifies as diarrhea, we also shared a visualization of stool types called the Bristol stool chart. Respondents then reported how many days each child under five years old experienced diarrhea in the past fourteen days. Diarrhea rates were higher than expected during the baseline period. Average rates of diarrhea in May are close to 3% of days per child in the literature, whereas households in our sample experienced diarrhea on 5.9% of child-days (Luby et al., 2006). There was an extreme heatwave ongoing in Karachi during the baseline period, which we believe explains the high rates of diarrhea (Xu et al., 2014). Diarrhea rates converged to the expected rate in the following month, without any increase in chlorine use.

[14]Anthropometric outcomes were only recorded at baseline and endline – child weight and mid-upper-arm circumference at baseline and endline, and child height and age at endline only. We calculated the children's weight as the difference between the weight of the mother alone and the weight of the mother while holding the child. For children older than 6 months, we also collected mid-upper-arm circumference. There were 338 households for whom we did not collect anthropometric data in the baseline visit due to issues with the measurement equipment. As such, we collected anthropometric data for these households in the following visit. We did not reveal treatment status to households until the first Phase 1 visit, and enumerators were blinded to treatment status during the baseline visit, so there is no reason to believe that treatment would differentially affect child anthropometrics between the baseline and the following visit. Furthermore, we visited these households at the beginning of the following visit to minimize the time between the end of the baseline survey and the time of measurement for any anthropometric outcomes that we consider as baseline measures.

positive for chlorine residual, as 98% of households reported that chlorine tablets were not available in local markets. Awareness of chlorine tablets in this community was likewise low: Twenty percent of the sample reported having heard of chlorine tablets, but forty percent of this sample could not properly identify what chlorine tablets were used for. Despite low awareness, eighty-seven percent reported that they would consider using chlorine tablets after enumerators explained the purpose of chlorine for purifying water.

We electronically randomized households to treatment groups between the baseline visit and the first biweekly visit. Eligible households were individually randomized into the four 450-person experimental arms as follows:

Control (C): Pure control (no offer of free chlorine tablets; chlorine testing)

Treatment 1 (T1): Free chlorine tablets and chlorine testing (Chlorine Only)

Treatment 2 (T2): Free chlorine tablets and chlorine testing + Info-tool (Info-tool)

Treatment 3 (T3): Free chlorine tablets and chlorine testing + financial incentive (Incentives)

Treatment groups are balanced on most, but not all, baseline measures. In Tables A.1 and A.2, we compare each treatment group with the pure control group. Children from T1 are reported to have fewer diarrhea days than the control group both when considering a binary measure for whether or not any child in the household had diarrhea in the past fourteen days (Table A.1), or when considering both a binary measure and the number of diarrhea-days at the child level (Table A.2). However, there are no significant differences in child anthropometrics, nor do we see lower levels of diarrhea among T1 households in the following visit (before households had received any treatment interventions), indicating these differences are likely spurious.

		Phase 1		**Phase 2**		
		Round 1	Round 2			
	Month 1	Pre-Treatment: Months 2-4	Treatment: Months 5-7	Post-Treatment: Months 8-18	Month 19	Month 24
Chlorine	Baseline Survey		Chlorine Distribution	Chlorine Distribution	Endline Survey 1	Endline Survey 2
Incentives			Chlorine Distribution + Incentives			
Info-Tool		Info-tool Training	Chlorine Distribution + Info-tool			
Control						
		Twice-monthly visits for child health data	Twice-monthly visits for child health and chlorine presence data	Once-monthly visits for child health and chlorine presence data		

Figure 1: Study Timeline

2.2 Phase 1: Biweekly Visits

Phase 1, during which we introduced the behavioral interventions and visited households every two weeks, consisted of two rounds. Throughout Round 1, households were aware of their treatment status and we trained T2 households on how to use the Info-tool chart. In Round 2, we distributed chlorine tablets to all three treatment groups and delivered the incentives treatment to T3 participants, while T2 households continued to use the Info-tool chart. Figure 1 lays out the study timeline and key events.

2.2.1 Round 1 of Phase 1

Throughout Round 1 of Phase 1 (June to August 2022), households were visited by surveyors every two weeks for a short round of data collection. These surveys were standardized and involved a few key components: measuring child diarrhea days over the past two weeks, testing for the presence of chlorine residual, and training caregivers in the T2 group on how to use the Info-tool chart.[15] Since chlorine distribution had not yet begun, and chlorine tablets are not widely available in local markets, ten percent of households were randomly selected for chlorine testing during every other visit during Round 1 of Phase 1. Chlorine testing in this period was simply to ensure that knowledge of treatment status did not affect people's ability to procure chlorine tablets on their own.

In the first Phase 1 visit, we revealed treatment status to caregivers. We explained each treatment group to caregivers and told them that a lottery would determine their treatment status. We explained that the lottery would run through the enumerator's tablet, and that whichever number would appear on the screen would determine their treatment status. These numbers were pre-loaded so as to deliver the treatment status we pre-assigned. After treatment status was revealed, we reminded participants about what being a part of their treatment group would entail.

Info-tool

To activate the 'learning' mechanism, participants in T2 received the Info-tool chart on the first visit of Phase 1.[16] The Info-tool chart was a simple pen-and-paper chart that allowed caregivers to track their children's diarrhea (Figure F.1). Each chart consisted of two bars to represent

[15] In the first biweekly survey, we also collected data on household bargaining power, as we were unable to incorporate this into our first baseline visit. Since we ask caregivers questions about decision-making power before revealing treatment status, we consider this as a baseline measure of household bargaining power. There were no differences across treatment groups in being the sole decision-maker, or a part of decision-making, in issues of child health, household purchases, or household visits. We also collected anthropometric data for 338 households for whom we were not able to collect anthropometric data in the baseline visit.

[16] The study was non-blinded with respect to study subjects and treatment administrators (CHWs). Enumerators were not informed by the research team on the specific treatment status of the households at baseline. While all enumerators were hired through our partner organization, Interactive Research and Development (IRD), we recruited a different team for the endline survey to ensure objectivity, and these enumerators were not informed of the participants' treatment status before beginning the survey. However, participants were not blind to their own treatment status, so it is possible that they discussed their treatment status with the enumerators and effectively rendered the endline survey non-blinded. Community health workers used electronic tablets and smartphones to collect data.

each month: one bar in which caregivers filled in a square for each child-day with diarrhea (for example, if two children had diarrhea on the same day, the caregiver would fill in two boxes), and one bar in which the enumerator filled in a benchmark diarrhea rate. The benchmark diarrhea rate was calculated in two ways: initially, by using the fourteen-day diarrhea rate in the pure control group across all children, multiplied by the number of children in the household; and subsequently, by using the monthly incidence of childhood diarrhea from the epidemiological literature (Luby et al., 2006). When we used the diarrheal incidence in the Control group, the benchmark was updated daily so as to reflect the past fourteen days of data collection from the pure control households. We switched from using the diarrhea rate in our Control group to the diarrhea rate in the epidemiological literature out of concern that the incidence of diarrhea in the Control group would be impacted by a reduced disease environment due to the presence of treated households in their community. Enumerators clearly explained to participants that the benchmark rate was the average rate of diarrhea from people in the community who did not use chlorine to purify their water.

This simple, visual paper-and-pencil tool was borrowed from Akram and Mendelsohn (2021), which was comfortable for low-literacy-and-numeracy caregivers to use. In this way, T2 caregivers were able to compare their children's diarrheal incidence with the average diarrheal incidence in households that did not use chlorine tablets, both before they were offered chlorine tablets (months 1-3 of the intervention) and after (months 4-6 of the intervention).

2.2.2 Round 2 of Phase 1

Round 2 of Phase 1 began in late August 2022 and continued for three months. Throughout this round, we continued to visit households every 2 weeks and commenced the distribution of chlorine tablets to all T1, T2, and T3 households. We continued to collect information on diarrhea prevalence and tested water for the presence of chlorine in all households. We tested for chlorine residual every other visit (once per month).

Incentives

At the beginning of Round 2, we also began the incentive treatment with T3 households. This treatment, which should build consumption stock, was designed to test learning against habit formation as the primary mechanism. Caregivers were offered tokens redeemable for child and household goods if they could show empty chlorine tablet wrappers as proof of usage. The chlorine tablets came individually wrapped and participants were instructed to save the empty wrappers in a small pouch provided for them. Each daily reward for proper chlorine use was equal to approximately 5 US cents (with 'proper use' calibrated to the household's pre-intervention water consumption). To hold income effects constant, we also gave comparable products to participants in the remaining groups – *unconditioned* on chlorine tablet wrappers – as a token of appreciation for participating in our surveys. Since T3 households could redeem tokens for household goods that had varying values, we implemented a lottery to determine the value of the unconditional gifts that non-T3 households received.

Info-tool

At the end of Round 2, we aggregated the monthly Info-tool chart statistics across three-month intervals and presented the aggregated data to T2 participants visually (Figure F.2). As such, T2 participants were able to visualize their own children's diarrhea rate relative to the average diarrhea rate among households that did not use chlorine in the three-month interval before chlorine tablet distribution, and make the same visual comparison for the three-month interval following chlorine tablet distribution: in effect, a difference-in-differences.

Throughout Round 2, we conducted random unscheduled audits where CHWs tested water for the presence of chlorine residual. These audits were conducted to ensure that households were not chlorinating their water in anticipation of our bi-weekly visits and to increase confidence that our scheduled monthly chlorine tests serve as a good proxy for chlorine use throughout the month (audit visits are discussed in more detail in Section J).

2.3 Phase 2: Monthly Visits

Phase 2 began in late November 2022 and continued through October 2023. In this phase, we visited households once per month. We continued to distribute chlorine tablets and test water for the presence of chlorine residual but ceased provision of incentives or gifts, no longer informed Info-tool participants of their own children's diarrhea rate relative to the benchmark diarrhea rate, and stopped providing Info-tool sheets or helping Info-tool participants track their children's diarrhea rates.

2.4 Endline Visits

The first endline survey was completed in December 2023 through January 2024. In this round, we collected all of the measures from our baseline: diarrhea rates; presence of chlorine residual; child anthropometrics, including weight, mid-upper arm circumference, and height; and household bargaining power. We additionally elicited willingness-to-pay for chlorine tablets using a take-it-or-leave-it offer with a randomized price.[17] Finally, we collected detailed social network data at endline to understand potential mechanisms for spillovers, including directly serving neighbors chlorinated water (through meal-sharing, for example), and indirectly through conversations about the project and lessons learned through the Info-tool.

We were able to find 87% of households for the endline survey. Incentives households were the most likely to attrit (14.9%), and were 4.2 percentage points more likely to attrit than Control ($p < 0.05$; Table A.1). However, across a range of baseline observables, the endline Incentives respondents do not statistically differ from their counterparts in the other arms (Tables E.5 and E.6). Moreover, none of our primary analyses at endline rely on a comparison between Incentives and Control. Our primary analyses compare Info-tool with Control or the Chlorine

[17] After providing respondents with a one-month supply of chlorine tablets, free of charge, we offered them a second month's supply of chlorine at a randomized price (market price, a 29% subsidy, or a 53% subsidy).

Only group, each of whose endline attrition is non-differentiable from the other.

We conducted a final short followup survey in June 2024, motivated by three reasons: (1) to provide participants with chlorine tablets for the summer months, the period in which diarrhea rates peak, (2) to monitor stockpiling of chlorine tablets, and (3) to ask additional survey questions to disentangle mechanisms.[18]

Taken together, we administered 26 surveys per household: a baseline survey, 11 rounds of bi-weekly visits in months 1-6 of the intervention (Phase 1), 12 monthly visits in months 7-18 of the intervention (Phase 2), one endline visit immediately after Phase 2, and a second endline visit four months later. Our primary outcome, chlorine residual in drinking water, was recorded in nineteen of these visits (baseline; bi-weekly visits 6, 8, 10, and 11; and every visit thereafter). We randomly tested households for chlorine residual in months 1-3 to ascertain that people never attained and used chlorine before we began distribution.

2.5 Outcomes

Our primary outcomes are chlorine use and child health. Our measure for chlorine use is a binary indicator for the presence of chlorine residual in drinking water. After we started distributing free chlorine tablets, CHWs tested for chlorine residual in each household's drinking water vessel every month using a simple test strip that, when dipped into a small cup of water, turns shades of blue depending on the chlorine concentration in the water (the minimum amount of chlorine we can detect with our test strips is 0.5 parts per molecule). The presence of chlorine residual is an objective measure of chlorine tablet use that cannot be manipulated. However, chlorine residual only presents itself if the chlorine tablet was added to the water in the past twenty-four hours, meaning we may not capture a treatment effect for households who imperfectly chlorinate their water.

Our measures of child health are self-reported diarrhea incidence and child anthropometrics. Diarrheal incidence is measured by the number of days of diarrhea across all children under five years old, which we collect in every visit. While we do collect data on the frequency of childhood diarrhea among households, we do not use diarrhea as our only measure of child health because it is subject to measurement and misclassification errors. Firstly, our interventions, particularly Info-tool, are likely to change the reporting of diarrhea in children, since we explicitly encourage participants to track diarrhea. Thus, we may capture a treatment effect on reporting, rather than a treatment effect on actual incidence of diarrhea. Secondly, the presence of non-infectious

[18]We sought to monitor stockpiling because there was a sharp drop-off in rates of chlorine detection in the last two rounds of surveying (the final survey of Phase 2, and the first endline survey), which was precisely when we began to tell respondents that the trial was nearly complete and that we would soon cease to provide free chlorine tablets. Since we ended the survey during the winter, the time of year when diarrhea rates are lowest, stockpiling chlorine tablets for the summer would have been a rational response from participants. We find some evidence of this: although only 3% of respondents reported that they still had chlorine tablets remaining in this final followup survey, 44% of respondents reported that they ran out of their chlorine tablets later than they should have if they were chlorinating their water daily, and for 24% of households the discrepancy was by greater than one month.

diarrhea and asymptomatic infection make diarrhea presence a poor outcome for measuring the effectiveness of water and sanitation interventions (Watson et al., 2022).

Frequent diarrhea can result in poor absorption of nutrients, leading to sub-optimal physical development in children. As such, we supplement diarrhea incidence with child anthropometrics, including height-for-age, weight-for-age, weight-for-height, and mid-upper-arm-circumference-for-age, which we collected at endline among children under 80 months old[19].

3 Model of Learning and Technology Adoption

We present a Bayesian model of learning about a technology, following the notation of Kondylis et al. (2023), but modify the model so that some signals are more "actionable" than others, regardless of the signal sign, magnitude, or precision. This does not affect explicit knowledge about the technology itself, but instead affects the way that posterior beliefs translate into action. Let A be a distribution of prior beliefs about the efficacy of the technology:

$$A \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

In each period, individuals observe or hear about a signal Y with precision given by σ_Y^2:

$$Y|A = a \sim \mathcal{N}(a, \sigma_Y^2)$$

Individuals then update their posterior beliefs according to Bayes rule:

$$A|Y \sim \mathcal{N}(M, \sigma^2)$$

where $M \equiv \ell Y + (1-\ell)\mu_0$ represents the updated expected returns to using the technology (in our case, M represents posterior beliefs about chlorine efficacy). Weights $\ell \equiv \sigma_0^2/(\sigma_0^2 + \sigma_Y^2)$ represent learning. Then, participants' updated uncertainty is $\sigma^2 \equiv (1-\ell)\sigma_0^2$.

Actions versus Beliefs

We will consider two margins along which information affects technology adoption. First, information makes people explicitly more knowledgeable about the efficacy of the technology, which they will incorporate into their posterior beliefs M. This explicit knowledge will be reflected in their stated beliefs. Second, information makes these beliefs more or less actionable. People can intellectually understand the returns to a technology, and even be confident in that information. However, how and from whom they get that information may affect the decision to act on it.

Let $\alpha_Y \in [0, 1]$ be a weight determining how actionable a signal is, where $\alpha_Y = 1$ is a fully actionable signal, and $\alpha_Y = 0$ is a completely non-actionable signal. How actionable an individual

[19]This is because, in baseline, we measured children who were under 5 years or 60 months old.

determines a signal to be is not related to the precision of the signal itself, which is modeled in the variance σ_Y^2. Instead, this weight accounts for an individual's willingness to act on information, *regardless* of the informational content or uncertainty in the information. Now, M still represents *stated* posterior beliefs about the returns to the technology (or explicit knowledge). Let $M_\alpha \equiv \ell\alpha_Y Y + (1 - \ell\alpha_Y)\mu_0$ be the action-weighted posterior belief. An individual adopts the technology when $M_\alpha > C$, where C is the cost of adopting the technology[20].

A model that would generate a similar pattern of adoption is one where the intervention leads certain signals to be more precise (lower σ_Y^2), rather than increasing the action-weights α_Y of those signals. However, with a constant C, this model requires that stated posterior beliefs and adoption move together. Conversely, our model *does not* imply this prediction, and instead allows stated posterior beliefs and adoption dynamics to follow different patterns from one another.

Let $\gamma_Y \in \{\Gamma\}$ be the source of the signal Y from a set of potential sources Γ. We make two key assumptions: α_Y depends on trust in γ_Y, defined as T_{γ_Y}; and T_{γ_Y} depends on the stock of previous personal experience acquiring signals from γ_Y, defined as E_{γ_Y}.

$$\alpha_Y = \alpha_Y(T_{\gamma_Y}, \cdot)$$

$$T_{\gamma_Y} = T_{\gamma_Y}(E_{\gamma_Y}, \cdot)$$

We assume that other reasons for an individual to be skeptical towards Y are exogenous. While we model action weights as a function of trust in the tools used to acquire that information, note that we define trust broadly, and there are related mechanisms that can make signals more or less actionable, independent of signal uncertainty. For example, an ownership effect (as in Conlon et al. (2022)), signal persuasiveness (as in Fafchamps et al. (2024)), or identity of the signal-sender (as in Alsan et al. (2019) and Alsan and Eichmeyer (2024)) could contribute to a signal's action weight.

First period adoption

At baseline, only 12.7% of respondent said that they had heard of chlorine before and could name that its purpose was to purify water, implying that initial priors are diffuse. We consider information sent by the Community Health Worker (CHW) as the first signal that participants receive about chlorine, Y_{CHW}. It is likely that $\sigma^2_{Y_{CHW}}$ is relatively low, considering that the CHW represents a trusted NGO and had already visited households for three months regularly prior to distributing chlorine. It is also likely that C is close to zero in the first period, since the largest reported cost associated with taking chlorine is an unpleasant taste if the chlorine dose is not titrated just right, which caregivers would not yet have experienced.

[20] In a standard model, participants adopt the technology when $M > C$

Interventions:

All chlorine groups: We reduce the cost of using chlorine C in all time periods, through free distribution and free delivery.

Incentives: We further reduce the cost of using chlorine C in the Incentives group for the first 3 months of chlorine distribution by providing monetary incentives for use.

Info-tool: We change the signal observation technology γ_Y in the first 3 months of chlorine distribution, which increases $E_{\gamma_Y}|\gamma_Y = IT$ in perpetuity. Consequently, we change trust in Info-Tool signals $T_{\gamma_Y}|\gamma_Y = IT$. This will be true both for signals personally observed via the Info-Tool, and signals sent by other Info-tool participants where the sender informs the receiver that she observed her signal through the Info-tool. Assuming that the Info-Tool is a trust-worthy source, the Info-Tool increases trust in all Info-Tool signals (acquired by oneself or through others).

Comparative statics:

Shock to C for Incentives:

Prediction 1: Increased contemporaneous adoption relative to Chlorine Only

A negative shock to C will increase the probability in any period that $M_\alpha > C$, leading to increased contemporaneous chlorine adoption. This effect will be immediate from the first period of chlorine distribution, but will fade out as soon as the CHW stops providing chlorination incentives.

Shock to $T_{\gamma_Y}|\gamma_Y = IT$ for Info-tool:

Prediction 2: Increased contemporaneous adoption relative to Chlorine Only among individuals with larger early-period health signals, with fade-out

The Info-Tool leads to an increase in $E_{\gamma_Y}|\gamma_Y = IT$. In the first 3 months of adoption, while personally observed signals are acquired via the Info-Tool, participants will weight their own signals with $\alpha_Y|\gamma_Y = IT$. Assuming that participants believe the Info-Tool is a trust-worthy technology, Info-Tool participants place a contemporaneously higher weight α_Y on signals acquired through observation during the treatment period. After the CHW stops assisting participants in the Info-tool, and stops providing gifts to the Incentives groups, costs are the same for all three groups. However, Info-tool participants' action-weighted prior will be higher at the time that the behavioral interventions end. Thus, we should see higher average use immediately after the behavioral interventions end.

Since individuals learn from short-term health signals, we should expect to see heterogeneity by the sign on individual contemporaneous health signals, and the Info-tool group should be the most sensitive to small differences in health signals. However, we should expect this effect to fade out as participants acquire new signals from their observations on which the Info-Tool

group no longer places higher weights. Furthermore, placing higher weights on early health signals is not the only way in which Info-tool participants differ from the other groups. Info-tool participants also place heavier weights on signals from *other* Info-Tool participants (described in Prediction 3 below), which may supersede heterogeneity by individual early health signals. As such, heterogeneity by early health signals is suggestive of acting on observed signals, but the absence of heterogeneity by health signals is not dispositive.

Prediction 3: Increased adoption relative to Chlorine Only among individuals with social connections who were also in the Info-tool group

As soon as an Info-tool participant has used the Info-tool, which is a shock to $\alpha_Y|\gamma_Y = IT$ for the signals participants collect with the Info-Tool *themselves*, this is also a positive shock to any other Info-Tool-acquired signal, including those acquired by *other* Info-Tool participants. Thus, Info-tool caregivers with access to another network connection who was in the Info-tool group will also be more responsive to those network connections' health signals than caregivers in other treatment groups. Furthermore, Info-tool participants should be responsive to the sign of the signals they receive from other Info-tool social network connections. This can have effects after the Info-Tool period has ended, if participants are sharing net information rather than period-specific information.

We test our model using simulated data and, under a set of reasonable parametric assumptions, find adoption patterns that closely follow our empirical data without any changes in posterior beliefs across groups (Section G).

4 Results: Chlorination

4.1 Overall Average Treatment Effects

We begin by plotting chlorine use by treatment groups across the study period (Figure 2). The short-run is defined as Phase 1 (weeks 12-20). The medium-run spans the first three months of Phase 2 (weeks 22-32). The long-run spans the remainder of Phase 2 (weeks 36-72).

Free chlorine distribution leads to a large increase in chlorination in all treated groups. As we will see in Section 6, there are large decreases in reported child diarrhea and improvements in child anthropometrics in all groups, indicating that free chlorine distribution alone, even with low levels of perfect compliance, can have substantive impacts on child health. On aggregate, the impacts of the behavioral interventions are small in magnitude and at times statistically insignificant. We highlight two important patterns: first, during the short-run, Incentives households chlorinate more than those in Info-tool or Chlorine. Immediately following incentive withdrawal (week 20) however, chlorination rates for incentive households plummet, while the decay among Info-tool households is slower. Chlorination patterns across the treatment groups appear, on average, to follow one another thereafter.

On average, chlorine residual is detected in the water of over 20% of the treatment households

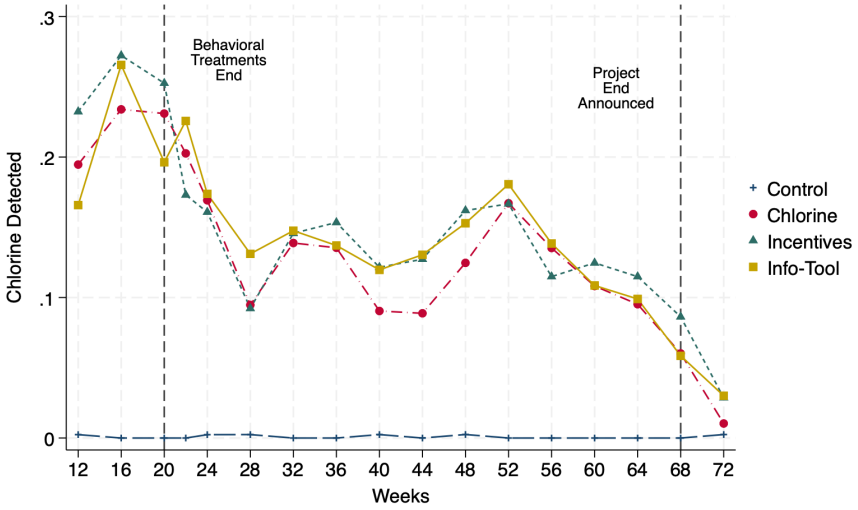


Figure 2: Raw Rates of Chlorine Detection

during the short-run (weeks 12-20). In the medium run (weeks 22-32, or over the three months post-intervention), the detection rate declines, especially among Incentives households. Info-tool households, in contrast, chlorinate more than the other two intervention groups throughout this period. In the succeeding period, however, Incentives households and Info-tool household converge around week 32, with households in Chlorine finally converging to their chlorination rate in week 52 (eight months after the end of interventions). Finally, there is a gradual decline in chlorination rates across all groups in the last four months of the study period. While our trial was not designed to identify the causes of chlorine use patterns across time, the serial pattern follows diarrhea rates, with all three groups chlorinating at higher rates during the high-diarrhea season. Furthermore, chlorination rates fell most sharply in the final two months of the study, which coincided with when we informed participants that the study was about to end and may have led them to stockpile chlorine tablets.[21]

[21]Overall trends in chlorine use might be explained by seasonality and by experimental changes that affected all participants.

Seasonality: Diarrheal prevalence peaks in the summer, during the hottest and rainiest months. We began our study at the peak season, which would allow chlorine use to generate starker changes in diarrhea rates (thereby giving the Info-Tool a greater chance of showing the efficacy of chlorine). Chlorine use tracked closely with self-reported diarrhea prevalence over the next few months, when both diarrhea rates and chlorine use rates fell (Figure C.1). In the following summer, diarrhea rates did not rise in the expected pattern, including in the Pure Control group, and chlorine use also did not reach its height of use in the previous year. It is possible that the intervention created herd immunity and reduced the aggregate disease burden in the neighborhood, or the community disease burden may have been lower because the summer rain was significantly milder than the previous year (flooding encourages the spread of infectious disease).

Experimental Changes: Chlorine rates fell sharply after the behavioral interventions ended (week 20), including in the Chlorine Only group, who did not receive a behavioral intervention. While diarrhea rates also fell abruptly at this time, there are two experimental reasons why chlorine rates may have fallen after week 20: (1) we began visiting households every month, rather than every two weeks, which means that each measure of chlorine residual came a longer time after the last opportunity to re-stock households with chlorine and remind participants about proper chlorine use; and (2) we stopped providing gifts to all households (conditional on chlorine use in the Incentives group, unconditional in the other groups), which may have had an effect on water chlorinating across all groups. In the last two months of the study, chlorine rates dropped to almost zero. While this sudden drop-off did not coincide with a sudden drop-off in diarrhea rates, diarrhea was low during this time period. Furthermore, these were the only visits since the beginning of the trial that we explicitly reminded

Our main empirical specification is the following:

$$Y_{i,s} = \beta_0 + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + Y_{i0} + X_{i0} + \delta_s + \gamma_b + \epsilon_{i,s}, \tag{1}$$

where $Y_{i,s}$ is an outcome for individual i measured in survey-round s; $T1_i$, $T2_i$, and $T3_i$ are binary variables representing Chlorine Only, Info-tool, and Incentive group participants, respectively; Y_{i0} is the baseline measurement of the outcome (whenever available); X_{i0} is a vector of unbalanced baseline covariates, the density of study participants within twenty meters[22], and other covariates that we select using the double-lasso method proposed by Urminsky et al. (2016); δ_s denotes survey-round fixed effects; and γ_b are block fixed effects (geographic units of stratification). We cluster standard errors at the household level.[23]

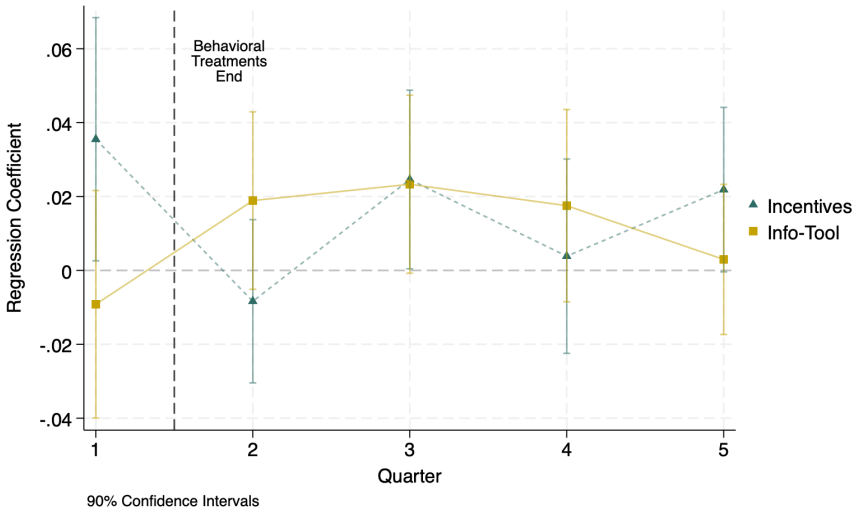


Figure 3: Regression Coefficients of Chlorine Detection (Omitted: Chlorine Only)

Figure 3 shows the regression results for chlorination using the panel dataset from household visits (also reported in Table C.1). Consistent with the observations from the raw chlorination plots, during the short-run period (weeks 12-20, or Quarter 1), households assigned to Incentives outperform those in Info-tool or Chlorine (p-values of 0.014 and 0.090, respectively). In the immediate post-intervention period (weeks 22-32, or Quarter 2), households in the Info-tool group chlorinate at a higher level than Incentives and Chlorine; however, contrary to the findings in AM, the difference in treatment effects between Info-tool and Chlorine is not statistically distinguishable (p-value = 0.142).

participants that the intervention was about to end. It is possible that households stockpiled chlorine tablets, knowing that they would soon lose access to a free supply of chlorine tablets and that the high diarrhea season was still several months away. We found some evidence of stockpiling in our second endline: although only 3% of respondents reported that they still had chlorine tablets remaining in this final followup survey, 44% of respondents reported that they ran out of their chlorine tablets later than they should have if they were chlorinating their water daily, and for 24% of households the discrepancy was by greater than one month.

[22]Section 4.2 details the selection of this bandwidth following Egger et al. (2022).

[23]If a participant could not be found or refused to let us test their water in any given survey round, we consider this observation as an attriter and drop it from the analysis.

4.2 Learning Spillovers

In our model, individuals update their beliefs about health technologies when they receive signals from themselves *and others*. Motivated by this theory, we investigate the effects of learning spillovers. Learning spillovers are also a natural potential explanation for why we are not able to replicate the large positive Info-Tool treatment effects that AM find, given that the key difference between the two studies is that we randomize the treatment at the individual level, and AM randomizes the Info-tool treatment at the neighborhood-block level.

Households are, on average (median), 13.6 (10.4) meters from the closest other participant; 16.0 (12.1) meters from the closest participant in any chlorine treatment group; and 27.5 (22.7) meters from the closest Info-tool participant. While caregivers are surrounded by other participants, only 43% of participants have a nearby neighbor (someone within twenty meters) in the Info-tool group. When we recenter this measure to account for endogenous neighborhood density (following Borusyak and Hull (2023)), only 35% of participants are exposed to a nearby neighbor in the Info-tool group by random variation in the treatment assignment (Table A.3 demonstrates that this measure is balanced across treatment arms).

We compute exposure to Info-Tool neighbors in the AM data using the same spillover definition. In that experiment, 49% of Info-tool participants were randomly exposed to other Info-tool participants by variation in the treatment assignment, while 2% of Chlorine Only participants were.[24] This means both that, in our trial, non-Info-tool participants are *more* exposed to other Info-tool participants than in AM, and that Info-tool participants are *less* exposed to other Info-tool participants. If there is a learning spillover from Info-tool participants onto non-Info-tool neighbors, our individual-level randomization would conceal Info-tool treatment effects; and if there is a learning spillover from Info-tool participants onto other Info-tool neighbors, the treatment would be less powerful in our setting.

Motivated by our theoretical framework of learning and the contextual differences between our study design and AM, we now test for learning spillovers from Info-tool participants onto their neighbors. We find that the Info-tool treatment has powerful spillovers, but these spillovers are only absorbed by other Info-tool participants. In other words, there is an Info-tool-to-Info-tool spillover.

First, we define our spillover measure. Because randomization was conducted at the individual level, individuals are randomly exposed to Info-tool neighbors. We consider somebody as being in the "spillover sample" if they have at least one Info-tool neighbor within an r-meter radius of where they live. First, we determine the number of Info-tool participants within each radius $r \in \{20(20)200\}$ meters. Next, we recenter these measures to purge our estimates of omitted variable bias that may arise from neighborhood density or other endogenous environmental factors, following Borusyak and Hull (2023). We construct $AnyT2_i^r$, a binary measure that indicates if the recentered number of Info-tool participants within radius r of individual i

[24] There were no Incentives or Pure Control groups in AM.

is greater than 0. We interpret this variable as indicating if participants were randomly more exposed to an Info-tool neighbor than they would be in expectation based on endogenous spatial factors. Following Egger et al. (2022), we select the R that minimizes the Schwarz Bayesian Information Criterion (BIC) for each $r \in \{20(20)200\}$ meters. Then, $\sum_{r=20}^{R} \theta_r(\overline{AnyT2_i^r})$ is our estimate of learning spillovers. For all of our specifications, the smallest radius (20 meters) minimizes the Schwarz BIC. The spillover and non-spillover samples are balanced on observable baseline variables, across the full sample (Table E.1), and within each individual treatment group (Tables E.2, E.3, E.4).

We test for the presence of learning spillover using the following specification:

$$\begin{aligned} Y_i =& \theta_0 + \theta_1 T1_i + \theta_2 T2_i + \theta_3 T3_i + \sum_{r=20}^{R} \theta_{4,r} AnyT2_i^r \\ &+ \sum_{r=20}^{R} \theta_{5,r} AnyT2_i^r \times T1_i + \sum_{r=20}^{R} \theta_{6,r} AnyT2_i^r \times T2_i + \sum_{r=20}^{R} \theta_{7,r} AnyT2_i^r \times T3_i \\ &+ Y_{i0} + X_{i0} + \gamma_{1,b} + \epsilon_i \end{aligned}$$

Then $\theta_{6,r}$ is our estimate of Info-tool to Info-tool spillovers. We use $\theta_{4,r}$, $\theta_{5,r}$ and $\theta_{7,r}$ to rule out spillovers from Info-tool onto other groups.

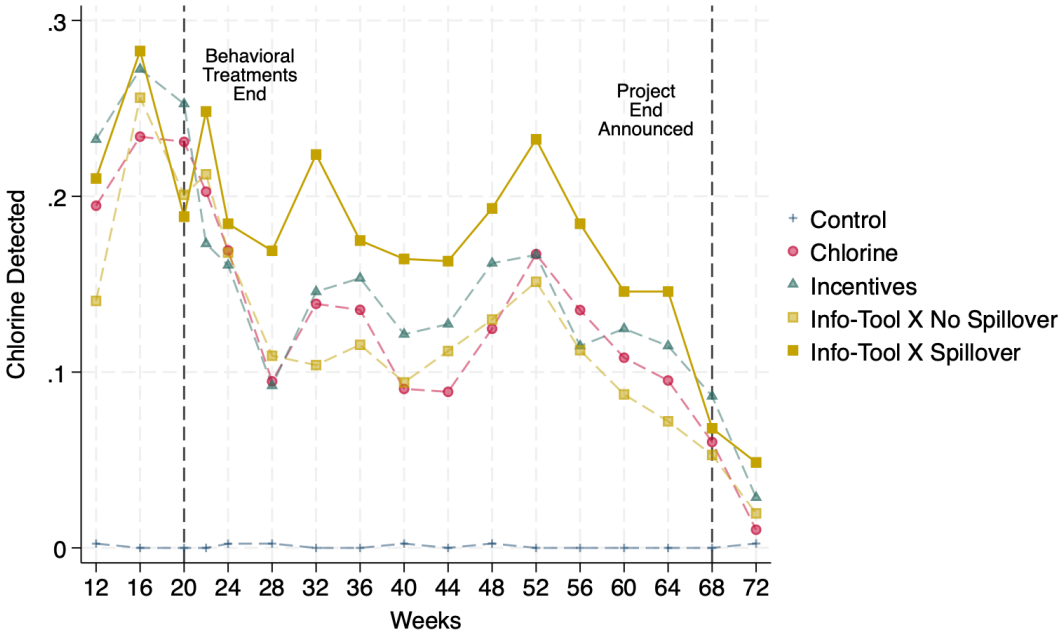
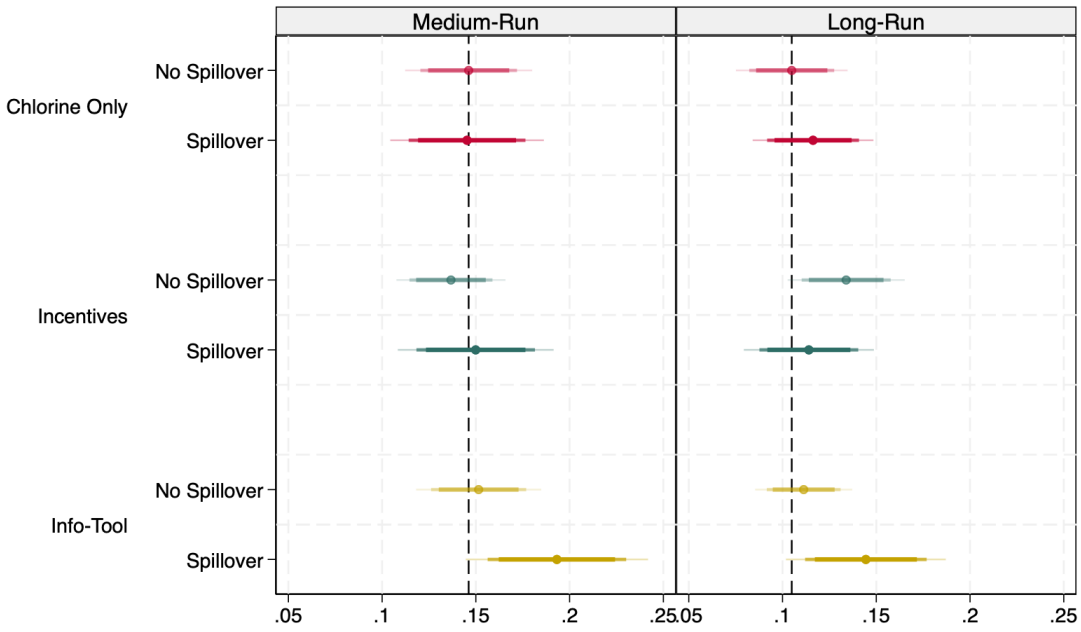


Figure 4: Raw Rates of Chlorine Detection
where "Spillover" is exposure to any Info-tool neighbor

While the presence of Info-tool participants within twenty meters does indeed increase chlorine detection, this is only true for *other Info-tool participants*. Figure 4 plots raw chlorination rates across these groups, while Table 1 and Figure 5 present the regression analogs. This pattern emerges in the medium-run (the quarter after behavioral treatments end), and persists through

the end of the study. In the medium-run, Info-tool participants with another Info-tool participant neighbor chlorinate 26% more often than participants with no Info-tool neighbor, and the difference is statistically distinguishable at conventional levels. There are no meaningful differences in use (statistically or in magnitude) by this spillover measure in either the Chlorine Only or Incentives groups. In the long-run, Info-tool participants with another Info-tool neighbor still chlorinate 30% more frequently. Again, there are no differences in the Chlorine Only or Incentives groups.

Figure 5: Regression Coefficients of Chlorine Detection (Omitted: Pure Control) where "Spillover" is exposure to any Info-tool neighbor



On aggregate across the study, Chlorine participants, Incentive participants, and Info-tool participants without another Info-tool participant in a twenty meter radius chlorinate 11.9 percent of the time – whereas Info-tool participants with at least one other Info-tool participant in a twenty-meter radius chlorinate 15 percent of the time. We conduct two tests for robustness. First, as a placebo test, we replicate our analysis to test for spillovers from Chlorine Only or Incentives participants onto their neighbors. There is no time period in which we find spillover treatment effects from Chlorine Only participants or Incentives participants onto any treatment group (Tables C.3 and C.4). Second, we show that chlorine use is constant in the number of Info-Tool neighbors among non-Info-Tool participants (Figure C.2). These placebo tests rule out alternative mechanisms driven by the diarrheal disease burden or observation of others using chlorine, strengthening our assertion that Info-Tool neighbors impact one another by sharing information about health signals.[25] Third, we reanalyze the AM data to test for

[25]Because Info-Tool participants are using chlorine at a higher rate in the medium-run, it is possible that the diarrheal burden is lower in neighborhoods with more Info-Tool participants during this time period, which Info-Tool participants are more likely to notice. However, in the short-run, Incentives are using chlorine at higher rates, so any mechanism related to aggregate chlorine use in the community should bear out in the short-run for participants with more exposure to Incentives participants. Similarly, since Info-Tool participants with

Table 1: Chlorine Detection (Household-Survey Panel)
where "Spillover" is exposure to any Info-tool neighbor

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × No Spillover	0.219***	(0.018)	0.146***	(0.013)	0.105***	(0.012)
Chlorine × Spillover	0.219***	(0.024)	0.145***	(0.016)	0.116***	(0.013)
Incentives × No Spillover	0.248***	(0.018)	0.137***	(0.011)	0.134***	(0.012)
Incentives × Spillover	0.259***	(0.024)	0.149***	(0.016)	0.114***	(0.014)
Info-Tool × No Spillover	0.202***	(0.017)	0.152***	(0.013)	0.111***	(0.010)
Info-Tool × Spillover	0.211***	(0.022)	0.192***	(0.019)	0.145***	(0.016)
Observations	4711		6354		14066	
P-values:						
Chlorine: No Spillover = Spillover	0.998		0.942		0.450	
Incentives: No Spillover = Spillover	0.726		0.521		0.232	
Info-Tool: No Spillover = Spillover	0.742		0.072		0.061	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

within-treatment Info-Tool spillover effects in Section B, and find comparable effects.

5 Mechanisms

Why does a simple pen-and-paper chart lead to changes in health behaviors? In our model, experience using the Info-tool teaches participants that signals generated with the Info-tool are trustworthy, so they place heavier weights on these signals when choosing their actions. Then, Info-tool participants are primed to learn from their own health signals while they are using the Info-Tool, and from other Info-tool participants' health signals.

We report estimates for our main model objects in Table 2: explicit knowledge or beliefs about chlorine efficacy M, action-weighted beliefs M_α, and behavioral decisions determined by $\mathbb{1}(M_\alpha > C)$. We find *no differences* in participants' belief that their children's diarrhea rate decreased after they began using chlorine (incentivized), our measure of M (we discuss this measure and its validity in detail in Section 5.2.4). However, we *do* find differences in the degree to which participants report that they are *motivated* to use chlorine to attain health for their family, our stated measure of M_α (we discuss this measure and its validity in detail in Section 5.1.2). This variation in M_α then translates into differences in chlorine adoption, which represents $\mathbb{1}(M_\alpha > C)$ (we discuss our results on chlorine adoption in Section 4).

an Info-Tool neighbor are the most likely participants to use chlorine, it is possible that these neighborhoods have an improved disease environment, and that changes in the disease environment create long-term use. If this is the case, then non-Info-Tool participants with access to *two or more* Info-Tool neighbors should respond similarly as Info-Tool participants with access to at least *one* Info-Tool neighbor.

Table 2: Model Parameters

	(1) Stated M: Diarrhea Went Down		(2) Stated M_α: Motivation: Health		(3) Observed $\mathbb{1}(M_\alpha > C)$: Times Detected Chlorine	
Chlorine × Spillover	0.049	(0.047)	-0.097	(0.093)	0.150	(0.202)
Incentives × No Spillover	0.014	(0.039)	0.053	(0.078)	0.210	(0.169)
Incentives × Spillover	0.013	(0.047)	0.054	(0.092)	0.112	(0.201)
Info-Tool × No Spillover	0.033	(0.040)	0.113	(0.078)	0.089	(0.169)
Info-Tool × Spillover	0.003	(0.047)	0.229**	(0.092)	0.587***	(0.202)
Observations	1111		1049		1116	
Chlorine × No Spillover Mean	0.711		6.081		1.543	
P-values:						
Incentives × No Spillover = Incentives × Spillover	0.987		0.986		0.632	
Info-Tool × No Spillover = Info-Tool × Spillover	0.530		0.215		0.015	

Standard errors in parentheses
Each observation is at the household level. The outcome for column (1) is an indicator for participants who said they believe that their childrens' diarrhea rate dropped after using chlorine in an incentivized elicitation about the efficacy of chlorine. The outcome for column (2) is a rating between 1 and 7 for how true the following statement felt: I use chlorine to achieve a standard of health for my family. This regression controls for the average motivation score the respondent gave across all motivation questions, and the order of questions (randomized). The outcome for column (3) is the number of times that we detected chlorine residual in the participant's water during the post-treatment period. This regression controls for the number of times we tested their water during the post-treatment period. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

5.1 Action Weights

We show two pieces of evidence that jointly suggest that participants apply higher weights to health signals when choosing to adopt chlorine tablets for water purification. First, we show that Info-tool participants' chlorine use is heterogeneous by the sign on their health signals (the expected rate of decline in diarrhea). Second, we show that the Info-Tool to Info-Tool spillover is completely explained by Info-Tool neighbors whose diarrhea rate was predicted to improve, indicating that Info-Tool participants only learn from their Info-Tool neighbors who received positive signals.

Next, we use evidence from stated beliefs to argue that Info-Tool participants apply higher weights to Info-Tool signals because they place more *trust* in these health signals; and that trust enters into the action weights on signals, rather than affecting signal uncertainty and posterior beliefs.

5.1.1 Responsiveness to Health Signals

Average Effects

First, we show that Info-tool participants who are predicted to experience improvements in their diarrhea rate during the time period where they are tracking their data – or, Info-tool participants who we expect to have a positive sign on the health signals they observe in their environment and place higher weights on – chlorinate at a higher rate immediately after the

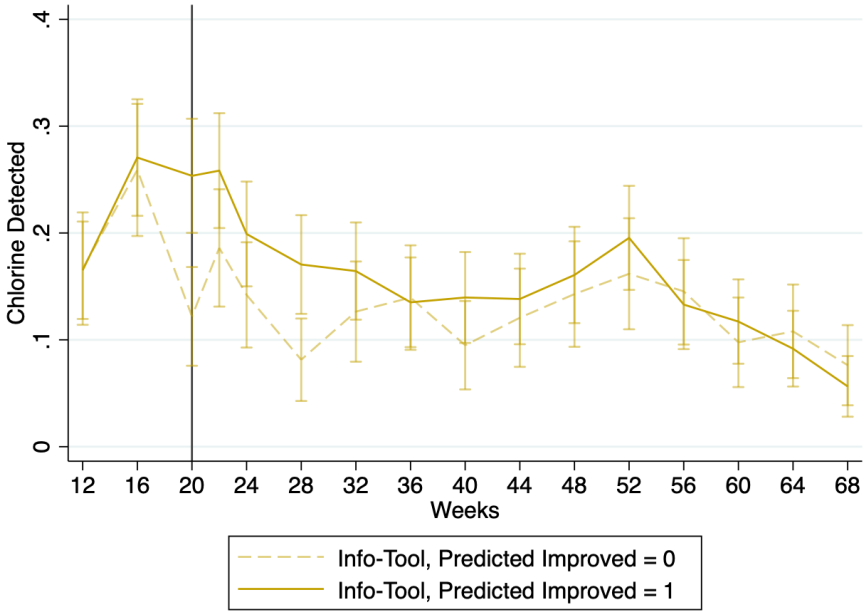


Figure 6: Info-tool Chlorine Detection by Predicted Health Improvement

Info-tool treatment ends (Table 3 and Figure 6). This effect is very valuable in the short-run: predicted-improved Info-tool participants chlorinate 29% more often than not-predicted-improved Info-tool participants in the short-run (p-value = 0.049). In the medium- and long-run, this effect becomes smaller and noisier. However, predicted-improved and not-predicted-improved Info-tool participants' chlorination rates converge not because people in the predicted-improved sample are learning less, but because people in the not-predicted-improved sample are learning more from their Info-tool neighbors, as evidenced by Figure C.5.

Actual health will be correlated with treatment adoption. To understand how treatment responds to health signals that are exogenous to treatment status, we use a predicted measure of health improvement. First, we define "health improvement" as the number of days of diarrhea in the three months before chlorine adoption minus the number of days of diarrhea in the first three months of chlorine adoption. The first three months of chlorine adoption is the relevant period of time to analyze health signals, because this is the time in which Info-tool participants receive salient information about their health. We consider someone to be "predicted improved" if the predicted diarrhea rate improvement is above the median.

We use lasso with the Pure Control sample to construct a measure of predicted health improvement (as defined above), using variables collected before treatment status was revealed to predict this measure (all variables from the baseline survey and the first half of the first follow-up survey). To avoid bias that can arise from endogenous stratification, we use the leave-one-out procedure proposed by Abadie et al. (2018).[26]

To test for individual learning from one's own health signals in the Info-tool group, we use the

[26]The correlation between "actual health improvement" and "predicted health improvement" is 0.39 in the Pure Control group, and 0.40 in the treatment groups.

Table 3: Chlorine Detection (Household-Survey Panel)
by predicted health improvement
Omitted group: Chlorine × Not Improved

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × Improved	0.033	(0.029)	0.035*	(0.021)	0.013	(0.016)
Incentives × Not Improved	0.045	(0.028)	0.002	(0.018)	0.021	(0.015)
Incentives × Improved	0.058**	(0.029)	0.019	(0.019)	0.027*	(0.016)
Info-Tool × Not Improved	-0.019	(0.026)	0.020	(0.019)	0.013	(0.015)
Info-Tool × Improved	0.034	(0.028)	0.052**	(0.022)	0.030**	(0.015)
Observations	3463		4689		10472	
P-values:						
Incentives: Not Improved = Improved	0.659		0.365		0.717	
Info-Tool: Not Improved = Improved	0.049		0.134		0.277	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). Improved is a binary indicator for if the participant's predicted improvement in health after the beginning of chlorine distribution was above the median. See Section 6 for a detailed explanation for how the predicted-health-improvement measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

following specification:

$$\begin{aligned} Y_i =& \theta_0 + \theta_1 T1_i + \theta_2 T2_i + \theta_3 T3_i + \theta_4 \hat{I}_i \\ &+ \theta_5 \hat{I}_i \times T1_i + \theta_6 \hat{I}_i \times T2_i + \theta_7 \hat{I}_i \times T3_i \\ &+ Y_{i0} + X_{i0} + \gamma_{1,b} + \epsilon_i \end{aligned}$$

where $\hat{I}_i = 1$ if an individual was predicted to improve (continuous predicted health improvement is above the median). Then θ_6 is our object of interest. We use θ_4, θ_5, and θ_7 to test if there are differences in long-term use between those predicted to improve and those not predicted to improve when the information is not made salient.

We find that Info-tool participants whose health was predicted to improve in the first three months of chlorine distribution use chlorine at a higher rate than Info-tool participants whose health was not predicted to improve. In the short run, predicted-improved Info-tool participants use chlorine 29% more often than Info-tool participants who are not predicted to improve. The two groups only diverge in their chlorine use in the third month, by which time they would have received and been able to respond to health signals that the Info-tool makes salient. There are no differences between predicted-improved and not-predicted-improved participants in any other treatment group in the short-run.

In the quarter immediately after behavioral treatments end, the predicted-improved Info-tool respondents continue using chlorine at a higher rate than not-predicted-improved Info-tool participants (23% higher rate of use, Table 3). However, we are not able to rule out these

differences at traditional thresholds of statistical significance (p-value = 0.134). [27]

Spillovers

If Info-tool participants are able to learn from their Info-tool neighbors' information about the efficacy of chlorine to a greater degree than other participants are, then the spillover should be explained by cases where information shared between two participants is likely about chlorine being highly efficacious. To test if the spillover treatment effect can be explained by positive signals from neighboring Info-tool participants we limit $AnyT2_i^r$ to only be equal to 1 if there is any Info-tool participant within radius r of participant i *who was predicted to see their health improve*. When we change the definition of the spillover from *any Info-tool neighbor* to *any predicted-improved Info-tool neighbor* the spillover estimates remain almost unchanged (comparing estimates in Table 4 with estimates in Table 1, estimates are *at least as big* in magnitude to the spillover treatment effect with the broader definition of spillovers).[28]

It is possible that participants who are predicted to improve have characteristics that make them more likely to influence their neighbors, regardless of treatment status or whether they share information that can be interpreted as a positive signal about chlorine. To rule out this possibility, we conduct a placebo test where we test heterogeneity by the total number of predicted-improved treated neighbors. In this placebo test, there are no differences in use by exposure to this spillover measure in the short-run or medium-run, in any group (Table C.5).

5.1.2 Stated Beliefs

We ask respondents several questions to interrogate three ways in which the Info-tool could have led participants to apply higher weights on health signals and increase their use of chlorine: (1) Info-tool increases the probability that participants believe that chlorine is effective; (2) Info-tool increases the probability that participants believe that their disease environment poses a risk to children's health; and (3) Info-tool increases the probability that participants act on their beliefs, without affecting beliefs themselves. We find evidence contrary to mechanisms (1)

[27] In this quarter, differences in use by predicted-improvement in the Chlorine Only group also emerge. This could be explained by some learning that happens through observation without the Info-tool, or by selection. For example, perhaps the characteristics of people who are predicted to improve also incline them more towards longer-term use without any learning. While these patterns fade in both groups in the long-run, the magnitude of the difference remains slightly larger in the Info-tool group.

[28] In the long-run, a *reverse* pattern emerges in the Incentives group, where people without access to a predicted-improved treated neighbor use chlorine at a higher rate. While we do not have a prediction for this result, these participants were more likely to report at endline that they had shared their tablets with others. Incentives participants in the spillover sample are in an environment where many others have a higher demand for chlorine than they do. We do not have any evidence as to why the Incentives group is more responsive to this higher aggregate demand than the Chlorine Only group, but we speculate that it could be due to two possibilities. Firstly, Incentives participants are used to receiving income for their chlorine tablets, so they might have sold their tablets. Only one person reports that they sold chlorine tablets at endline (the rest say they shared them for free), but it could be that participants do not wish to admit they sold tablets or forgot. Secondly, Incentives participants may wish to share chlorine tablets out of equity concerns. Although the participants not in the Incentives group did receive gifts via a lottery, it could be that a feeling of inequity arises between the Incentives group and other households because they were chosen for a group where receipt of gifts was within the caregivers' control and explicitly linked to the treatment assignment.

Table 4: Chlorine Detection (Household-Survey Panel)
where "Spillover" is exposure to any Info-tool neighbor predicted to improve
Omitted group: Chlorine × No Spillover

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × Spillover	-0.037	(0.033)	0.010	(0.024)	0.022	(0.017)
Incentives × No Spillover	0.039*	(0.023)	-0.006	(0.015)	0.030**	(0.013)
Incentives × Spillover	-0.012	(0.032)	-0.003	(0.023)	-0.000	(0.017)
Info-Tool × No Spillover	-0.020	(0.022)	0.011	(0.016)	0.012	(0.012)
Info-Tool × Spillover	-0.011	(0.031)	0.051*	(0.027)	0.047**	(0.021)
Observations	3463		4689		10472	
P-values:						
Incentives: No Spillover = Spillover	0.114		0.883		0.074	
Info-Tool: No Spillover = Spillover	0.781		0.134		0.092	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) whose health was predicted to improve than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined. See Section 6 for a detailed explanation for how the predicted-health-improvement measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

and (2), and supportive evidence of mechanism (3).

Motivation

Although at endline households in all treatment groups non-differentially report that they believe chlorine to be an effective technology, and unclean water to be a source of sickness, Info-tool participants are more likely to report that attempting to achieve health for the family is the *motivating reason* for why they use chlorine (Table 5).[29] Info-tool households rated "to attain a certain level of health for my family" as a stronger motivating factor for using chlorine, controlling for their ratings of other potential motivating factors. Although almost everyone believes that chlorine is an effective technology, and that they live in a risky disease environment where that technology is needed, the Info-tool gives people the motivation to act upon this information. Thus, our evidence suggests that Info-tool influences action-weighted beliefs, rather than beliefs alone.

[29] At endline, we asked caregivers a series of questions to understand their motivation behind using chlorine (question formulation adapted from Tremblay et al. (2009)). We read several statements describing reasons why someone might use chlorine, and asked women to rate on a scale from 1 to 7 how true that statement was for them during the times when they have ever used chlorine. The statements described a motivation related to health ("Because using treatments like chlorine tablets help me to attain a certain level of health for my family"), habit formation ("Because it has become a fundamental part of my routine"), income generation ("Because it could allow me to earn money"), and intrinsic motivation ("Because I derive pleasure from trying new things", "Because I want to be very good at taking care of my family, otherwise I would be very disappointed", and "For the satisfaction I experience when I am successful at doing difficult tasks").

Our model predicts that action-weighted beliefs should update more among Info-tool participants who receive signals about the efficacy of chlorine from other Info-tool participants, but that participants in other groups should not respond to exposure to Info-tool neighbors. Indeed, we see that the treatment effect on the motivation score that participants give "health" is twice as large in the spillover sample in the Info-tool group, as compared to Info-tool participants who are not in the spillover sample (Column 2, Table 2). These treatment effects are not statistically distinguishable and therefore are not conclusive on their own, but it is encouraging that the patterns we see are aligned with our model predictions. Chlorine Only and Incentives participants do not rate health as a motivating factor differentially by access to neighbors from the Info-tool group.

Trust

Why does Info-tool motivate people to act upon something that, on the surface, everybody knows? To understand if trust is the key mechanism, we reminded all participants about the three treatment groups in our study: people receiving chlorine only, people receiving chlorine as well as gifts in exchange for using chlorine, and people receiving chlorine as well as a pen and paper chart with which they tracked their children's diarrhea rates. We asked participants to imagine a neighbor in each of these three groups and asked: who among these three neighbors do you think knows more about children's health? Differencing out the propensity to choose one's own group using Chlorine-Only households, participants in the Info-tool group were 13.3 percentage points ($p = 0.004$) more likely to list the Info-tool respondent (Table 6).[30] While all participants are able to repeat the information that they have received throughout the study – that their children are exposed to unclean water which makes them sick, and chlorine tablets can purify their water and make their children healthier – the Info-tool makes this information trustworthy and actionable.

Our model suggests that experience with the Info-Tool alone, regardless of exposure to other Info-Tool participants, generates more trust in signals generated with the Info-Tool. Indeed, we find that trust in other Info-Tool participants' information about child health is higher than other groups regardless of access to an Info-tool spillover – in fact, Info-Tool participants *without* access to an Info-tool spillover are slightly more likely to rate other Info-tool participants as the most knowledgeable (61% increase, $p = .094$) – indicating that personal experience with the Info-tool itself is enough to change participants' trust in Info-tool as an information source (Table C.6). Then, Info-tool participants are uniquely primed to trust health signals from other Info-tool participants. Only when an individual has access to these signals, by being randomly assigned an Info-tool neighbor, do they receive new trust-worthy and *actionable* information that leads their behaviors to change.

[30]The corresponding estimate is 10 percentage points ($p = 0.046$) if we use the Incentive households to net out the propensity to choose one's own group.

Table 5: Endline Stated Motivations
Omitted group: Chlorine Only

	(1)		(2)		(3)		(4)	
	Motivation: Health		Motivation: Habit		Motivation: Money		Motivation: Intrinsic	
Incentives	0.088	(0.062)	-0.031	(0.064)	-0.020	(0.063)	-0.023	(0.032)
Info-Tool	0.191***	(0.062)	-0.049	(0.064)	-0.068	(0.062)	-0.027	(0.033)
Observations	1049		1057		1051		1058	
P-values:								
Incentives = Info-Tool	0.094		0.769		0.444		0.880	

Standard errors in parentheses
Each observation is at the household level. The outcome for column (1) is a rating between 1 and 7 for how true the following statement felt: I use chlorine to achieve a standard of health for my family. The outcome for column (2) is a rating between 1 and 7 for how true the following statement felt: Because it has become a fundamental part of my routine. The outcome for column (3) is a rating between 1 and 7 for how true the following statement felt: Because it could allow me to earn money. The outcome for column (4) is the average of the ratings between 1 and 7 for how true each of the following statements felt: Because I derive pleasure from trying new things; Because I want to be very good at taking care of my family, otherwise I would be very disappointed; and For the satisfaction I experience when I am successful at doing difficult tasks. All regressions control for the average motivation score the respondent gave across all motivation questions, the order of questions (randomized), neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Endline Stated Trust
Omitted group: Pure Control

	(1)	(2)	(3)
	Info-Tool Knows Most	Incentives Knows Most	Chlorine Knows Most
Chlorine Only	-0.043	-0.079**	0.122***
	(0.027)	(0.031)	(0.034)
Incentives	-0.004	0.088***	-0.085**
	(0.027)	(0.031)	(0.034)
Info-Tool	0.222***	-0.083***	-0.138***
	(0.027)	(0.031)	(0.034)
Observations	1516	1516	1516
Control Mean	0.196	0.276	0.527
DID Estimate:			
Info-Tool picks Info-Tool			
– Other Group picks Own Group		0.133 [p= 0.004]	0.100 [p= 0.046]

Standard errors in parentheses
Each observation is at the household level. The outcome for column (1) is an indicator for if the respondent chose a hypothetical Info-tool respondent as the person most likely to be knowledgeable about child health (rather than a Chlorine Only or Incentives participant). The outcome for column (2) is an indicator for if the respondent chose a hypothetical Incentives respondent as the person most likely to be knowledgeable about child health (rather than a Chlorine Only or Info-Tool participant). The outcome for column (3) is an indicator for if the respondent chose a hypothetical Chlorine Only respondent as the person most likely to be knowledgeable about child health (rather than an Incentives or Info-Tool participant). All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

5.2 Competing Mechanisms

5.2.1 Early Adoption

Alternatively, the Info-tool might lead participants to adopt chlorine early on at a higher rate than they would without the Info-tool. Long-term adoption could then be explained by early adoption leading to habit formation, or intertemporal complementarities in chlorine use. We use the Incentives arm to rule out stories of chlorine use related to higher rates of early adoption.

Our Incentives arm builds on Hussam et al. (2022), in which persistence in handwashing behavior is engendered through exogenous short-run financial incentives to handwash in West Bengal.[31] Our Incentives arm serves as a parallel intervention: households are incentivized to chlorinate their water daily, receiving tokens that can be exchanged for household goods for each day of empty chlorine wrappers they present to enumerators. Should water chlorination, which is a repeated act performed at the same time and place each day, indeed be habit forming, then this exogenous increase in initial consumption stock via financial incentives will activate the intertemporal complementarities in use, thereby generating long-run use even after incentives are withdrawn. Alternatively, a higher rate of use may lead households to acquire more signals in a shorter period of time than in the other two groups, leading participants to learn from their own experience how the tablets improve their children's health. We do not attempt to distinguish between habit formation and learning through rapid accumulation of health signals in the Incentives arm. Instead, we use the Incentives arm to rule out either of these two mechanisms relative to the role of salient signals.

We observe that Incentives arm households use chlorine at a higher rate than either of the Chlorine or Into-Tool arms during the behavioral treatment period. In order to ensure that Incentives households are actually chlorinating at a higher rate, rather than chlorinating just on the days where the enumerator is present, we do not directly incentivize the water chlorination test (household incentives are tied to presenting empty chlorine tablet wrappers), and we conduct unscheduled audit tests. In these audit tests, we detected chlorine at a higher rate in Incentives than in Info-tool. Furthermore, we detected chlorine at a much higher rate during the audit visits than during the regularly scheduled visits (45% detection rate in the Incentives arm during audit visits), further suggesting that Incentives participants were not more likely to use chlorine when they expected that we would come to test their water than at other times.

Since the Incentives arm built up a higher stock of chlorine use in the first three months of chlorine distribution, any theory of sustained behavioral change relating to early adoption will be born out in the long term in the Incentives arm. In the three months immediately after the behavioral interventions ended, the Incentives arm reverted to the same rate of use as the Chlorine Only group, while the Info-tool group chlorinated 18% more often in the medium-run, the quarter after behavioral interventions ended, than the Incentives group ($p = 0.055$; we

[31] Hussam et al. (2022) finds that households who experienced larger health improvements (either across weeks or in aggregate) from the intervention did not exhibit differentially greater persistence in handwashing, and therefore attribute the long run behavior change to habit formation rather than learning.

cannot, however, rule out equality between the Info-Tool group and the Chlorine Only group during this time period with $p = 0.142$). Thus, the mechanism that helped the Info-tool sustain chlorine adoption in the months following the behavioral interventions cannot be explained by habit formation or any story related to early adoption.[32] For a more detailed discussion on the roles of learning and habit formation, see Section I.2.

5.2.2 More Interactions and Mimickry

It is possible that the novelty of the Info-tool leads participants to interact with one another more when they are both in this group, and that they simply mimic one another's behaviors rather than learn from their information. For example, perhaps they assist one another in filling out the chart. However, we do not have any evidence to suggest that participants interacted in this way, with only 12% of Info-tool respondents reporting that they ever discussed the Info-tool with anybody else. Info-tool participants' endline social networks were not larger than any other groups' (Table C.7). Furthermore, we have no evidence of Info-tool participants mimicking one another on other behaviors, including take-up of other health programs that were present during our study trial, or choice of household savings technology (Table C.10).

5.2.3 More Conversations

We asked participants if they ever discussed health or water purification with members of their social network. On average, participants discussed health with one person, and this did not differ across any treatment group. Control group participants were much less likely to discuss water purification than treated participants, but Info-tool participants were no more likely to discuss water purification with their social network connections than the other two groups. If anything, the Incentives group participants were the most likely to discuss water purification with others (Table C.7).

5.2.4 Signal Uncertainty

In an alternate model, the Info-Tool acts as an education tool that increases participants' ability to comprehend signals that they observe or hear. Although this alternate model is attractive in its simplicity, since this mechanism could alone explain the complementarity result, the data do not support it. In this model, the Info-Tool first facilitates participants in comprehending the signals that they observe, leading them to update their beliefs whereas the participants in other treatment groups do not; then, because Info-Tool participants' beliefs now differ from the beliefs of participants in other groups, the information they share about chlorine efficacy

[32] Average rates of chlorine use in the Incentives and Info-tool groups converge at 32 weeks. Notably, the Info-Tool participants in the Spillover sample continue to chlorinate at a higher rate. This could be due to forgetting in the Info-tool arm without an Info-Tool neighbor, or it could be due to some catching-up in the Incentives group (cyclical habit formation, learning over a different time horizon in the absence of salient signals, etc.). We do not attempt to disentangle potential reasons why Incentives and Info-tool eventually converge on average because, ultimately, the higher rate of use that Info-tool experienced, though temporary, resulted in greater child health improvements. Furthermore, the Info-Tool group in the Spillover sample continued to chlorinate at a higher rate after the two groups converged on average.

Table 7: Midline Memory/Beliefs about Chlorine Efficacy

	(1) Correct Answer	(2) Went Down	(3) Didn't Change	(4) Went Up	(5) Don't Know
Chlorine Only	-0.006	0.069**	-0.039**	-0.005	-0.023**
	(0.035)	(0.031)	(0.019)	(0.027)	(0.009)
Incentives	0.034	0.064**	-0.036*	0.001	-0.027***
	(0.035)	(0.031)	(0.019)	(0.027)	(0.009)
Info-Tool	0.043	0.073**	-0.016	-0.030	-0.026***
	(0.035)	(0.031)	(0.019)	(0.027)	(0.009)
Observations	1494	1494	1494	1494	1494
Control Mean	0.503	0.651	0.108	0.206	0.031
P-values:					
Incentives = Chlorine Only	0.251	0.875	0.883	0.833	0.697
Info-Tool = Chlorine Only	0.155	0.891	0.218	0.367	0.743
Info-Tool = Incentives	0.781	0.768	0.277	0.264	0.952

Standard errors in parentheses
Each observation is at the household level. This regression includes neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

also differs. Because Info-Tool participants gained skills in comprehension from the treatment in this model, they are also uniquely enabled to understand the signals from their Info-Tool neighbors, generating the complementarity that we observe.

If this is the true model that explains our results, the following conditions would need to be true: (1) the Info-Tool led participants to better comprehend the signals about chlorine that they observed; and (2) the Info-Tool participants shared different information than participants in the other treatment groups.

No Evidence of Differential Comprehension

Halfway through Phase 2 (the week 44 survey), we conducted an incentivized belief elicitation activity. We simply asked participants to tell us whether their children's diarrhea rates increased, decreased, or remained the same during the three months following the start of chlorine distribution, compared to the three months prior. If participants answered correctly, they received a gift (children's goods such as pencils or a notepad). If the Info-Tool improves signal comprehension, then participants should be more likely to answer the question correctly. However, there were no differences in correctly answering the question across any groups (Table 7). There were also no differences in which answer the participants gave across the three groups that received chlorine, indicating non-differential knowledge or optimism about chlorine efficacy.[33]

To ascertain that answering the question correctly is sensitive to comprehension skills, we test if response accuracy is differential by education. Reassuringly, participants in any treatment group who had ever been to school (sixty-four percent of the sample had received zero years of education) were 5.9 percentage points (12%, $p = 0.061$) more likely to answer the question

[33]Control participants were less likely to say that their diarrhea rate had improved than the treatment groups (accurately so), but there were no differences between any groups that received chlorine.

Table 8: Midline Memory/Beliefs about Chlorine Efficacy: By Education

	(1) Correct Answer	(2) Went Down	(3) Didn't Change	(4) Went Up	(5) Don't Know
Any Education	-0.128**	0.017	0.008	-0.050	0.033**
	(0.053)	(0.048)	(0.029)	(0.042)	(0.014)
Any Treatment Group	-0.047	0.069**	-0.024	-0.032	-0.010
	(0.036)	(0.033)	(0.020)	(0.029)	(0.009)
Any Treatment Group × Any Education	0.187***	0.019	-0.033	0.049	-0.041***
	(0.061)	(0.055)	(0.033)	(0.048)	(0.016)
Observations	1427	1427	1427	1427	1427
Control Mean (No Education)	0.551	0.650	0.115	0.214	0.019
P-values:					
Any Education + Any Treatment Group × Any Education = 0	0.063	0.213	0.154	0.968	0.363

Standard errors in parentheses
Each observation is at the household level. This regression includes neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

correctly (Table 8).[34] Interestingly, Control participants with any education were *less* likely to answer the question correctly, partially explained because educated Control participants were more likely to answer "I don't know" than any other group. This suggests that access to technological experimentation alone *does* lead participants to increase the attention they pay to signals from their observed environment, especially among people with education. Jointly, this set of results demonstrates that the Info-Tool does not actually increase the accuracy of participants' knowledge or the optimism of their beliefs about chlorine efficacy. At endline, we asked several more non-incentivized questions about beliefs in chlorine efficacy and the riskiness of the disease environment, none of which yielded differential responses across treatment groups.[35]

No Evidence of Differential Information Sharing

We can rule out that Info-Tool participants talked to more people, or discussed water purification or child health with more people (see Section 5.2.2 and Section 5.2.3). Since we do not observe participants' conversations, we cannot determine exactly *what* they say about water purification or child health in these conversations. However, it is encouraging that Info-Tool participants were no more likely to answer positively or accurately about chlorine efficacy, including when correct answers are incentivized. Furthermore, we asked a few questions at

[34]The education treatment effect was non-differential across treated arms among participants who received any chlorine treatment.

[35]At endline, we find no differences across treatment groups in beliefs about how many child-days of diarrhea a hypothetical household would experience after using chlorine relative to before using chlorine; no differences in the number of child-days of diarrhea participants believe their household would experience in the absence of chlorine, in either the summer or winter seasons; no differences in reporting "unclean water" as a primary cause of illness in the household prior to using chlorine tablets (unprompted); and no differences in how high they rank unclean water as a primary source of child illness prior to using chlorine tablets, relative to other potential sources of illness (prompted). It is possible that beliefs updated immediately after the Info-tool treatment but then converged with time, with social learning superseding individual learning, such that we do not observe differences in beliefs at endline, so we take these null results as consistent with our results but not conclusive.

endline to understand the nature of participants' conversations about water purification, and there are no differences across treatment groups.

We asked participants to guess the water purification method of each member of their social network. Using each participant's name, nickname, husband's name, and neighborhood block, we are able to link social network nodes with participants in our sample. If Info-Tool participants are sharing information about chlorine tablets that is more favorable towards chlorine than other groups, we might expect participants to believe that *their* Info-Tool friends are chlorine users. We see no differences in participants' beliefs about their treated friends' chlorine tablet use (Table 9). Interestingly, participants *are* more likely to believe that their friends in *any* treatment group are more likely to use chlorine than friends in the Control group or outside the study sample. This implies that participants have some awareness of who uses chlorine tablets, and that this question captures relevant information about participants' propensity to share information about the water treatment methods they use.

It is possible that some Info-Tool participants share especially optimistic information about chlorine tablets, and some share especially pessimistic information. Then, these two forces might counteract each other so that participants believe their Info-Tool and other-treatment-group friends use chlorine at the same rates on average. If this is the case, then participants should have more *accurate* information about Info-Tool participants chlorine use, even if not more optimistic. Using our participants' self-reported water purification methods at endline, we can determine if participants accurately guessed the water purification method of their within-sample network connections. We find no evidence that participants, regardless of their treatment group, have a better understanding of their Info-Tool friends' water purification methods than the purification methods that any of their other treatment-group friends use (Table C.11).[36] Furthermore, Info-Tool participants do not do a better job guessing than any other group.

Finally, we also asked participants how long ago they first and last discussed water purification with each social network connection. Info-Tool participants did not begin engaging in conversations about water purification earlier than any of the other groups (Chlorine began engaging in these conversations the earliest), nor did they have a conversation more recently (Incentives had the most recent water purification discussions) (Table C.12).

This evidence supports our model, where treated participants are all equally aware of the benefits of chlorine tablet water purification, and share this information non-differentially in frequency or substance. Trust in signals acquired with the Info-Tool makes the belief in chlorine's efficacy actionable.

[36]The same holds true when we use other measures of "true treatment method" using our objective measures of water chlorine detection.

Table 9: Endline Beliefs About Friends' Chlorine Use

	(1) Believes Uses Chlorine: Full Sample		(2) Believes Uses Chlorine: Info-Tool Only	
Control Friend	-0.010	(0.038)	0.065	(0.067)
Chlorine Only Friend	0.086**	(0.041)	0.137	(0.087)
Incentives Friend	0.013	(0.032)	0.016	(0.068)
Info-Tool Friend	0.035	(0.046)	0.093	(0.089)
Observations	2614		617	
Non-Participant Sample Mean	0.085		0.052	
P-values:				
Info-Tool Friend = Chlorine Friend	0.281		0.713	

Standard errors in parentheses
Standard errors are clustered at the household level. Each observation is at the network-link level, for observations where the node is within our sample. This regression include neighborhood block fixed effects, unbalanced baseline controls, the number of study participants within twenty meters, lasso-selected baseline controls, and participant fixed effects. The outcome in column (1) is an indicator for if the participant's guess about if her friend uses chlorine aligns with what that friend reported using in the endline survey. The outcome in column (2) is an indicator for if the participant's guess about if her friend uses chlorine aligns with what we objectively observed (did we ever detect chlorine in the friend's water). The outcome in column (3) is an indicator for if the participant guessed that her friend uses chlorine tablets for water purification, and we detected chlorine in that participant's water at least three times.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

5.2.5 Social Norms

It is possible that participants act on what they think *others* believe, via a preference to conform to social norms, rather than their own beliefs. We asked participants at endline if they believe a guest at their home would drink chlorinated water if they were to serve it. While the majority of the Control group believed a guest would accept chlorinated water (67%), there was a large effect of being in any treatment group on this measure (Table C.13). However, the Info-tool group was no more likely to believe a guest would accept water than the other two groups. If anything, Incentives participants were the most likely to believe a guest would accept chlorinated water. We also asked participants to report to the best of their ability how each of their network connections purifies their water. Control group participants were very unlikely to believe that their network connections purified their water with chlorine (6.4% believed one network connection purified their water with chlorine), but the rate was 92.9% higher for participants in any treated group. Again, Info-tool was no more likely to believe their network connections purified water than any of the other treated groups (Table C.13).

6 Child Health

6.1 Diarrhea

We find substantial impacts of water chlorination on child health. Diarrhea rates over time offer a lens into how the various interventions impact children's health dynamically over the course of the experiment. Treatment group participants report 36% (Chlorine Only), 26% (Info-Tool), and 38% (Incentives) reductions in diarrhea in the short-run relative to the Pure Control group (Table C.2). These relative reductions mirror relative rates of chlorine use during this time period. Reductions in diarrhea rates continue to mirror relative levels of chlorine use

in the medium- and long-run, with the Incentives groups experiencing the smallest reductions in diarrhea prevalence (11% reduction in the medium-run and 18% reduction in the long-run, neither statistically distinguishable from the Pure Control group); and the Info-Tool group experiencing the largest reductions in diarrhea prevalence (38% reduction in the medium-run, and 33% reduction in the long-run, both statistically distinguishable from the Pure Control group at conventional levels of significance). More details on measurement and results on child diarrhea are reported in Section K.

As a self-reported measure, children's diarrhea rate is subject to respondent bias; that it is directly connected to the substance of the Info-tool also leaves room for concerns of endogeneity in outcome reporting. As an objective and therefore our preferred measure of health, we measure child anthropometrics at endline in an index that combines four standardized measures: weight-for-age (WAZ), weight-for-height (WHZ), height-for-age (HAZ), and MUAC-for-age (mid-upper arm circumference-for-age) z-scores.

6.2 Child Anthropometrics

6.2.1 ITT Results: Any Treatment

To analyze the impact of our treatment on child anthropometrics, we use the following specification:

$$Y_c, h = \beta_0 + \beta_1 T_h + X_{h0} + \gamma_b + \epsilon_{c,h}$$

where $Y_{c,h}$ is health outcome Y for child c in household h, T_h is treatment status of household h, and X_{h0} are household-level baseline control variables. We cluster standard errors at the household level for all child-level specifications.

Table 10 reports the ITT effects of *any* treatment on the various measures of child anthropometry, whereas Table 12 reports the treatment-specific estimates. Following Anderson (2008), we create a summary index combining height-for-age, weight-for-height, weight-for-age, and MUAC-for-age z-scores. For households in any of the three treatment arms, the index increases by over 7% of a standard deviation ($p < 0.05$).[37]

6.2.2 IV Results: Effects of Water Purification

Treatment compliance was not perfect – we only were able to detect chlorine in the water of our treated participants 13% of the time across the whole study, and 47% of treatment participants said that they were using chlorine to treat their water at endline. To understand the effect of purified water on child health, we use an instrumental variables specification. We consider compliers to be any individual who at endline reported that they boil, bleach, or chlorinate their

[37]For households receiving any of the three treatments, the estimates for height-for-age and weight-for-height are close to zero whereas those for weight-for-age and MUAC-for-age are 0.112 ($p < 0.1$) and 0.055 (not significant).

Table 10: ITT Impacts on Endline Child Health (Pooled Treatments)

	(1) Index (Anthro-pometry)	(2) Height-for-Age	(3) Weight-for-Height	(4) Weight-for-Age	(5) MUAC-for-Age
Any Treatment	0.074**	0.016	-0.005	0.112*	0.055
	(0.034)	(0.066)	(0.081)	(0.065)	(0.054)
Observations	2616	2371	2439	2492	1954
Endline Control Mean	-0.019	-1.773	-0.291	-1.407	-1.453

Standard errors in parentheses
Any treatment is an indicator for not being in the Control group (ever receiving chlorine). Child-level cross-section of the endline survey. Standard errors are clustered at the household level. All regressions include baseline unbalanced household-level controls, baseline child anthropometrics and diarrhea rates, child gender, child age, neighborhood block fixed effects, and lasso-selected controls. Indeces are constructed using following Anderson (2008).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: IV Impacts on Endline Child Health (Instrument: Any Chlorine Treatment)

	(1) Index (Anthro-pometry)	(2) Height-for-Age	(3) Weight-for-Height	(4) Weight-for-Age	(5) MUAC-for-Age
Boils, Bleaches, or Chlorinates Water	0.240**	0.050	-0.017	0.371*	0.198
	(0.112)	(0.207)	(0.260)	(0.212)	(0.184)
Observations	2616	2371	2439	2492	1954
Endline Control Mean	-0.019	-1.773	-0.291	-1.407	-1.453
Weak-IV robust F statistic	119.93	127.97	121.36	122.46	95.19
C-statistic p-value	0.027	0.491	0.392	0.139	0.056

Standard errors in parentheses
Any treatment (i.e., received chlorine) is an instrument for if the respondent reported at endline that she boils, bleaches, or chlorinates her water. Child-level cross-section of the endline survey. Standard errors are clustered at the household level. All regressions include baseline unbalanced household-level controls, baseline child anthropometrics and diarrhea rates, child gender, child age, and neighborhood block fixed effects. Indeces are constructed using following Anderson (2008).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

water. While this is a self-reported measure of compliance, we think this is a better measure of purified water for two reasons: (1) our objective measure of chlorine residual detection is an underestimate and changes with time, and (2) participants may have learned about the negative effects of impure water through the study and changed their water treatment behavior in some way *other* than chlorinating their water. Indeed, we find that Control participants are more likely to boil their water at endline. If we consider treatment assignment to be an instrument for using chlorine to purify water, then the exclusion restriction does not hold because the treatment also affects participant's use of other water-treatment methods. If we instead consider treatment to be an instrument for using any effective water purification technique, which includes boiling, then higher rates of boiling in the Control group is not a threat to the exclusion restriction and the Control participants who boil their water can be considered always-takers.

Table 11 reports the results of an IV exercise where we use the assignment to any treatment as an instrument for whether the household reported to be boiling, bleaching, or chlorinating water at endline. Forty-eight percent of the sample report using one of these effective water purification technologies at endline. We find that effective water purification leads to a 0.24 SD increase in the index of child anthropometrics ($p < 0.05$).

6.3 Heterogeneity by Treatment

6.3.1 ITT Results

Table 12 shows that the improvements in anthropometric outcomes are consistently higher for households in the Info-tool group. Furthermore, the treatment effects are more than twice the magnitude among Info-tool households who have another Info-tool neighbor, relative to Info-tool households without access to this spillover (Table 13). Comparing only the spillover-sample Info-tool participants with the rest of the sample, there was a 0.12 SD increase in the index of child anthropometry ($p < 0.05$).

6.3.2 IV Results

The IV results may also be heterogeneous with respect to the specific treatment group and access to an Info-tool neighbor. Our instrumented measure (self-reported use of boiling, bleach, or chlorine for water purification) is self-reported and is an extensive margin measure of using an effective water purification method, which does not imply perfect compliance.[38] Furthermore, even with perfect compliance, individuals who use effective water purification methods at endline may have purified their water at different rates throughout the study; since health is a stock, this historical use matters. However, we cannot use a cumulative measure, such as the number of visits in which we detected chlorine in participants' water, because these measures will not be excludable if the treatments changed water purification methods aside from chlorinating, for which we don't have cumulative measures.

[38] Among the people who reported that they chlorinate their water at endline, we only detected chlorine in the water of 3.8%.

Table 12: ITT Impacts on Endline Child Health (Separated by Treatment)

	(1) Index (Anthro-pometry)	(2) Height-for-Age	(3) Weight-for-Height	(4) Weight-for-Age	(5) MUAC-for-Age
Chlorine-Only	0.081**	0.020	0.007	0.080	0.025
	(0.039)	(0.084)	(0.099)	(0.081)	(0.064)
Incentives	0.030	0.022	-0.098	0.065	0.061
	(0.042)	(0.078)	(0.101)	(0.079)	(0.066)
Info-Tool	0.108***	0.006	0.072	0.187**	0.080
	(0.039)	(0.080)	(0.101)	(0.080)	(0.066)
Observations	2616	2371	2439	2492	1954
Endline Control Mean	-0.019	-1.773	-0.291	-1.407	-1.453
P-values:					
Chlorine = Incentives	0.186	0.979	0.310	0.853	0.572
Chlorine = Info-Tool	0.455	0.867	0.523	0.193	0.386
Incentives = Info-Tool	0.044	0.836	0.104	0.130	0.761

Standard errors in parentheses
Child-level cross-section of the endline survey. Standard errors are clustered at the household level. All regressions include baseline unbalanced household-level controls, baseline child anthropometrics and diarrhea rates, child gender, child age, neighborhood block fixed effects, and lasso-selected controls. Indeces are constructed using following Anderson (2008).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Child Index of Anthropometry (ITT):
where "Spillover" is exposure to any Info-tool neighbor

	(1) Ommitted Group: Pure Control		(2) Ommitted Group: Everyone Else	
Chlorine × No Spillover	0.097**	(0.045)		
Chlorine × Spillover	0.056	(0.056)		
Incentives × No Spillover	0.029	(0.045)		
Incentives × Spillover	0.020	(0.054)		
Info-Tool × No Spillover	0.085*	(0.044)		
Info-Tool × Spillover	0.163***	(0.057)	0.120**	(0.060)
Observations	2616		2616	
P-values:				
Chlorine: No Spillover = Spillover	0.494			
Incentives: No Spillover = Spillover	0.876			
Info-Tool: No Spillover = Spillover	0.202			

Standard errors in parentheses
Child-level cross-section of the endline survey. Standard errors are clustered at the household level. All regressions include baseline unbalanced household-level controls, baseline child anthropometrics and diarrhea rates, child gender, child age, neighborhood block fixed effects, and lasso-selected controls. The index is constructed using following Anderson (2008) from the following variables: WAZ, WHZ, HAZ, and Muac-for-age. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: IV Impacts on Endline Child Health:
where "Spillover" is exposure to any Info-tool neighbor

	(1) Instrument: Any Treatment	(2) Instrument: Info-Tool	(3) Instrument: Info-Tool × Spillover
Boils, Bleaches, or Chlorinates Water	0.240**	0.329***	0.506***
	(0.113)	(0.126)	(0.185)
Observations	2616	2616	2616
Endline Control Mean	-0.017	0.023	0.031
First-Stage Coefficient	0.312	0.329	0.312
Weak-IV robust F statistic	119.931	86.571	35.258
C-statistic p-value	0.027	0.005	0.002

Standard errors in parentheses
Child-level cross-section of the endline survey. Standard errors are clustered at the household level. Any treatment (i.e., received chlorine) is an instrument for if the respondent reported at endline that she boils, bleaches, or chlorinates her water. All regressions include baseline unbalanced household-level controls, baseline child anthropometrics and diarrhea rates, child gender, child age, neighborhood block fixed effects, and lasso-selected controls. The index is constructed using following Anderson (2008) from the following variables: WAZ, WHZ, HAZ, and Muac-for-age. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

The Info-tool treatment, and especially being in the Info-tool group with a neighbor in the Info-tool group, led to a higher rate of cumulative use of chlorine throughout the study. Therefore, it is likely that someone in these groups who reports purifying her water at endline had a higher rate of historical use than other treated participants, and that the effect of being someone who reports purifying her water at endline has a different effect on child health. To understand if there is heterogeneity in the causal effect of water purification on child health, we test three sets of compliers by varying our instrument.

Column (1) of Table 14 replicates column (1) of Table 11: assignment to any treatment group is the instrument for using an effective water purification method, and the compliers are any participants in the Chlorine Only, Incentives, or Info-tool group who were induced to report using an effective water purification method due to random assignment into any treatment group. In Table 14 column (2), the instrument is an indicator for being in the Info-tool group, and the compliers are individuals who were induced to report using an effective water purification method due to random assignment into the Info-tool group (the regression controls for being in the Incentives or Chlorine-Only groups). The estimate is 37% higher (0.329 SD increase in child anthropometrics), suggesting that the higher intensity of chlorine use that we detect among Info-tool throughout the study period translates into greater improvements in child health. Finally, in Table 14 column (3), the instrument is an indicator for being in the Info-tool group *and* having random exposure to an Info-tool neighbor, and the compliers are individuals who were induced to report using an effective water purification method due to randomly being in the Info-tool group and having an Info-tool neighbor (the regression controls for being in the Incentives or Chlorine-Only groups, and for being in the spillover sample).

The estimate of water purification on child health is 111% higher than the any-chlorine-group treatment effect, and 54% higher than the average Info-tool treatment effect (0.506 SD increase in child anthropometrics).[39]

These heterogeneous results further point towards full saturation of the Info-tool treatment as the optimal policy. Comparing our results on anthropometrics with results from other recent programs in South Asia, our results on weight-for-age are significantly larger than those from handwashing, hygiene, nutrient supplements, or early childhood education programs (Table 15). We do not see any results on height-for-age, unlike the handwashing, hygiene, and nutrient supplement programs.

Table 15: Benchmarking Child Health Estimates

Intervention	**Paper**	**HAZ**	**WHZ**	**WAZ**	**Muac-for-age**
Chlorine: Average	Table 11	0.050	-0.017	0.371*	0.198
Chlorine: Saturated Info-tool	Table D.7	-0.083	0.584	0.789**	-0.428
Handwashing	Hussam et al. (2022)	0.272	–	0.203	0.078
Hygiene	Bennett et al. (2018)	0.290	–	0.270	–
Nutrient Supplements	Sazawal et al. (2013)	0.180	–	0.030	–
	Soofi et al. (2022)	0.290	0.050	0.260	–
ECD	Bos et al. (2024)	-0.024	0.230	0.137	–

7 Conclusion

We offer three policy takeaways from this study. The first directly addresses the question of scale. When experiential and social learning are complementary, there is high added value in saturating a learning or information treatment. Indeed, we find that the cost per DALY averted is *decreasing* in the density of Info-Tool participants in an area (Figure 7).[40] We estimate that the cost per DALY averted for an Info-Tool participant with at least one other Info-Tool neighbor is 2778 USD, while the cost per DALY averted for all other participants is 38% to 131% larger (depending on the treatment group and density of Info-Tool neighbors). It is important that policymakers not assume that these interventions will work similarly if disseminated sparsely within a population, for example, by treating or seeding knowledge among a specific sub-population.

Second, we add to the literature documenting the necessity of trust in information that individuals receive about health in order for it to impact behavior, a phenomenon with specific implications for historically marginalized communities. We operate in a setting where people lack formal consumer protection and have recent experience with insincere and politicized health campaigns (Martinez-Bravo and Stegmann, 2022).[41] This adds to a body of research

[39]Table D.7 reports the corresponding estimates for the four components of the index.

[40]We assume that every time we detect chlorine is equivalent to 30 days of water chlorination, and that 11,000 days of water chlorination equates to 1 DALY averted (International, 2017). Each treatment household cost 24.78 USD. Each Info-Tool household cost an additional 7.84 USD (paper materials and CHW time), and each Incentives household cost on average an additional 0.88 USD (short-term incentives provision).

[41]It is important to note that information coming from neighbors was effective above and beyond the in-

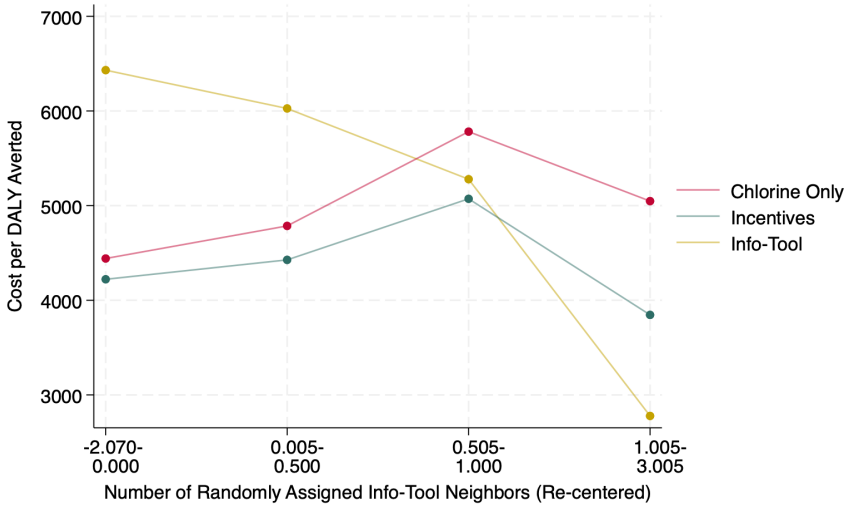


Figure 7: Cost-Effectiveness Analysis

showing that, while the identity of someone delivering information is important for generating trust, *how* they deliver that information is likewise critical. As outsiders in these communities, researchers may not have the relevant knowledge and skills to craft the messages that will be most effective. Allowing community members to experience and learn for themselves, and then craft their own messages, might be a more effective information campaign strategy.

Third, we add to a small but growing body of evidence that the underlying mechanism behind behavioral change or technological adoption is more complex than a mere shift in explicit knowledge about the returns to the behavior or technology (Hussam et al., 2022; Conlon et al., 2022; Fafchamps et al., 2024); rather, in our context, learning appears to be most effective when it is action-based and similarly experienced by others in one's network.

Our results speak to public health programs across a wide geographic space. The take-up of water purification products is low in many developing countries. In Pakistan, only 0.3% of the population report usage of chlorine tablets, with 7.1% adopting any purification technology (Pakistan DHS 2017-18). The cost and availability of such tablets is an important barrier to adoption. In our endline take-it-or-leave-it willingness-to-pay exercise, only 2.1% of the sample were willing and able to pay *anything* for a one-month supply of chlorine tablets, and only 1.7% were willing and able to pay the market price. This is likely driven by inability to pay rather than low valuation of chlorine, because participants demonstrated demand for chlorine in other ways. Seven percent of participants demonstrated willingness to give up their time for the chance to purchase chlorine by asking the enumerator to return at a later date to try selling again, hoping that they would have cash on hand available at another time. In the same visit, 37% reported that they currently used chlorine to purify their water,[42] and 77% accepted free chlorine tablets.

formation coming from the Community Health Workers, most of whom live in the same community as the respondents.

[42]The correlation between self-reported chlorine use at endline and the probability that we ever detected chlorine residual in the prior three months was 0.22.

However, free provision induces only partial usage, and low demand is a barrier *conditional* on accessible supply.[43] Despite extremely low rates of chlorine tablet purchasing in our take-it-or-leave it exercise, we find the highest rates of take-up among the Info-Tool group, which is statistically distinguishable from take-up in the Chlorine Only group, who had the lowest willingness-to-pay (Table C.15). Many features of the study context are characteristic of low- and middle-income countries across South Asia and Sub-Saharan Africa (Supply and Programme, 2014). Inadequate public WASH infrastructure and low use of cheap point-of-use water purification technology, such as chlorine tablets, lead to environments with water contamination and high diarrhea prevalence. Full subsidization of these technologies has not been successful in bringing about substantive increases in take-up (Akram and Mendelsohn (2021) in Pakistan; Dupas et al. (2016) in Kenya). The materials that we use in our intervention are reasonable to use outside of an experiment and in other contexts. The Info-tool is a simple and cheap pencil-and-paper intervention that low-literacy-and-numeracy individuals can easily use. Community health workers are a common feature of health systems across low- and middle-income countries, and our field protocols fit into the typical health worker workflow; by integrating into existing workflows, this intervention does not impose added human resource burdens on institutions working in this space. As such, this learning tool is potentially scalable.

Aside from point-of-use chlorination for water purification, for what types of technologies might there be a complementarity between experiential and social learning? Our model suggests that the best candidates are technologies that are costly to adopt, require persistent use, and produce observable yet noisy treatment effects. Technologies that require persistent use provide users with many opportunities for adoption and observation. However, with sufficiently noisy treatment effects and sufficiently high costs, experiential learning may need reinforcement through trust-worthy sources. Examples of health technologies that meet these criteria include: sanitation (e.g., hand-washing and proper treatment of food), mental health care practices (e.g., therapy and meditation), lifestyle changes (e.g., diet and exercise), and annual vaccinations (e.g., influenza and COVID-19). Examples of non-health technologies that meet these criteria include: agricultural practices, childcare practices[44], and learning or study habits.[45] A promising avenue for future research is to test for within-treatment spillover effects of programs encouraging take-up of technologies that, according to our criteria, are candidates for experiential and social learning complementarities.

[43] For example, free delivery of chlorine solution in Kenya yielded a usage rate of 34% (Dupas et al., 2016).

[44] An illustrative example is the parental education groups organized in many countries under the coordination of trained health-workers. These groups provide a centralized source of reliable information, communicated by the worker, while also fostering peer-to-peer learning. Parents share specific details and insights about various practices based on their personal experiences, thereby complementing the centralized guidance with practical, experiential knowledge (Bunting, 2004).

[45] List and Uchida (2024) finds within-treatment (but no cross-treatment) spillover effects of early childhood education on cognition, driven by social network effects. While they do not specify what behaviors drive these cognitive gains, it is conceivable that students learn behaviors that promote cognitive development in preschool, which the students reinforce in one another if they are in classrooms together in kindergarten and grade school.

References

Abadie, Alberto, Matthew M. Chingos, and Martin R. West, "Endogenous Stratification in Randomized Experiments," *Review of Economics and Statistics*, 2018, *C* (4), 567–580.
AEA RCT Registry Dataverse

AEA RCT Registry Dataverse, *https://www.socialscienceregistry.org/site/csv 2024. Accessed: 2024-11-12.*

Aggarwal, Shilpa, Rebecca Dizon-Ross, and Ariel D. Zucker, *"Incentivizing behavioral change: The role of time preferences," NBER Working Paper No. w27079, 2020.*

Akram, A. A. and R. Mendelsohn, *"Diaries to Increase the Adoption of Chlorine Tablets for Water Purification by Poor Households," Water Economics and Policy (WEP), 2021, 7 (2), 1–34.*

Allcott, Hunt and Todd Rogers, *"The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," American Economic Review, 2014, 104 (10), 3003–3037.*

Alpízar, Francisco, Maria Bernedo Del Carpio, Paul J. Ferraro, and Ben S. Meiselman, *"Exposure-enhanced goods and technology disadoption," Working Paper., 2022.*

Alsan, Marcella and Marianne Wanamaker, *"Tuskegee and the Health of Black Men," The Quarterly Journal of Economics, 2018, 133 (1), 407–455.*

_ and Sarah Eichmeyer, *"Experimental Evidence on the Effectiveness of Nonexperts for Improving Vaccine Demand," American Economic Journal: Economic Policy, 2024, 16 (1), 394–414.*

_ , Owen Garrick, and Grant Graziani, *"Does Diversity Matter for Health? Experimental Evidence from Oakland," American Economic Review, 2019, 109, 4071–4111.*

Anderson, Michael L., *"Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," Journal of the American Statistical Association, 2008, 103 (484), 1481–1495.*

Arrow, Kenneth, *"The Economic Implications of Learning by Doing," The Review of Economic Studies, 1962, 29 (3), 155–173.*

Banerjee, Abhijit, Abhijit Chowdhury, Jishnu Das, Jeffrey Hammer, Reshmaan Hussam, and Aakash Mohpal, *"The Market for Healthcare in Low Income Countries," Working Paper, 2023.*

_ , **Arun G. Chandrasekhar, Esther Duflo, and Matthew O Jackson**, *"Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials," The Review of Economic Studies, 2019, 86 (6), 2453–2490.*

_ , **Emily Breza, Arun Chandrasekhar, and Ben Golub**, *"When Less is More: Experimental Evidence on Information Delivery during India's Demonization," Review of Economic Studies, 2024, 91, 1884–1992.*

Beaman, Lori, Ariel BenYishay, Jeremey Magruder, and Ahmed Mushfiq Mobarak, *"Can Network Theory-Based Targeting Increase Technology Adoption?," American Economic Review, 2021, 111 (6).*

Becker, Gary S and Kevin M Murphy, *"A theory of rational addiction," Journal of political Economy, 1988, 96 (4), 675–700.*

Bennett, Daniel, Asjad Naqvi, and Wolf-Peter Schmidt, *"Learning, Hygiene and Traditional Medicine," The Economic Journal, 2018, 128 (612), F545–74.*

Bikhchandani, Sushil, David Hirshleifer, Omer Tamuz, and Ivo Welch, *"Information Cascades and Social Learning," Journal of Economic Literature, 2024, 62, 1040–1093.*

Borusyak, Kirill and Peter Hull, *"Nonrandom Exposure to Exogenous Shocks," Econometrica, 2023, 91, 2155–2185.*

Bos, Johannes M., Abu S. Shonchoy, Saravana Ravindran, and Akib Khan, *"Early childhood human capital formation at scale," Journal of Public Economics, 2024, 231.*

Breza, Emily and Arun Chandrasekhar, *"Social Networks, Reputation, and Commitment: Evidence from a Savings Monitors Experiment," Econometrica, 2019, 87 (1), 175–216.*

Bunting, Lisa, *"Parenting programmes: The best available evidence," Child Care in Practice, 2004, 10 (4), 327–343.*

Calónico, Sebastian, Rafael Di Tella, and Juan Cruz Lopez del Valle, *"The Political Economy of a "Miracle Cure": The Case of Nebulized Ibuprofen and its Diffusion in Argentina," NBER Working Paper 31781, 2023.*

Caro-Burnett, Johan, Judith A. Chevalier, and Ahmed Mushfiq Mobarak, *""Is Habit a Powerful Policy Instrument to Induce Prosocial Behavioral Change?," Cowles Foundation Discussion Papers, 2021, 2600.*

Celhay, Pablo A., Paul J. Gertler, Paula Giovagnoli, and Christel Vermeersch, *"Long-Term Effects of Temporary Incentives on Productivity in Medical Care Clinics," American Economic Journal: Applied Economics, 2015, 11 (3), 92–127.*

Chandrasekhar, Arun G., Esther Dufo, Michael Kremer, João F. Pugliese, Jonathan Robinson, and Frank Schilbach, *"Blue Spoons: Sparking Communication about Appropriate Technology Use," NBER Working Paper 30423, 2022.*

Chen, Yiqun, Petra Persson, and Maria Polyakova, *"The Roots of Health Inequality and the Value of Intrafamily Expertise," American Economic Journal: Applied Economics, 2022, 14 (3), 185–223.*

Conley, Timothy G. and Chris Udry, *"Learning about a New Technology: Pineapple in Ghana," The American Economic Review, 2010, 100 (1), 35–69.*

Conlon, John, Manvika Mani, Gautam Rao, Matthew Ridley, and Frank Schilbach, *"Not Learning from Others," Revise Resubmit, Econometrica, 2022.*

Corno, Lucia, *"Learning (or Not) in Health-Seeking Behavior: Evidence from Rural Tanzania," Economic Development and Cultural Change, 2014, 63 (1), 27–72.*

D'Acunto, Francesco, Ulrike Malmendier, Juan Ospina-Tejeiro, and Michael Weber, *"Exposure to Grocery Prices and Inflation Expectations," Journal of Political Economy, 2021, 129, 1615–1639.*

Darden, Michael E. and Mario Macis, *"Trust and Health Care-Seeking Behavior," NBER Working Paper 32028, 2024.*

de Janvry, Alain, Karen Macours Macours, and Elisabeth Sadoulet, *"Learning for adopting: Technology adoption in developing country agriculture," FERDI, 2016.*

Duhigg, Charles, *The power of habit: Why we do what we do in life and business, Vol. 34, Random House, 2012.*

Dupas, Pascaline, Vivian Hoffman, Michael Kremer, and Alix Peterson Zwane, *"Targeting Health Subsidies through a Nonprice Mechanism: A Randomized Controlled Trial in Kenya," Science, 2016, 353 (6302), 889–895.*

Dussault, Gilles and Maria Cristina Franceschini, *"Not enough there, too many here: understanding geographical imbalances in the distribution of the health workforce," Human Resources for Health, 2006, 4 (12).*

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker, *"General equilibrium effects of cash transfers: experimental evidence from Kenya," Econometrica, 2022, 90 (6), 2603–2643.*

Facchini, Gabriel, *"Forgetting-by-not-doing: The case of surgeons and cesarean sections," Health Economics, 2022, 31 (3), 481–495.*

Fafchamps, Marcel, Asad Islam, Debayan Pakrashi, and Denni Tommasi, *"Diffusion in social networks: Experimental evidence on information sharing vs persuasion," Working Paper, 2024.*

Foster, Andrew D. and Mark R. Rosenzweig, *"Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," The Journal of Political Economy, 1994, 103 (6), 1176–1209.*

Halm, Ethan A., Clara Lee, and Mark R. Chassin, *"Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature," Annals of Internal Medicine, 2002, 137 (6).*

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, *"Learning Through Noticing: Theory and Evidence from a Field Experiment," The Quarterly Journal of Economics, 2014, 129 (3), 1311–1353.*

Haushofer, Johannes, Michael Kremer, Ricardo Maertens, and Brandon Joel Tan, *"Water treatment and child mortality: Evidence from kenya," Technical Report, National Bureau of Economic Research 2021.*

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol, *"Rational Habit Formation: Experimental Evidence from Handwashing in India," American Economic Journal: Applied Economics, 2022, 14 (1), 1–41.*

_ , Kailash Pandey, Abu Shonchoy, and Chikako Yamauchi, *"Translating Information into Action," Revise and resubmit, American Economic Review, 2023.*

International, Population Services, *"PSI Impact Calculator,"* `http://impactcalculator.psi.org/`. *2017.*

Khandelwal, Vatsal, *"Learning in networks with idiosyncratic agents," Games and Economic Behavior, 2024, 144, 225–249.*

Kondylis, Florence, John Ashton Loeser, Mushfiq Mobarak, Maria Ruth Jones, and Daniel Kevin Stein, *"Learning from Self and Learning from Others : Experimental Evidence from Bangladesh (English).," Policy Research working paper ; no. WPS 10545; Impact Evaluation series Washington, D.C. : World Bank Group., 2023.*

Kremer, Michael, Stephen P Luby, Ricardo Maertens, Brandon Tan, and Witold Wiecek, *"Water Treatment And Child Mortality: A Meta-Analysis And Cost-effectiveness Analysis," Technical Report, National Bureau of Economic Research 2023.*

List, John A. and Haruka Uchida, *"Here Today, Gone Tomorrow? Toward an Understanding of Fade-Out in Early Childhood Education Programs," NBER Working Paper 33027, 2024.*

Lowes, Sara and Eduardo Montero, *"The Legacy of Colonial Medicine in Central Africa," American Economic Review, 2021, 111 (4), 1284–1314.*

Luby, Stephen P, Mubina Agboatwalla, John Painter, Arshad Altaf, Ward Billhimer, Bruce Keswick, and Robert M. Hoekstra, *"Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial," The American Journal of Tropical Medicine and Hygiene, 2006, 11 (4), 479–489.*

Malmendier, Ulrike and Stefan Nagel, *"Learning from Inflation Experiences," The Quarterly Journal of Economics, 2015, 131 (1), 53–87.*

Martinez-Bravo, Monica and Andreas Stegmann, *"In Vaccines We Trust? The Effects of the CIA's Vaccine Ruse on Immunization in Pakistan," Journal of the European Economic Association, 2022, 20 (1), 150–186.*

Miguel, Edward and Michael Kremer, *"Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," Econometrica, 2004, 72, 159–217.*

Muralidharan, Karthik and Venkatesh Sundararaman, *"Teacher Performance Pay: Experimental Evidence from India," Journal of Political Economy, 2011, 119 (1), 39–77.*

Neal, David, Jelena Vujcic, Orlando Hernandez, Wendy Wood, J Vujcic, O Hernandez, and W Wood, *"The science of habit: creating disruptive and sticky behavior change in handwashing behavior," Washington DC, USA. USAID/WASHplus Project, 2015.*

Quick, Robert E., Akiko Kimura, Angelica Thevos, Mathias Tembo, Isidore Shamputa, Lori Hutwagner, and Eric Mintz, *"Diarrhea Prevention through Household-Level Water Disinfection and Safe Storage in Zambia," The American Journal of Tropical Medicine and Hygiene, 2002, 66 (5), 584–589.*

Quick, Robert E, LV Venczel, ED Mintz, L Soleto, J Aparicio, M Gironaz, L Hutwagner, K Greene, C Bopp, K Maloney et al., *"Diarrhoea prevention in Bolivia through point-of-use water treatment and safe storage: a promising new strategy," Epidemiology & Infection, 1999, 122 (1), 83–90.*

Reller, Megan E, Carlos E Mendoza, M Beatriz Lopez, Maricruz Alvarez, Robert M Hoekstra, Christy A Olson, Kathleen G Baier, Bruce H Keswick, and Stephen P Luby, *"A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala," American Society of Tropical Medicine and Hygiene, 2003, 69 (4), 411–419.*

Royer, Heather, Mark Stehr, and Justin Sydnor, *"Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company," American Economic Journal: Applied Economics, 2015, 7 (3), 51–84.*

Sazawal, Sunil, AKM Ahsan Habib, Usha Dhingra, Arup Dutta, Pratibha Dhingra, Archana Sarkar, Saikat Deb, Jahangir Alam, Asmaul Husna, and Robert E. Black, *"Impact of micronutrient fortification of yoghurt on micronutrient status markers and growth – a randomized double blind controlled trial among school children in Bangladesh," BMC Public Health, 2013, 13 (514).*

Scharf, Rebecca J, Mark D DeBoer, and Richard L Guerrant, *"Recent advances in understanding the long-term sequelae of childhood infectious diarrhea," Current infectious disease reports, 2014, 16, 1–7.*

Simonsohn, Uri, Niklas Karlsson, George Loewenstein, and Dan Ariely, *"The tree of experience in the forest of information: Overweighing experienced relative to observed information," Games and Economic Behavior, 2008, 62, 263–286.*

Soofi, Sajid Bashir, Gul Nawaz Khan, Shabina Ariff, Yasir Ihtesham, Mahamadou Tanimoune, Arjumand Rizvi, Muhammad Sajid, Cecilia Garzon, Saskia de Pee, and Zulfiqar A. Bhutta, *"Effectiveness of nutritional supplementation during the first 1000-days of life to reduce child undernutrition: A cluster randomized controlled trial in Pakistan," The Lancet Regional Health – Southeast Asia, 2022, 4, 1–11.*

Supply, WHO/UNICEF Joint Water and Sanitation Monitoring Programme, *Progress on drinking water and sanitation: 2014 Update, World Health Organization, 2014.*

Tremblay, Maxime A., Céline M. Blanchard, Sara Taylor, Luc G. Pelletier, and Martin Villeneuve, *"Work Extrinsic and Intrinsic Motivation Scale: Its Value for Organizational Psychology Research," Canadian Journal of Behavioural Science, 2009, 41 (4), 213–226.*

UNICEF and WHO, *"Progress on Sanitation and Drinking Water: 2015 Update and MDG Assessment," Geneva, 80 pages, 2015.*

Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov, *"Using double-lasso regression for principled variable selection," Working Paper, 2016.*

Watson, Samuel I, Ryan TT Rego, Timothy Hofer, and Richard J Lilford, *"Evaluations of water, sanitation and hygiene interventions should not use diarrhoea as (primary) outcome," BMJ Global Health, 2022, 7 (5), e008521.*

Wellsjo, Alexandra Steiny, *"Simple Actions, Complex Habits: Lessons from Hospital Hand Hygiene," Technical Report, Working paper 2021.*

WHO, *"Diarrhoeal disease factsheet 2017," 2017.*

Xu, Zhiwei, Yang Liu, Zongwei Ma, Ghasem (Sam) Toloo, Wenbiao Hu, and Shilu Tong, *"Assessment of the temperature effect on childhood diarrhea using satellite imagery," Scientific Reports, 2014, 4 (5389).*

Online Supplementary Material

Appendix A Baseline Balance

Table A.1: Baseline Balance: Adult Outcomes

Variable	(1) Control Mean/SE	(2) Chlorine-tablets Mean/SE	(3) Incentives Mean/SE	(4) Info-tool Mean/SE	(5) Total Mean/SE	T-test Difference (1)-(2)	(1)-(3)	(1)-(4)
Any Child Had Motions	0.283	0.356	0.304	0.291	0.309	-0.073**	-0.021	-0.008
	(0.021)	(0.023)	(0.022)	(0.021)	(0.011)			
Reported Highest Diarrhea in Summer	0.836	0.844	0.825	0.820	0.831	-0.008	0.011	0.016
	(0.017)	(0.017)	(0.018)	(0.018)	(0.009)			
Number of Children <5	1.496	1.517	1.479	1.507	1.499	-0.021	0.017	-0.011
	(0.034)	(0.038)	(0.031)	(0.037)	(0.017)			
Has Heard of Chlorine	0.188	0.194	0.193	0.240	0.204	-0.006	-0.005	-0.052**
	(0.018)	(0.019)	(0.019)	(0.020)	(0.009)			
Would consider using chlorine	0.869	0.880	0.887	0.878	0.878	-0.010	-0.017	-0.008
	(0.016)	(0.015)	(0.015)	(0.015)	(0.008)			
Enumerator Observes Dirt in Water	0.164	0.160	0.173	0.156	0.163	0.003	-0.009	0.008
	(0.017)	(0.017)	(0.018)	(0.017)	(0.009)			
Reports Dirt in Water	0.752	0.753	0.743	0.751	0.750	-0.001	0.009	0.001
	(0.020)	(0.020)	(0.021)	(0.020)	(0.010)			
Boils, Bleaches, or Chlorinates Water	0.148	0.131	0.131	0.147	0.139	0.017	0.017	0.002
	(0.017)	(0.016)	(0.016)	(0.017)	(0.008)			
Strains or Filters Water	0.633	0.657	0.619	0.633	0.635	-0.024	0.014	-0.001
	(0.023)	(0.022)	(0.023)	(0.023)	(0.011)			
Believes Chlorine is for Water Purification	0.188	0.198	0.220	0.218	0.206	-0.010	-0.031	-0.030
	(0.018)	(0.019)	(0.020)	(0.019)	(0.010)			
Caretaker Asked about Chlorine Test	0.142	0.176	0.140	0.144	0.150	-0.034	0.002	-0.003
	(0.016)	(0.018)	(0.016)	(0.017)	(0.008)			
Attrited (endline)	0.106	0.136	0.149	0.113	0.126	-0.030	-0.042**	-0.007
	(0.015)	(0.016)	(0.017)	(0.015)	(0.008)			
N	452	449	451	450	1802			
F-test of joint significance (F-stat)						0.977	0.748	0.430

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are robust. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.2: Baseline Balance: Child Outcomes

	(1) Control		(2) Chlorine-tablets		(3) Incentives		(4) Info-tool		T-test Difference		
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)	(1)-(3)	(1)-(4)
Child Weight (kg)	672	10.461	675	10.417	664	10.621	670	10.624	0.044	-0.159	-0.163
	[451]	(0.111)	[448]	(0.109)	[448]	(0.131)	[449]	(0.111)			
Child MUAC (cm)	676	14.504	678	14.428	666	14.474	675	14.546	0.076	0.030	-0.041
	[452]	(0.048)	[449]	(0.042)	[450]	(0.049)	[450]	(0.050)			
Number of Motion Days	676	0.761	679	0.928	667	0.886	672	0.779	-0.167*	-0.125	-0.019
	[452]	(0.067)	[449]	(0.077)	[451]	(0.080)	[450]	(0.074)			
Child Had > 0 Motion Days	676	0.214	680	0.278	667	0.226	675	0.230	-0.063***	-0.012	-0.015
	[452]	(0.016)	[449]	(0.019)	[451]	(0.016)	[450]	(0.017)			

Notes: The value displayed for t-tests are the differences in the means across the groups. Standard errors are clustered at variable HHID. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A.3: Treatment Balance: Spillover

	(1) Spillover Sample		(2) Spillover Sample	
Any Treatment Group	0.019	(0.025)		
Chlorine Only			0.018	(0.031)
Incentives			0.019	(0.031)
Info-Tool			0.019	(0.031)
Observations	1690		1690	
P-values:				
Incentives = Chlorine Only			0.975	
Info-Tool = Chlorine Only			0.981	
Info-Tool = Incentives			0.995	

Standard errors in parentheses
Each observation is at the household level. This regression includes neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Appendix B Akram and Mendelsohn (2021) Reanalysis

We reanalyze the data from Akram and Mendelsohn (2021), computing the same measure of spillovers that we compute in our data. We find that Info-tool participants with another Info-tool participant within twenty meters accept chlorine 17% more often than in other households (Akram and Mendelsohn (2021) used chlorine tablet acceptance as their primary outcome throughout the trial, and only tested water for chlorine residual in the final endline visit.)

Similar to our study, the spillover and non-spillover samples diverge most starkly soon after the period ends where the participants are using the Info-tool, and after they see the cumulative bar chart (Figure B.1). Also similar to our study, the spillover and non-spillover samples converge towards the end of the trial.

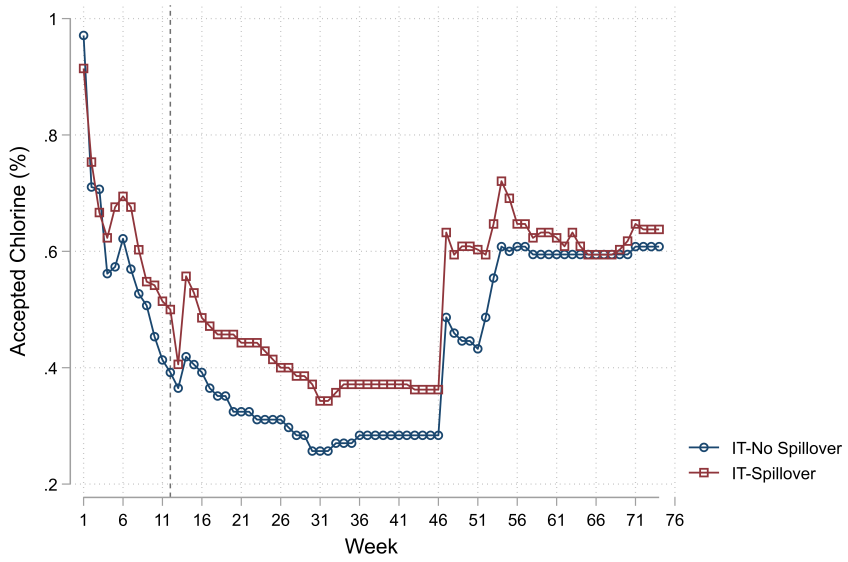


Figure B.1: Akram and Mendelsohn (2021) Reanalysis
Rates of Chlorine Acceptance
where "Spillover" is exposure to any Info-tool neighbor

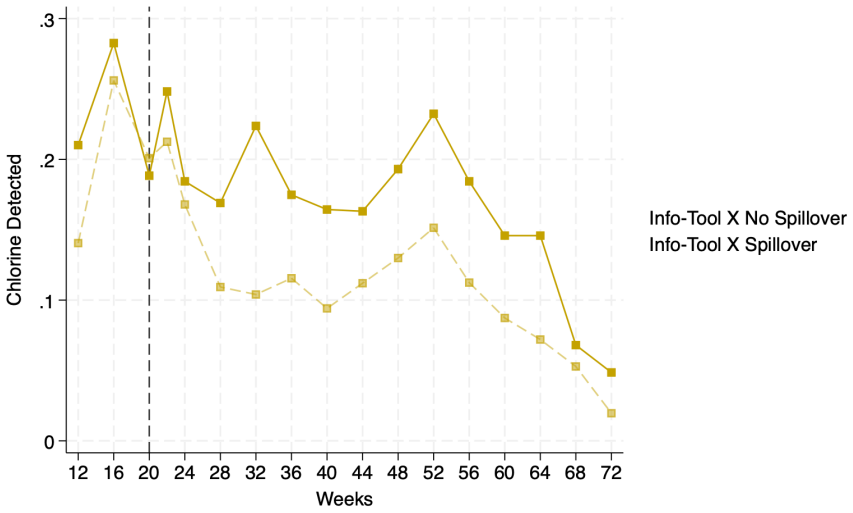


Figure B.2: Our Study
Rates of Chlorine Detection
where "Spillover" is exposure to any Info-tool neighbor

Appendix C Additional Figures and Tables

Table C.1: Chlorine Detection (Household-Survey Panel)

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine Only	0.221***	(0.015)	0.147***	(0.011)	0.108***	(0.009)
Incentives	0.255***	(0.015)	0.142***	(0.010)	0.126***	(0.010)
Info-Tool	0.208***	(0.014)	0.168***	(0.011)	0.122***	(0.009)
Observations	4711		6354		14066	
P-values:						
Chlorine = Incentives	0.090		0.689		0.117	
Chlorine = Info-Tool	0.484		0.142		0.201	
Incentives = Info-Tool	0.014		0.055		0.749	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: Child-Days of Diarrhea (Household-Survey Panel)

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine Only	-0.186***	(0.062)	-0.070	(0.045)	-0.091***	(0.032)
Incentives	-0.198***	(0.057)	-0.027	(0.045)	-0.050	(0.033)
Info-Tool	-0.136**	(0.062)	-0.092**	(0.042)	-0.091***	(0.031)
Observations	7872		6354		14066	
Control Mean	0.515		0.240		0.277	
P-values:						
Chlorine = Incentives	0.821		0.271		0.187	
Chlorine = Info-Tool	0.378		0.528		0.990	
Incentives = Info-Tool	0.228		0.070		0.167	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a continuous measure of the total child-days of diarrhea that the household reported over the preceding two weeks (aggregated across all children). All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of children in the household under five in that survey round, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

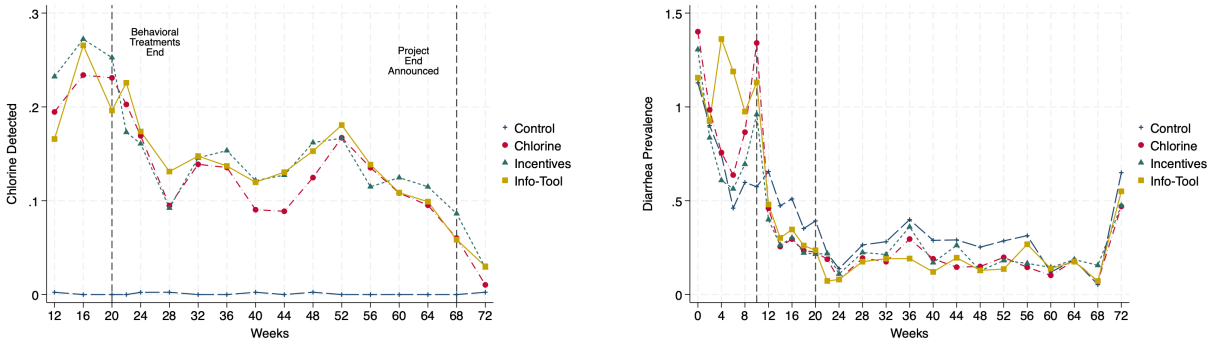


Figure C.1: Raw Average: Chlorine Detection and Diarrhea Prevalence

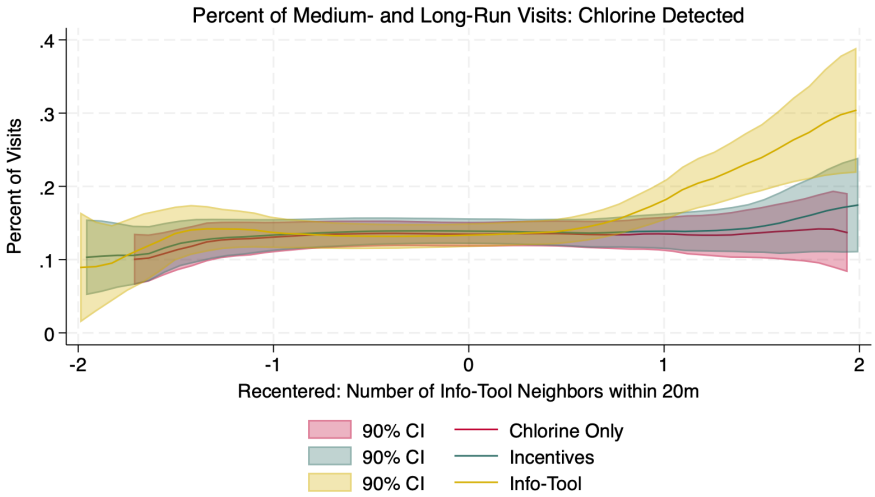


Figure C.2: Raw Rates of Chlorine Detection by Spillover Exposure (Continuous)

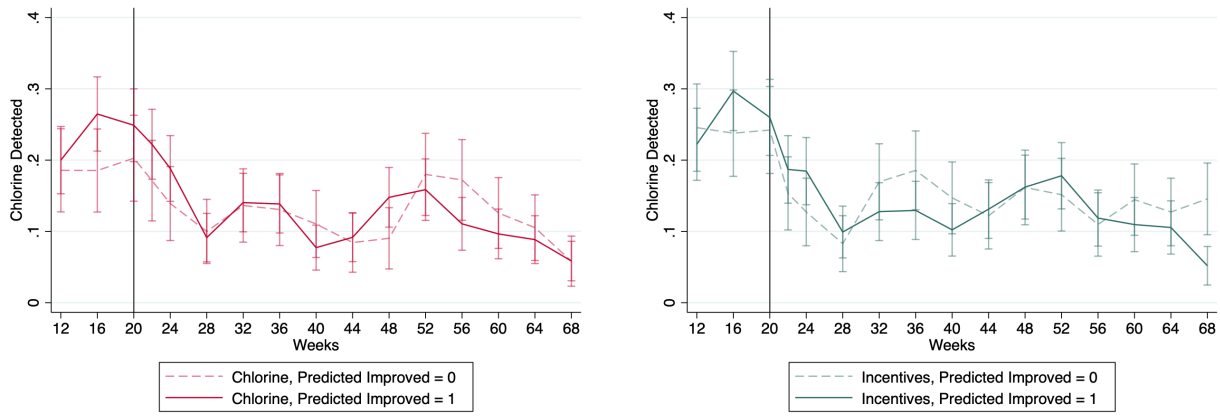


Figure C.3: Heterogeneity: Predicted Health Improvement

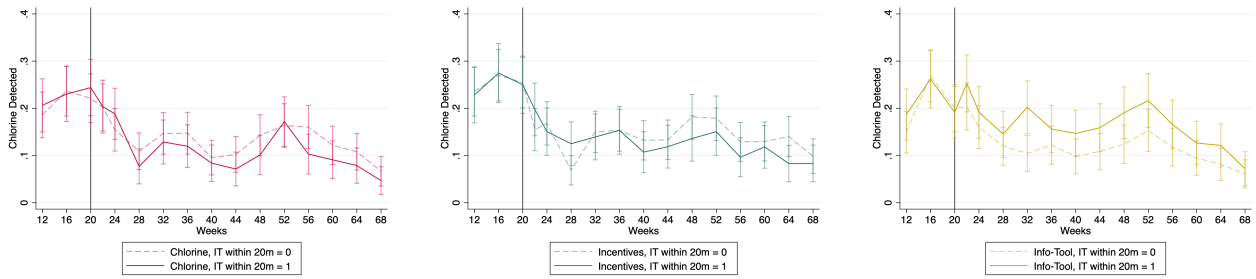


Figure C.4: Heterogeneity: Info-tool Neighbors (Re-centered)

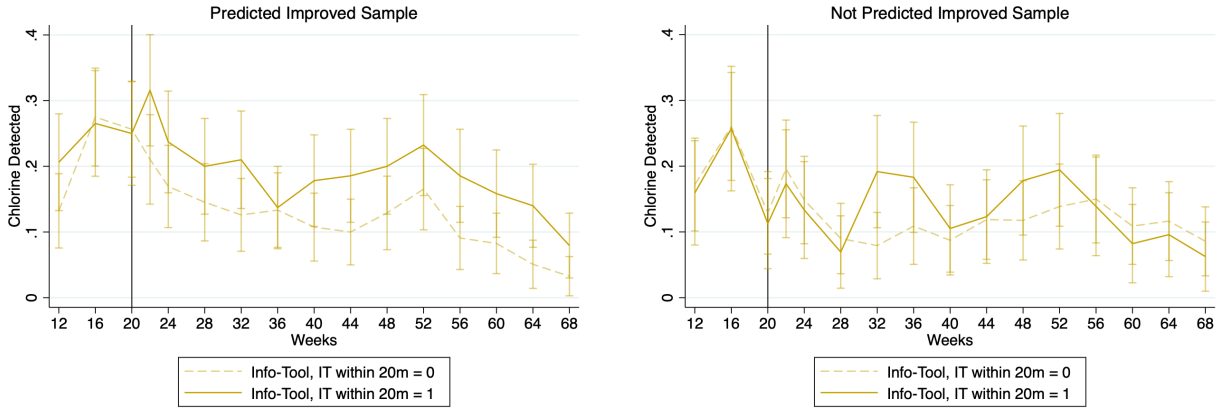


Figure C.5: Heterogeneity: Info-tool Neighbors (Re-centered) × Predicted Health Improvement

Table C.3: Placebo Test – Chlorine Detection (Household Survey Panel)
where "Spillover" is exposure to any Chlorine Only neighbor
Omitted group: Chlorine × No Spillover

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × No Spillover	0.208***	(0.018)	0.133***	(0.013)	0.108***	(0.011)
Chlorine × Spillover	0.237***	(0.026)	0.169***	(0.017)	0.112***	(0.017)
Incentives × No Spillover	0.250***	(0.019)	0.136***	(0.012)	0.131***	(0.013)
Incentives × Spillover	0.255***	(0.025)	0.150***	(0.016)	0.120***	(0.015)
Info-Tool × No Spillover	0.212***	(0.017)	0.166***	(0.014)	0.127***	(0.012)
Info-Tool × Spillover	0.193***	(0.022)	0.166***	(0.018)	0.117***	(0.015)
Observations	4711		6354		14066	
P-values:						
Chlorine: No Spillover = Spillover	0.364		0.102		0.831	
Incentives: No Spillover = Spillover	0.875		0.456		0.607	
Info-Tool: No Spillover = Spillover	0.511		0.999		0.593	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to a Chlorine Only neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

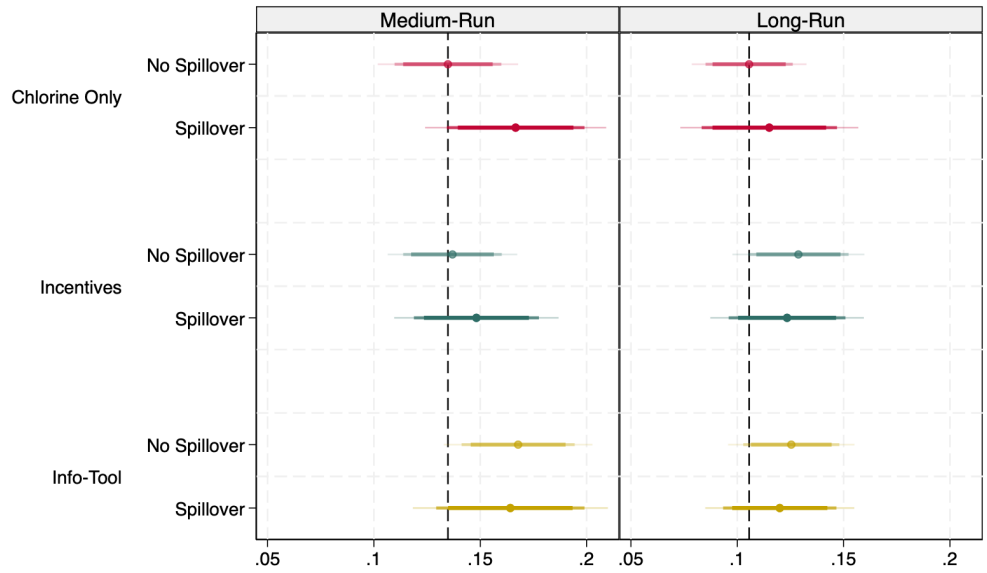


Figure C.6: Placebo Test – Chlorine Detection (Household Survey Panel) where "Spillover" is exposure to any Chlorine Only neighbor

Table C.4: Placebo Test – Chlorine Detection (Panel Specification) where "Spillover" is exposure to any Incentives neighbor Omitted group: Chlorine × No Spillover

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × No Spillover	0.212***	(0.018)	0.141***	(0.013)	0.101***	(0.011)
Chlorine × Spillover	0.231***	(0.026)	0.154***	(0.017)	0.123***	(0.018)
Incentives × No Spillover	0.264***	(0.019)	0.148***	(0.012)	0.135***	(0.013)
Incentives × Spillover	0.231***	(0.024)	0.130***	(0.015)	0.112***	(0.015)
Info-Tool × No Spillover	0.194***	(0.016)	0.163***	(0.014)	0.123***	(0.012)
Info-Tool × Spillover	0.223***	(0.024)	0.172***	(0.019)	0.125***	(0.016)
Observations	4711		6354		14066	
P-values:						
Chlorine: No Spillover = Spillover	0.555		0.564		0.290	
Incentives: No Spillover = Spillover	0.291		0.363		0.250	
Info-Tool: No Spillover = Spillover	0.326		0.702		0.952	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to an Incentives neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

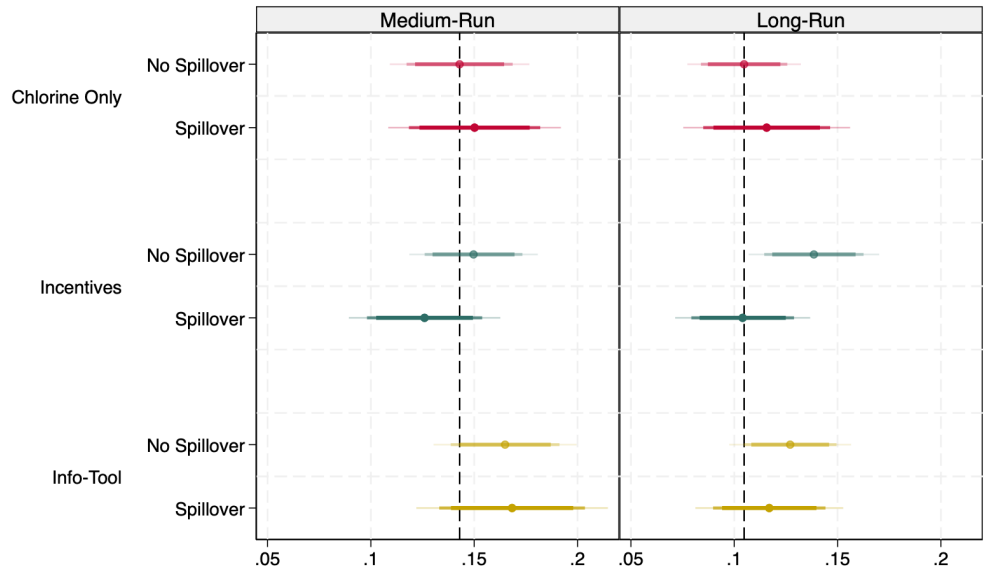


Figure C.7: Placebo Test – Chlorine Detection (Household Survey Panel) where "Spillover" is exposure to any Incentives neighbor

Table C.5: Placebo Test – Chlorine Detection (Household-Survey Panel) where "Spillover" is exposure to any treated neighbor predicted to improve Omitted group: Chlorine × No Spillover

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × Spillover	0.004	(0.030)	0.008	(0.021)	0.012	(0.016)
Incentives × No Spillover	0.058**	(0.027)	-0.007	(0.018)	0.030**	(0.015)
Incentives × Spillover	0.008	(0.028)	-0.002	(0.020)	0.012	(0.017)
Info-Tool × No Spillover	-0.022	(0.024)	0.013	(0.019)	0.018	(0.014)
Info-Tool × Spillover	0.014	(0.028)	0.032	(0.023)	0.022	(0.017)
Observations	3463		4689		10472	
P-values:						
Incentives: No Spillover = Spillover	0.095		0.770		0.275	
Info-Tool: No Spillover = Spillover	0.204		0.382		0.845	

Standard errors in parentheses

Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not chlorine residual was detected in the respondent's water during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to any treatment neighbor (someone within 20m) who was predicted to have their health improve than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined. See Section 6 for a detailed explanation for how the predicted-health-improvement measure is defined.

* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: Endline Stated Trust
Omitted group: Pure Control

	(1) Info-Tool Knows Most	(2) Incentives Knows Most	(3) Chlorine Knows Most
Chlorine × No Spillover	-0.023	-0.105***	0.128***
	(0.034)	(0.038)	(0.042)
Chlorine × Spillover	-0.079*	-0.030	0.109*
	(0.046)	(0.053)	(0.058)
Incentives × No Spillover	0.001	0.069*	-0.070*
	(0.034)	(0.038)	(0.042)
Incentives × Spillover	-0.013	0.125**	-0.112*
	(0.046)	(0.052)	(0.058)
Info-Tool × No Spillover	0.256***	-0.115***	-0.141***
	(0.034)	(0.039)	(0.043)
Info-Tool × Spillover	0.159***	-0.025	-0.134**
	(0.046)	(0.052)	(0.058)
Observations	1516	1516	1516
P-values:			
Outcome Group: Spillover = No Spillover	0.091	0.394	0.792
DID Estimate:			
Info-Tool Spillover Effect			
– Other Group Spillover Effect		-0.152 [p= 0.110]	-0.078 [p= 0.468]

Standard errors in parentheses
Each observation is at the household level. The outcome for column (1) is an indicator for if the respondent chose a hypothetical Info-tool respondent as the person most likely to be knowledgeable about child health (rather than a Chlorine Only or Incentives participant). The outcome for column (2) is an indicator for if the respondent chose a hypothetical Incentives respondent as the person most likely to be knowledgeable about child health (rather than a Chlorine Only or Info-Tool participant). The outcome for column (3) is an indicator for if the respondent chose a hypothetical Chlorine Only respondent as the person most likely to be knowledgeable about child health (rather than an Incentives or Info-Tool participant). All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Endline Social Networks

	(1) Number in Social Network	(2) Number Discuss Health	(3) Number Discussed Water Purification
Incentives	-0.051	-0.022	0.114**
	(0.076)	(0.054)	(0.045)
Info-Tool	-0.119	-0.004	0.005
	(0.076)	(0.054)	(0.045)
Observations	1116	1116	1116
P-values:			
Incentives = Info-Tool	0.377	0.730	0.015

Standard errors in parentheses
Each observation is at the household level. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Endline Social Networks where "Spillover" is exposure to any Info-tool neighbor

	(1) Number in Social Network	(2) Number Discuss Health	(3) Number Discussed Water Purification
Chlorine × Spillover	0.110	-0.065	0.133**
	(0.128)	(0.081)	(0.067)
Incentives × No Spillover	-0.019	-0.050	0.148***
	(0.096)	(0.068)	(0.056)
Incentives × Spillover	0.000	-0.037	0.186***
	(.)	(0.080)	(0.067)
Info-Tool × No Spillover	-0.048	0.004	0.037
	(0.097)	(0.068)	(0.057)
Info-Tool × Spillover	-0.131	-0.081	0.083
	(0.127)	(0.081)	(0.067)
Observations	1116	1116	1116
P-values:			
Incentives × No Spillover = Incentives × Spillover	0.845	0.877	0.576
Info-Tool × No Spillover = Info-Tool × Spillover	0.605	0.297	0.502

Standard errors in parentheses
Each observation is at the household level. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: Mimicry of Other Behaviors

	(1) Take-up of Other Programs: Vitamin A	(2) Take-up of Other Programs: Deworming	(3) Take-up of Other Programs: Iron	(4) Number of Friends Believes Uses Same Savings Technology
Chlorine Only	0.031	0.004	0.011	-0.001
	(0.031)	(0.023)	(0.010)	(0.057)
Incentives	0.067**	0.001	0.010	0.005
	(0.031)	(0.022)	(0.010)	(0.057)
Info-Tool	0.084***	0.004	0.015	-0.114**
	(0.031)	(0.023)	(0.010)	(0.057)
Observations	1512	1512	1512	1503
P-values:				
Chlorine = Incentives	0.246	0.906	0.889	0.910
Chlorine = Info-Tool	0.085	1.000	0.685	0.051
Incentives = Info-Tool	0.566	0.905	0.584	0.039

Standard errors in parentheses
Each observation is at the household level. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.10: Mimicry of Other Behaviors
where "Spillover" is exposure to any Info-tool neighbor

	(1) Take-up of Other Programs: Vitamin A	(2) Take-up of Other Programs: Deworming	(3) Take-up of Other Programs: Iron	(4) Number of Friends Believes Uses Same Savings Technology
Chlorine × No Spillover	0.069*	-0.001	0.013	0.041
	(0.038)	(0.026)	(0.011)	(0.065)
Chlorine × Spillover	-0.041	0.011	0.008	-0.079
	(0.053)	(0.032)	(0.014)	(0.080)
Incentives × No Spillover	0.046	0.017	0.002	0.033
	(0.038)	(0.025)	(0.011)	(0.065)
Incentives × Spillover	0.102*	-0.028	0.025*	-0.045
	(0.052)	(0.031)	(0.013)	(0.080)
Info-Tool × No Spillover	0.102***	0.016	0.012	-0.086
	(0.038)	(0.026)	(0.011)	(0.065)
Info-Tool × Spillover	0.050	-0.018	0.021	-0.166**
	(0.053)	(0.032)	(0.014)	(0.080)
Observations	1512	1512	1512	1503
P-values:				
Chlorine × No Spillover = Chlorine × Spillover	0.092	0.722	0.689	0.165
Incentives × No Spillover = Incentives × Spillover	0.388	0.182	0.104	0.373
Info-Tool × No Spillover = Info-Tool × Spillover	0.424	0.314	0.510	0.353

Standard errors in parentheses
Each observation is at the household level. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.11: Endline Accurate Beliefs About Friends' Water Purification Methods

	(1) Correct Guess (User or Not: SR)		(2) Correct Guess (User or Not: OB)		(3) Correct Guess (Is a Super User)	
Info-Tool Participant	-0.042***	(0.012)	0.020	(0.022)	-0.006	(0.009)
Info-Tool Friend	-0.054***	(0.019)	0.029	(0.038)	-0.003	(0.013)
Info-Tool Participant × Info-Tool Friend	-0.004	(0.039)	0.019	(0.063)	0.007	(0.021)
Observations	2644		2824		2824	
No Info-Tool × No Info-Tool Friend Mean	0.965		0.388		0.042	
P-values:						
Info-Tool Friend + Info-Tool × Info-Tool Friend = 0	0.083		0.407		0.843	

Standard errors in parentheses
Standard errors are clustered at the household level. Each observation is at the network-link level, for observations where the node is within our sample. This regression include neighborhood block fixed effects, unbalanced baseline controls, the number of study participants within twenty meters, lasso-selected baseline controls, participant treatment group fixed effects, and network-link treatment group fixed effects. The outcome in column (1) is an indicator for if the participant's guess about if her friend uses chlorine aligns with what that friend reported using in the endline survey. The outcome in column (2) is an indicator for if the participant's guess about if her friend uses chlorine aligns with what we objectively observed (did we ever detect chlorine in the friend's water). The outcome in column (3) is an indicator for if the participant guessed that her friend uses chlorine tablets for water purification, and we detected chlorine in that participant's water at least three times.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.12: Months Since Water Purification Discussions

	(1) First Water Discussion		(2) Last Water Discussion	
Chlorine	0.314	(0.201)	-1.311*	(0.704)
Incentives	0.060	(0.158)	-2.208***	(0.715)
Info-Tool	-0.044	(0.166)	-0.896	(0.706)
Observations	1357		1357	
Control Mean	1.027		1.027	
P-values:				
Incentives = Chlorine Only	0.154		0.224	
Info-Tool = Chlorine Only	0.050		0.570	
Info-Tool = Incentives	0.460		0.077	

Standard errors in parentheses
Each observation is at the household level. This regression include neighborhood block fixed effects, unbalanced baseline controls, and lasso-selected baseline controls. For links where participants responded that they had never discussed water, we impute with the minimum (0), and with the maximum (24 months).
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.13: Endline Social Norms

	(1) Believes Neighbor Would Accept Chlorinated Water	(2) Number of Friends Believes Chlorinates
Chlorine Only	0.082***	0.041*
	(0.030)	(0.022)
Incentives	0.139***	0.079***
	(0.030)	(0.022)
Info-Tool	0.098***	0.061***
	(0.030)	(0.022)
Observations	1503	1690
P-values:		
Chlorine = Incentives	0.061	0.073
Chlorine = Info-Tool	0.601	0.347
Incentives = Info-Tool	0.176	0.403

Standard errors in parentheses
Each observation is at the household level. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.14: Endline Social Norms where "Spillover" is exposure to any Info-tool neighbor

	(1) Believes Neighbor Would Accept Chlorinated Water	(2) Number of Friends Believes Chlorinates
Chlorine × No Spillover	0.079**	0.041*
	(0.037)	(0.025)
Chlorine × Spillover	0.088*	0.040
	(0.051)	(0.030)
Incentives × No Spillover	0.157***	0.072***
	(0.037)	(0.025)
Incentives × Spillover	0.109**	0.092***
	(0.050)	(0.030)
Info-Tool × No Spillover	0.060	0.049**
	(0.037)	(0.025)
Info-Tool × Spillover	0.164***	0.082***
	(0.051)	(0.030)
Observations	1503	1690
P-values:		
Chlorine × No Spillover = Chlorine × Spillover	0.892	0.969
Incentives × No Spillover = Incentives × Spillover	0.445	0.531
Info-Tool × No Spillover = Info-Tool × Spillover	0.100	0.316

Standard errors in parentheses
Each observation is at the household level. All specifications include neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table C.15: Endline Willingness-to-Pay

	(1) Purchased Tablets	(2) Price Accepted	(3) Hypothetical WTP	(4) Sold or Revisited
Chlorine Only	-0.017*	-2.472*	0.124	-0.044**
	(0.010)	(1.444)	(1.472)	(0.019)
Incentives	-0.010	-1.693	0.215	-0.037*
	(0.010)	(1.448)	(1.482)	(0.019)
Info-Tool	0.004	0.288	0.598	-0.020
	(0.010)	(1.444)	(1.494)	(0.019)
Observations	1503	1503	1099	1503
Control Mean	0.027	4.059	5.217	0.106
P-values:				
Incentives = Chlorine Only	0.474	0.594	0.949	0.685
Info-Tool = Chlorine Only	0.037	0.058	0.742	0.204
Info-Tool = Incentives	0.172	0.175	0.791	0.389

Standard errors in parentheses
Each observation is at the household level. This regression includes neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. The outcome for column (1) is an indicator for if the household ultimately purchased chlorine tablets. The outcome for column (2) is the price at which the household purchased chlorine tablets (the price is 0 for households who did not purchase tablets). The outcome for column (3) is a hypothetical price at which the household said they would be willing to purchase chlorine tablets, if they refused the offered price. The outcome for column (4) is an indicator for if the household purchased the chlorine tablets *or* asked the enumerator to return at a later date when they expected to have cash on hand.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Appendix D Main Tables: Robustness to Alternate Specification

For robustness, we also estimate the following specification using an aggregated version of a given outcome:

$$Y_i = \beta_0 + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + NV_i + X_{i0} + \delta_s + \gamma_b + \epsilon_i, \tag{2}$$

where Y_i represents the aggregated outcome measurements over relevant visits (e.g., total number of visits with chlorine presence detected in the water), and NV_i captures the number of visits with non-missing outcome measurements. All of our tables are replicated using this specification in the Appendix.

Table D.1: Chlorine Detection: Aggregate
where "Spillover" is exposure to any Info-tool neighbor

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × No Spillover	0.641***	(0.056)	0.541***	(0.051)	0.884***	(0.099)
Chlorine × Spillover	0.640***	(0.070)	0.560***	(0.062)	0.955***	(0.120)
Incentives × No Spillover	0.743***	(0.056)	0.506***	(0.051)	1.091***	(0.098)
Incentives × Spillover	0.743***	(0.068)	0.528***	(0.062)	0.937***	(0.119)
Info-Tool × No Spillover	0.617***	(0.057)	0.543***	(0.052)	0.918***	(0.099)
Info-Tool × Spillover	0.631***	(0.070)	0.716***	(0.063)	1.200***	(0.121)
Observations	1599		1690		1690	
P-values:						
Chlorine: No Spillover = Spillover	0.988		0.779		0.585	
Incentives: No Spillover = Spillover	0.998		0.740		0.234	
Info-Tool: No Spillover = Spillover	0.849		0.012		0.034	

Standard errors in parentheses
Each observation is at the household level. The outcome is the aggregate of the total number of times that chlorine was detected across all visits in the specified time period. All specifications include a control for the number of visits that the household was surveyed during the specified time period, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: Chlorine Detection: Aggregate by predicted health improvement

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × Improved	0.088	(0.082)	0.150**	(0.073)	0.152	(0.132)
Incentives × Not Improved	0.138*	(0.081)	0.008	(0.072)	0.168	(0.130)
Incentives × Improved	0.168**	(0.081)	0.079	(0.074)	0.231*	(0.132)
Info-Tool × Not Improved	-0.044	(0.081)	0.088	(0.073)	0.104	(0.131)
Info-Tool × Improved	0.101	(0.083)	0.180**	(0.074)	0.269**	(0.134)
Observations	1188		1261		1261	
P-values:						
Incentives: Not Improved = Improved	0.707		0.332		0.630	
Info-Tool: Not Improved = Improved	0.081		0.217		0.222	

Standard errors in parentheses
Each observation is at the household level. The outcome is the aggregate of the total number of times that chlorine was detected across all visits in the specified time period. All specifications include a control for the number of visits that the household was surveyed during the specified time period, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). Improved is a binary indicator for if the participant's predicted improvement in health after the beginning of chlorine distribution was above the median. See Section 6 for a detailed explanation for how the predicted-health-improvement measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table D.3: Chlorine Detection: Aggregate where "Spillover" is exposure to any Info-tool neighbor predicted to improve

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × Spillover	-0.091	(0.097)	0.033	(0.085)	0.196	(0.153)
Incentives × No Spillover	0.130**	(0.064)	-0.029	(0.058)	0.227**	(0.104)
Incentives × Spillover	-0.054	(0.095)	-0.016	(0.086)	-0.016	(0.155)
Info-Tool × No Spillover	-0.042	(0.065)	0.023	(0.059)	0.074	(0.106)
Info-Tool × Spillover	-0.028	(0.097)	0.191**	(0.086)	0.403***	(0.155)
Observations	1188		1261		1261	
P-values:						
Incentives: No Spillover = Spillover	0.054		0.874		0.118	
Info-Tool: No Spillover = Spillover	0.883		0.055		0.035	

Standard errors in parentheses
Each observation is at the household level. The outcome is the aggregate of the total number of times that chlorine was detected across all visits in the specified time period. All specifications include a control for the number of visits that the household was surveyed during the specified time period, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to an Info-Tool neighbor (someone within 20m) whose health was predicted to improve than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined. See Section 6 for a detailed explanation for how the predicted-health-improvement measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table D.4: Chlorine Detection: Aggregate
where "Spillover" is exposure to any treatment neighbor predicted to improve

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine × Spillover	-0.006	(0.084)	0.032	(0.075)	0.163	(0.134)
Incentives × No Spillover	0.185**	(0.074)	-0.039	(0.067)	0.269**	(0.120)
Incentives × Spillover	0.006	(0.080)	-0.001	(0.072)	0.083	(0.130)
Info-Tool × No Spillover	-0.054	(0.073)	0.034	(0.066)	0.168	(0.118)
Info-Tool × Spillover	0.031	(0.083)	0.125*	(0.076)	0.180	(0.136)
Observations	1188		1261		1261	
P-values:						
Chlorine: No Spillover = Spillover	0.939		0.670		0.225	
Incentives: No Spillover = Spillover	0.028		0.608		0.162	
Info-Tool: No Spillover = Spillover	0.313		0.231		0.930	

Standard errors in parentheses
Each observation is at the household level. The outcome is the aggregate of the total number of times that chlorine was detected across all visits in the specified time period. All specifications include a control for the number of visits that the household was surveyed during the specified time period, neighborhood block fixed effects, unbalanced baseline controls, the total number of study participants within twenty meters, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months). The Spillover Sample is defined as participants who were randomly more exposed to any treatment neighbor (someone within 20m) who was predicted to have their health improve than they would be in expectation based on endogenous spatial factors. See Section 5 for a detailed explanation of how the measure is defined. See Section 6 for a detailed explanation for how the predicted-health-improvement measure is defined.
* $p < .1$, ** $p < 0.05$, *** $p < 0.01$

Table D.5: Chlorine Detection: Aggregate

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine Only	0.641***	(0.050)	0.578***	(0.046)	0.933***	(0.090)
Incentives	0.743***	(0.049)	0.545***	(0.046)	1.085***	(0.089)
Info-Tool	0.622***	(0.050)	0.647***	(0.046)	1.059***	(0.089)
Observations	1599		1661		1653	
P-values:						
Chlorine = Incentives	0.039		0.490		0.091	
Chlorine = Info-Tool	0.706		0.139		0.162	
Incentives = Info-Tool	0.015		0.029		0.770	

Standard errors in parentheses
Each observation is at the household level. The outcome is the aggregate of the total number of times that chlorine was detected across all visits in the specified time period. All specifications include a control for the number of visits that the household was surveyed during the specified time period, neighborhood block fixed effects, unbalanced baseline controls, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.6: Diarrhea: Aggregate

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Chlorine Only	-0.868***	(0.278)	-0.264*	(0.157)	-0.807***	(0.265)
Incentives	-0.980***	(0.276)	-0.103	(0.156)	-0.444*	(0.262)
Info-Tool	-0.696**	(0.279)	-0.334**	(0.156)	-0.775***	(0.262)
Observations	1599		1661		1653	
Control Mean	2.505		0.949		2.348	
P-values:						
Chlorine = Incentives	0.687		0.307		0.172	
Chlorine = Info-Tool	0.542		0.661		0.903	
Incentives = Info-Tool	0.309		0.144		0.208	

Standard errors in parentheses
Each observation is at the household level. The outcome is a continuous measure of the total child-days of diarrhea that the household reported over the entirety of the short-run, medium-run, or long-run period (aggregated across all children). All specifications include a control for the number of visits that the household was surveyed during the specified time period, neighborhood block fixed effects, unbalanced baseline controls, the number of children under five the household had at baseline, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.7: IV Impacts on Endline Child Health (Instrument: Info-tool × Spillover) where "Spillover" is exposure to any Info-tool neighbor

	(1) Index (Anthro-pometry)	(2) Height-for-Age	(3) Weight-for-Height	(4) Weight-for-Age	(5) MUAC-for-Age
Boils, Bleaches, or Chlorinates Water	0.506***	-0.083	0.584	0.789**	-0.428
	(0.183)	(0.376)	(0.516)	(0.398)	(0.337)
Observations	2616	2371	2439	2492	1954
Endline Control Mean	-0.019	-1.773	-0.291	-1.407	-1.453
Weak-IV robust F statistic	35.26	34.36	33.85	36.26	27.02
C-statistic p-value	0.002	0.986	0.471	0.062	0.379

Standard errors in parentheses
Assignment to the Info-Tool treatment *and* randomly being exposed to an Info-Tool neighbor within 20 meters is an instrument for if the respondent reported at endline that she boils, bleaches, or chlorinates her water. Child-level cross-section of the endline survey. Standard errors are clustered at the household level. All regressions include baseline unbalanced household-level controls, baseline child anthropometrics and diarrhea rates, child gender, child age, and neighborhood block fixed effects. Indeces are constructed using following Anderson (2008).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.8: Chlorine Acceptance

	(1) Short-Run		(2) Medium-Run		(3) Long-Run	
Incentives	0.069***	(0.023)	0.080***	(0.024)	0.063**	(0.028)
Info-Tool	0.020	(0.024)	0.016	(0.026)	0.023	(0.029)
Observations	5545		4609		10455	
Chlorine-only Mean	0.794		0.756		0.740	

Standard errors in parentheses
Each observation is a household-level survey visit. The outcome is a binary indicator for whether or not the respondent accepted free chlorine during the survey visit. All specifications include survey-round fixed effects, neighborhood block fixed effects, unbalanced baseline controls, and lasso-selected baseline controls. Short-run is defined as the first three months of chlorine distribution (while the behavioral interventions were ongoing); Medium-run is defined as the next three months of chlorine distribution (immediately after the behavioral interventions ended); and Long-run is defined as the remainder of the trial (nine months).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix E Baseline Balance: Spillover and Endline Sample

Table E.1: Baseline Balance: Spillover Sample

Variable	(1) No Spillover Mean/SE	(2) Spillover Sample Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Any Child Had Motions	0.295	0.332	0.309	-0.036
	(0.013)	(0.019)	(0.011)	
Reported Highest Diarrhea in Summer	0.817	0.856	0.831	-0.039
	(0.011)	(0.014)	(0.009)	
Number of Children <5	1.491	1.514	1.499	-0.023
	(0.020)	(0.033)	(0.017)	
Has Heard of Chlorine	0.212	0.188	0.204	0.024
	(0.012)	(0.015)	(0.009)	
Would consider using chlorine	0.872	0.890	0.878	-0.019
	(0.010)	(0.012)	(0.008)	
Enumerator Observes Dirt in Water	0.164	0.162	0.163	0.002
	(0.011)	(0.014)	(0.009)	
Reports Dirt in Water	0.750	0.750	0.750	-0.000
	(0.013)	(0.017)	(0.010)	
Boils, Bleaches, or Chlorinates Water	0.143	0.133	0.139	0.010
	(0.010)	(0.013)	(0.008)	
Strains or Filters Water	0.620	0.662	0.635	-0.042
	(0.014)	(0.019)	(0.011)	
Believes Chlorine is for Water Purification	0.217	0.187	0.206	0.030
	(0.012)	(0.015)	(0.010)	
Caretaker Asked about Chlorine Test	0.143	0.164	0.150	-0.021
	(0.010)	(0.015)	(0.008)	
Attrited (endline)	0.125	0.128	0.126	-0.003
	(0.010)	(0.013)	(0.008)	
N	1154	648	1802	
F-test of joint significance (F-stat)				1.110

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are robust. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table E.2: Baseline Balance: Spillover Sample (Info-tool Only)

Variable	(1) Info-Tool (no IT neighbor) Mean/SE	(2) Info-Tool (with IT neighbor) Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Any Child Had Motions	0.253	0.358	0.291	-0.105**
	(0.026)	(0.037)	(0.021)	
Reported Highest Diarrhea in Summer	0.793	0.867	0.820	-0.074
	(0.024)	(0.027)	(0.018)	
Number of Children <5	1.498	1.521	1.507	-0.023
	(0.038)	(0.076)	(0.037)	
Has Heard of Chlorine	0.235	0.248	0.240	-0.013
	(0.025)	(0.034)	(0.020)	
Would consider using chlorine	0.867	0.897	0.878	-0.030
	(0.020)	(0.024)	(0.015)	
Enumerator Observes Dirt in Water	0.172	0.127	0.156	0.045
	(0.022)	(0.026)	(0.017)	
Reports Dirt in Water	0.754	0.745	0.751	0.009
	(0.026)	(0.034)	(0.020)	
Boils, Bleaches, or Chlorinates Water	0.126	0.182	0.147	-0.056
	(0.020)	(0.030)	(0.017)	
Strains or Filters Water	0.618	0.661	0.633	-0.043
	(0.029)	(0.037)	(0.023)	
Believes Chlorine is for Water Purification	0.239	0.182	0.218	0.057
	(0.025)	(0.030)	(0.019)	
Caretaker Asked about Chlorine Test	0.123	0.182	0.144	-0.059
	(0.019)	(0.030)	(0.017)	
Attrited (endline)	0.105	0.127	0.113	-0.022
	(0.018)	(0.026)	(0.015)	
N	285	165	450	
F-test of joint significance (F-stat)				1.382

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are robust. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table E.3: Baseline Balance: Spillover Sample (Incentives Only)

Variable	(1) Incentives (no IT neighbor) Mean/SE	(2) Incentives (with IT neighbor) Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Any Child Had Motions	0.315	0.285	0.304	0.030
	(0.028)	(0.035)	(0.022)	
Reported Highest Diarrhea in Summer	0.836	0.806	0.825	0.030
	(0.022)	(0.031)	(0.018)	
Number of Children <5	1.486	1.467	1.479	0.019
	(0.040)	(0.050)	(0.031)	
Has Heard of Chlorine	0.220	0.145	0.193	0.075*
	(0.025)	(0.028)	(0.019)	
Would consider using chlorine	0.895	0.873	0.887	0.022
	(0.018)	(0.026)	(0.015)	
Enumerator Observes Dirt in Water	0.185	0.152	0.173	0.034
	(0.023)	(0.028)	(0.018)	
Reports Dirt in Water	0.773	0.691	0.743	0.082
	(0.025)	(0.036)	(0.021)	
Boils, Bleaches, or Chlorinates Water	0.140	0.115	0.131	0.025
	(0.021)	(0.025)	(0.016)	
Strains or Filters Water	0.612	0.630	0.619	-0.018
	(0.029)	(0.038)	(0.023)	
Believes Chlorine is for Water Purification	0.234	0.194	0.220	0.040
	(0.025)	(0.031)	(0.020)	
Caretaker Asked about Chlorine Test	0.133	0.152	0.140	-0.019
	(0.020)	(0.028)	(0.016)	
Attrited (endline)	0.154	0.139	0.149	0.014
	(0.021)	(0.027)	(0.017)	
N	286	165	451	
F-test of joint significance (F-stat)				0.776

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are robust. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table E.4: Baseline Balance: Spillover Sample (Chlorine Only)

Variable	(1) Chlorine (no IT neighbor) Mean/SE	(2) Chlorine (with IT neighbor) Mean/SE	(3) Total Mean/SE	T-test Difference (1)-(2)
Any Child Had Motions	0.338	0.388	0.356	-0.050
	(0.028)	(0.038)	(0.023)	
Reported Highest Diarrhea in Summer	0.835	0.861	0.844	-0.026
	(0.022)	(0.027)	(0.017)	
Number of Children <5	1.475	1.588	1.517	-0.113
	(0.042)	(0.071)	(0.038)	
Has Heard of Chlorine	0.190	0.200	0.194	-0.010
	(0.023)	(0.031)	(0.019)	
Would consider using chlorine	0.870	0.897	0.880	-0.027
	(0.020)	(0.024)	(0.015)	
Enumerator Observes Dirt in Water	0.137	0.200	0.160	-0.063
	(0.020)	(0.031)	(0.017)	
Reports Dirt in Water	0.739	0.776	0.753	-0.036
	(0.026)	(0.033)	(0.020)	
Boils, Bleaches, or Chlorinates Water	0.144	0.109	0.131	0.035
	(0.021)	(0.024)	(0.016)	
Strains or Filters Water	0.641	0.685	0.657	-0.044
	(0.029)	(0.036)	(0.022)	
Believes Chlorine is for Water Purification	0.180	0.230	0.198	-0.051
	(0.023)	(0.033)	(0.019)	
Caretaker Asked about Chlorine Test	0.165	0.194	0.176	-0.028
	(0.022)	(0.031)	(0.018)	
Attrited (endline)	0.123	0.158	0.136	-0.034
	(0.020)	(0.028)	(0.016)	
N	284	165	449	
F-test of joint significance (F-stat)				0.636

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are robust. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table E.5: Baseline Balance: Adult Outcomes (Endline Sample)

Variable	(1) Control Mean/SE	(2) Chlorine-tablets Mean/SE	(3) Incentives Mean/SE	(4) Info-tool Mean/SE	(5) Total Mean/SE	T-test Difference (1)-(2)	(1)-(3)	(1)-(4)
Any Child Had Motions	0.285	0.345	0.292	0.278	0.300	-0.061*	-0.007	0.006
	(0.022)	(0.024)	(0.023)	(0.022)	(0.012)			
Reported Highest Diarrhea in Summer	0.837	0.832	0.820	0.835	0.831	0.004	0.016	0.002
	(0.018)	(0.019)	(0.020)	(0.019)	(0.009)			
Number of Children <5	1.493	1.497	1.479	1.511	1.495	-0.005	0.013	-0.019
	(0.035)	(0.039)	(0.034)	(0.039)	(0.018)			
Has Heard of Chlorine	0.200	0.204	0.201	0.238	0.211	-0.003	-0.000	-0.038
	(0.020)	(0.020)	(0.020)	(0.021)	(0.010)			
Would consider using chlorine	0.866	0.879	0.880	0.872	0.874	-0.013	-0.014	-0.006
	(0.017)	(0.017)	(0.017)	(0.017)	(0.008)			
Enumerator Observes Dirt in Water	0.146	0.155	0.169	0.158	0.157	-0.009	-0.023	-0.012
	(0.018)	(0.018)	(0.019)	(0.018)	(0.009)			
Reports Dirt in Water	0.748	0.747	0.758	0.752	0.751	0.000	-0.010	-0.004
	(0.022)	(0.022)	(0.022)	(0.022)	(0.011)			
Boils, Bleaches, or Chlorinates Water	0.141	0.121	0.130	0.148	0.135	0.020	0.011	-0.007
	(0.017)	(0.017)	(0.017)	(0.018)	(0.009)			
Strains or Filters Water	0.651	0.660	0.622	0.639	0.643	-0.009	0.029	0.012
	(0.024)	(0.024)	(0.025)	(0.024)	(0.012)			
Believes Chlorine is for Water Purification	0.198	0.196	0.232	0.216	0.210	0.002	-0.034	-0.018
	(0.020)	(0.020)	(0.022)	(0.021)	(0.010)			
Caretaker Asked about Chlorine Test	0.136	0.173	0.143	0.140	0.148	-0.037	-0.007	-0.004
	(0.017)	(0.019)	(0.018)	(0.017)	(0.009)			
N	404	388	384	399	1575			
F-test of joint significance (F-stat)						0.640	0.606	0.256

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are robust. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table E.6: Baseline Balance: Child Outcomes (Endline Sample)

	(1) Control		(2) Chlorine-tablets		(3) Incentives		(4) Info-tool		T-test Difference		
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)	(1)-(3)	(1)-(4)
Female	572	0.479	541	0.457	541	0.481	565	0.485	0.022	-0.002	-0.006
	[386]	(0.022)	[370]	(0.024)	[370]	(0.022)	[383]	(0.022)			
Age (in months)	568	30.885	539	30.638	539	30.226	564	31.599	0.247	0.659	-0.715
	[384]	(0.584)	[370]	(0.563)	[368]	(0.600)	[383]	(0.605)			
Child Weight (kg)	568	10.562	539	10.581	539	10.674	564	10.691	-0.019	-0.112	-0.130
	[384]	(0.119)	[370]	(0.122)	[368]	(0.129)	[383]	(0.120)			
Child MUAC (cm)	572	14.345	540	14.274	542	14.317	565	14.395	0.072	0.028	-0.050
	[386]	(0.053)	[370]	(0.049)	[371]	(0.057)	[383]	(0.058)			
Number of Motion Days	572	0.769	541	0.852	542	0.769	565	0.796	-0.083	-0.001	-0.027
	[386]	(0.074)	[370]	(0.080)	[371]	(0.081)	[383]	(0.084)			
Child Had > 0 Motion Days	572	0.215	541	0.262	542	0.218	565	0.219	-0.047*	-0.003	-0.004
	[386]	(0.018)	[370]	(0.020)	[371]	(0.018)	[383]	(0.018)			

Notes: The value displayed for t-tests are the differences in the means across the groups. Standard errors are clustered at variable HHID. Fixed effects using variable block are included in all estimation regressions. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Appendix F Photos of Intervention Materials

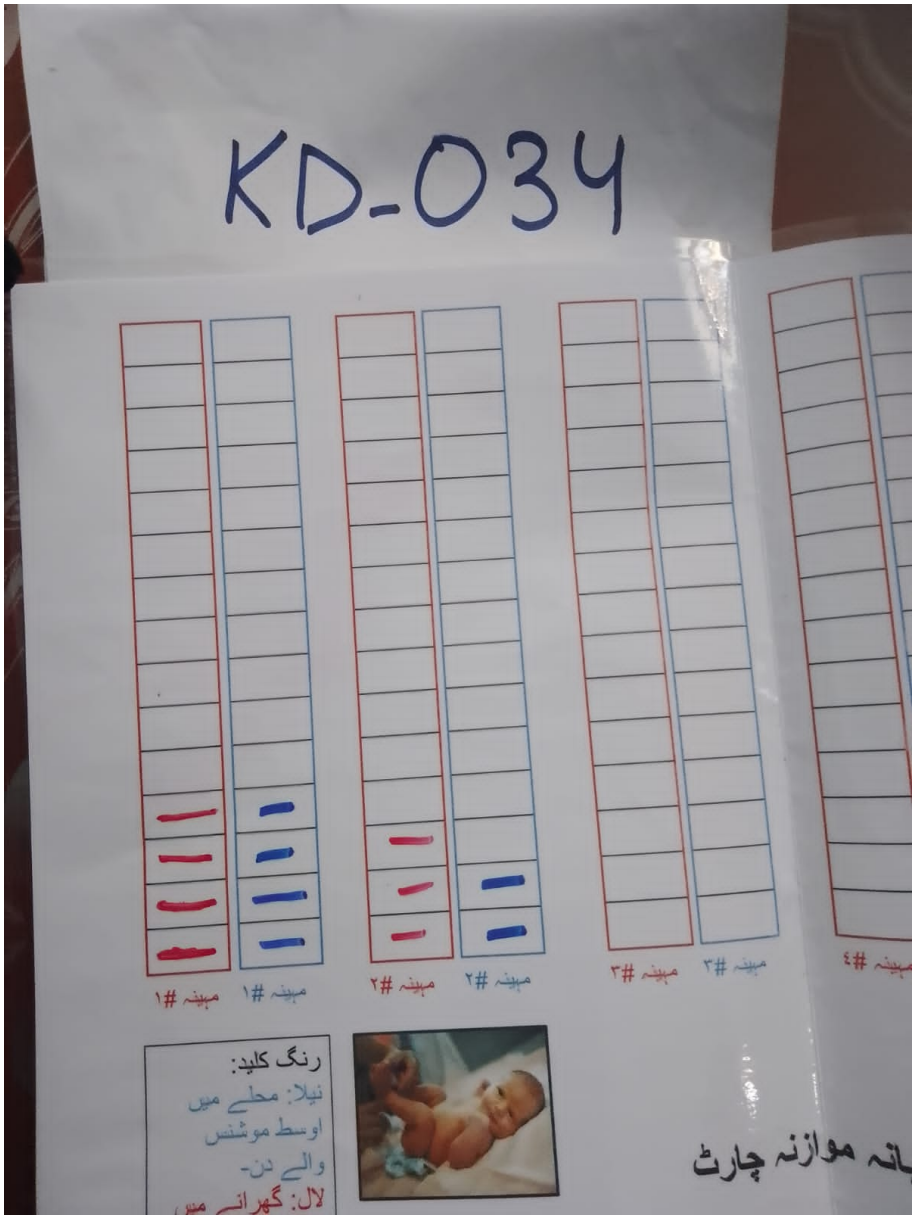


Figure F.1: Experiential Learning Intervention: Info-Tool

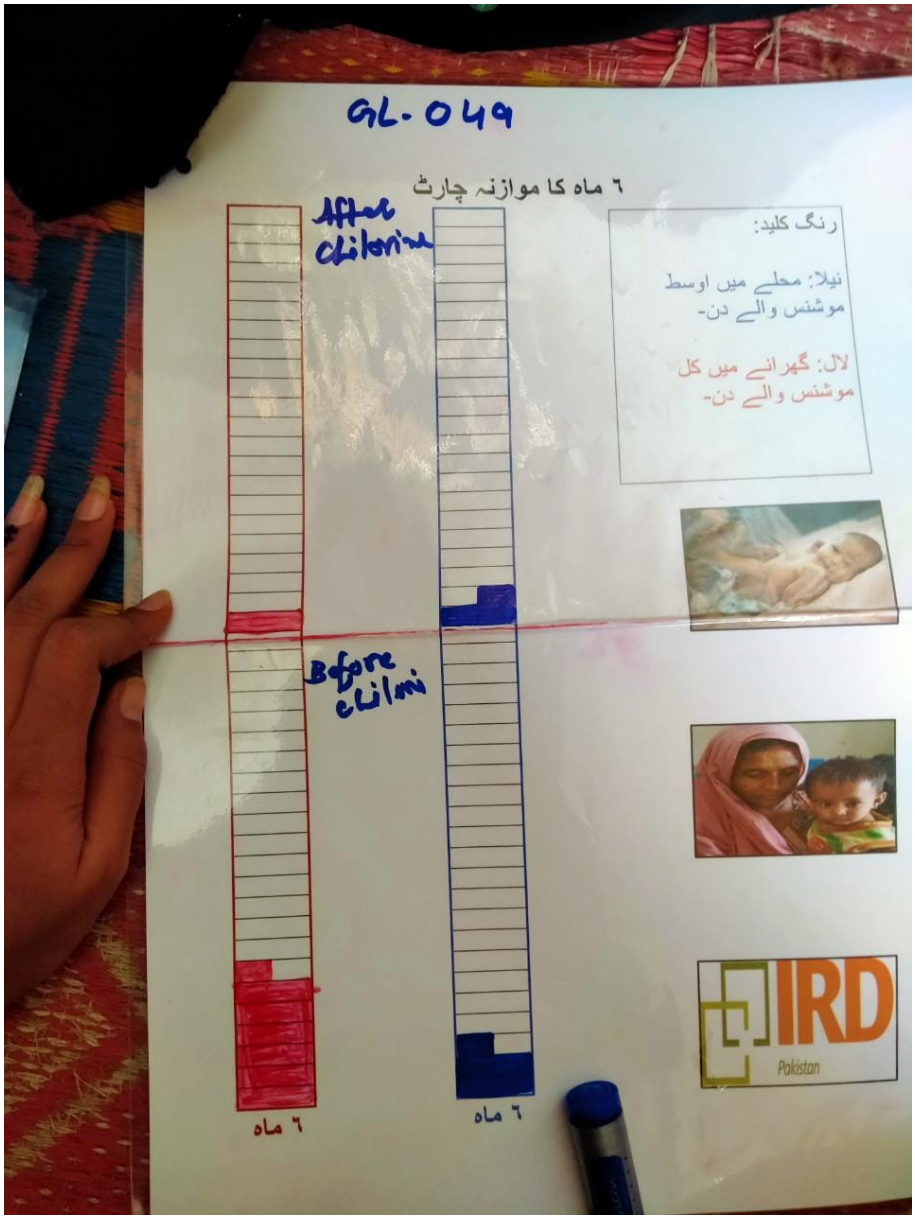


Figure F.2: Experiential Learning Intervention: Info-Tool (Difference-in-Differences Comparison)



Figure F.3: Sample Water Vessel

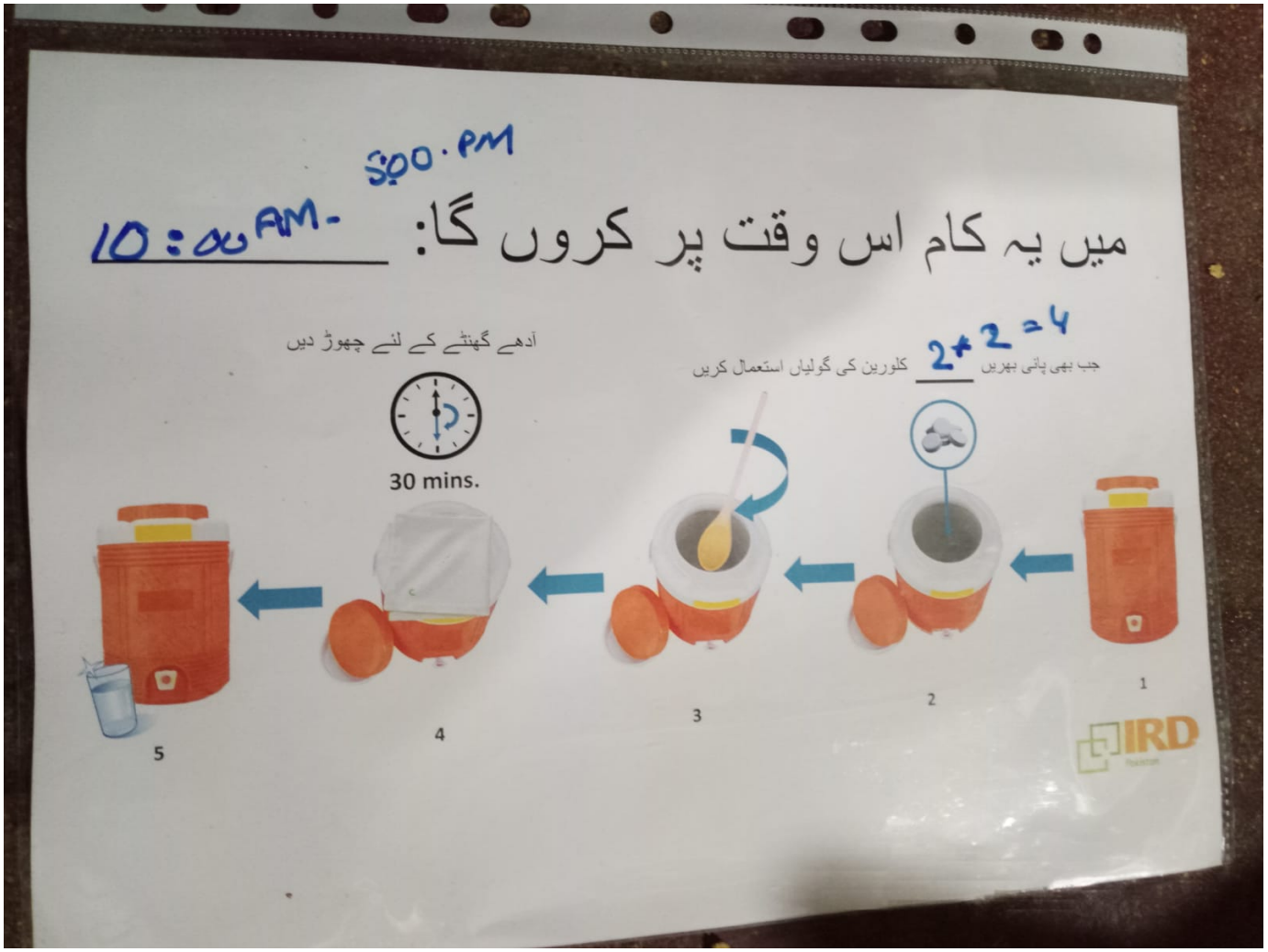


Figure F.4: Leaflet with Instructions on Chlorination

Appendix G Model Simulation

Under a reasonable set of parameters, our model generates dynamic predictions that match our empirical data. We simulate individuals' learning and technology adoption over sixteen periods, where the first four periods represent the "treatment period" (the behavioral interventions), and the following twelve periods represent the "post-treatment period" (free chlorine distribution in all groups). First, we simulate a dataset of 1350 individuals and divide them into three groups (450 per group, as in our experimental design). All three groups undergo the same learning and adoption process, with the following exceptions: one group, modeled on the Incentives group, faces lower costs of technology adoption in the first four periods; and another group, modeled on the Info-Tool group, places heavier action-weights on the signals observed through the Info-Tool in the first four periods (whether those signals are their own or their Info-Tool friends').

G.1 Simulated Data Set-Up

Individuals have a prior belief about chlorine efficacy, μ_0, which is distributed in the population as: $\mu_0 \sim \mathcal{N}(.3, .3)$. We let variance $\sigma_0^2 = 0.3$ be constant in the population. We assume that prior beliefs are largely positive and certain because most participants' only information about chlorine is from the CHW, who sends a positive signal about chlorine and represents a trusted information source.

In every period where participants adopt chlorine, they receive a signal about chlorine efficacy from nature $Y \sim \mathcal{N}(0.58, 0.49)$ in the population. This is modeled on the experimental sample mean and standard deviation of the probability that participants' diarrhea rate decreased in the three months after chlorine distribution relative to the three months before. We assume that every participant tries chlorine at least once in period 0 (which represents "day 1" of the treatment period), so that every participant begins period 1 with a signal Y. We model participants' uncertainty about this belief with $\sigma_Y^2 = 1$. Participants apply an *action-weight* to this belief, $\alpha_Y = 0.3$.

Social Learning

We randomly assign friends using $Pr(F = 1) = 0.0015$, sampling with replacement. On average, participants have 2.7 friends each (maximum = 9, minimum = 0), and 0.6 Info-Tool friends each (maximum = 4, minimum = 0). In each period, participants talk with their friends about chlorine with probability $Pr(T = 1) = 0.2$. In period t, the friend sends a signal, which is the average of the signals about chlorine that they received in the previous three periods $Y_{\bar{t}}$:

$$Y_{\bar{t}} = \frac{\sum_{p=t-2}^{p=t} Y_p \cdot \mathbb{1}(M_\alpha^p > C^p)}{\sum_{p=t-2}^{p=t} \mathbb{1}(M_\alpha^p > C^p)}$$

This means that, if a friend has not adopted chlorine in the preceding three periods, they do not share information about chlorine, even if they talk. When participants learn from themselves

(signals from nature), they only update with the information they gain in this particular period, Y_t. Different from friends, who we assume only learn from each others' signals if they talk (which happens with some random probability), individuals will *always* incorporate information that they generate if they adopt chlorine. In other words, there is no analog for randomly talking in the way that people incorporate their own signals.

Technology Adoption

In each period, i incorporates signals into her prior (one at a time) from herself and her neighbors. After incorporating all the new signals in the time period, she decides to adopt chlorine if $M_\alpha > C$, where we assume that $C = 0.6$. Recall that, if she does *not* adopt chlorine, she will not gain any new information about chlorine in the next time period. However, she can still learn from her friends about chlorine (if they talk in that time period, and if her friend has adopted chlorine in the past three time periods).

Interventions

For 450 individuals, we let $C = 0.5$ in the first four time periods (the Incentives group). For another 450 individuals, we let $\alpha_Y = 1$ for signals that are acquired with the Info-Tool (signals they receive from themselves in the first four periods, or signals they receive from others in the same group in the first 7 periods, since individuals share prior information up to three periods later). The remaining 450 participants undergo "standard" learning, and represent the Chlorine Only group.

G.2 Dynamic Simulations

Summary of base parametric assumptions:

$$\begin{aligned}
\mu_0 &\sim \mathcal{N}(0.3, 0.3) \\
\sigma_0 &= 0.3 \\
Y &\sim \mathcal{N}(0.58, 0.49) \\
\sigma_Y^2 &= 1 \\
\alpha_Y &= 0.3 \\
C &= 0.6
\end{aligned}$$

Summary of interventions:

$$\begin{aligned}
\text{Incentives: } & C = 0.5, \forall t <= 4 \\
\text{Info-Tool: } & \alpha_Y = 1, \forall Y_{t<=4 \ \& \ i=j} \\
\text{Info-Tool: } & \alpha_Y = 1, \forall Y_{t<=7 \ \& \ j \in \{IT\} \ \& \ i \neq j}
\end{aligned}$$

Our model generates a pattern of adoption that resembles our raw data (Figure G.1).

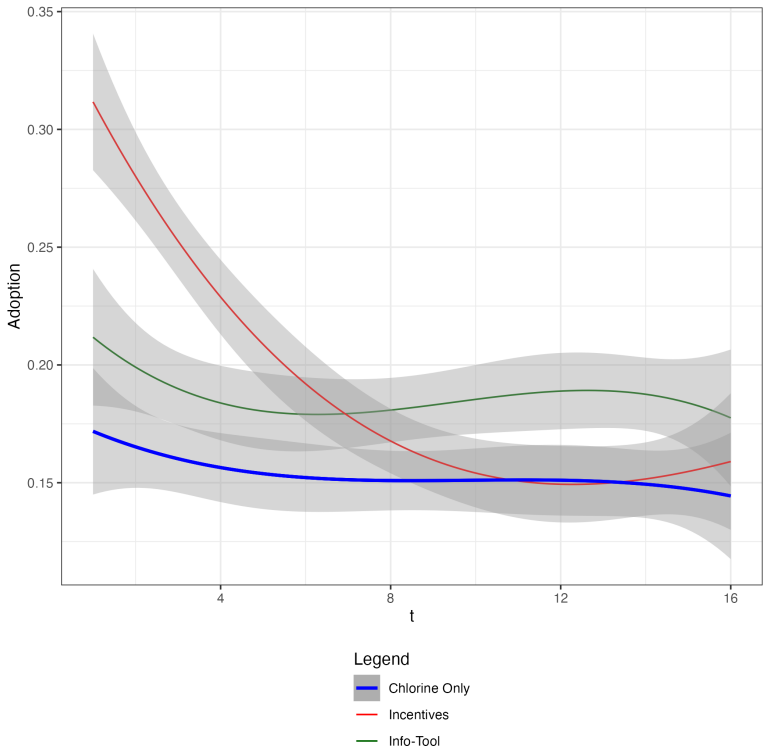


Figure G.1: Model Simulation: Chlorine Adoption Over Time

We observe a positive and *increasing* relationship between the number of Info-Tool friends and chlorine adoption within the Info-Tool group. However, this relationship is not present in the other groups. (Figure G.2).

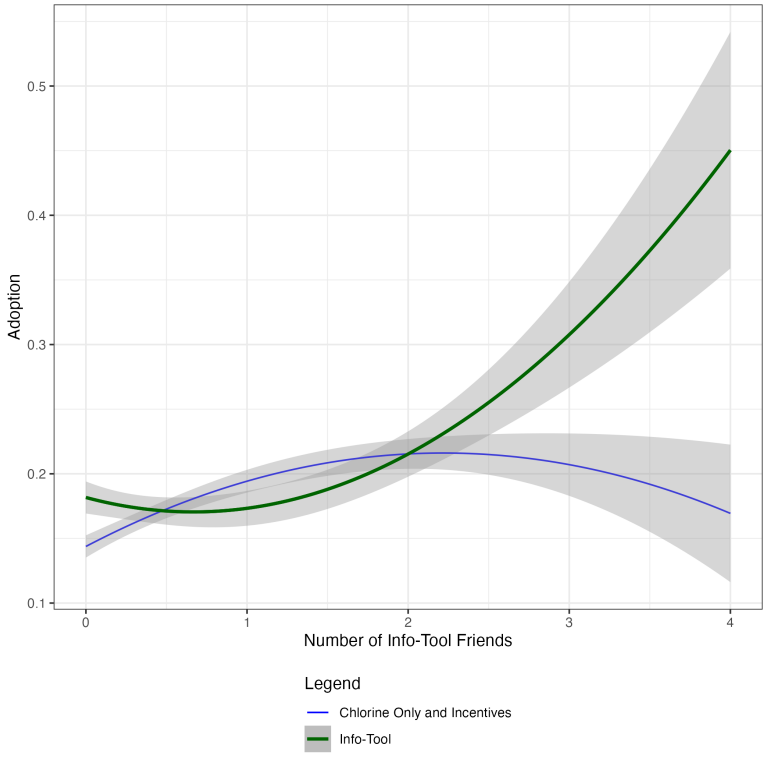


Figure G.2: Model Simulation: Chlorine Adoption as a Function of Info-Tool Friends

While adoption dynamics differ by intervention groups, and according to the number of Info-Tool neighbors, *stated* posterior beliefs M (i.e. explicit knowledge) do not (Figures G.3 and G.4).

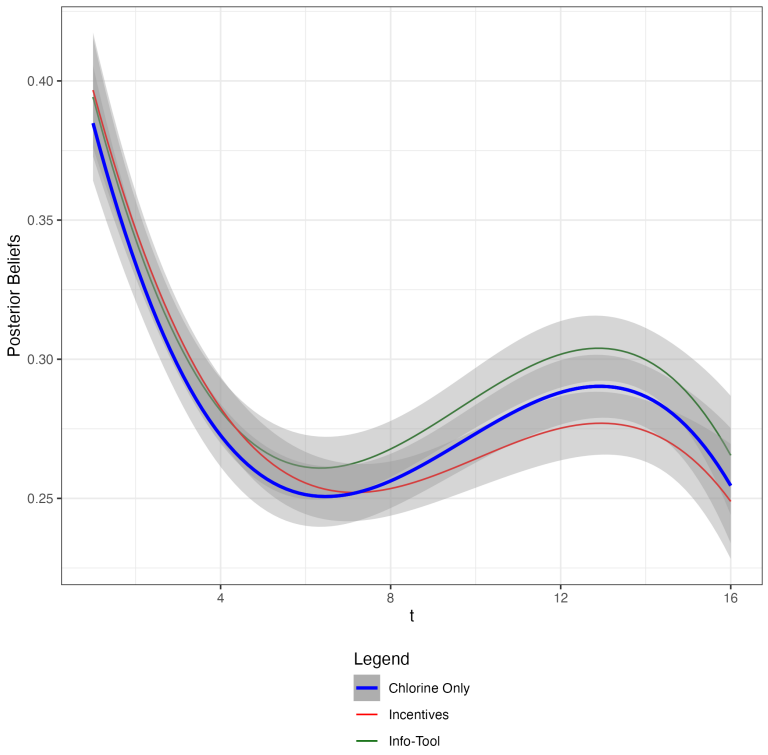


Figure G.3: Model Simulation: Stated Posterior Beliefs Over Time

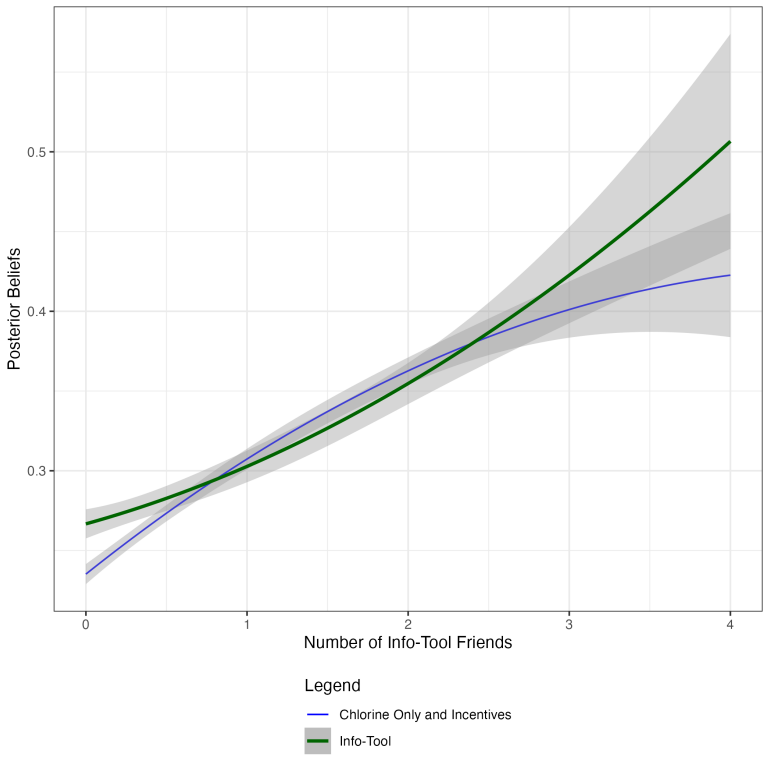


Figure G.4: Model Simulation: Stated Posterior Beliefs as a Function of Info-Tool Friends

Appendix H AEA Registry Analysis

H.1 Background

Researchers identify or negate cross-treatment spillovers through two primary methods: analyzing the treatment effect of distance (geographic or social) to treated units among control units, and through cluster-randomized designs. In designs with random variation in distance to treated units among control units, cross-treatment spillovers are identified but not neutralized (designs such as Miguel and Kremer (2004)). In cluster-randomized designs where every unit is treated in some clusters and no units are treated in other clusters, cross-treatment spillovers are neutralized but not identified (designs that randomize at a group level such as school-level randomization in Muralidharan and Sundararaman (2011)). Designs that randomize treatment saturation across clusters, and then randomize individuals to treatment or control within clusters according to the predetermined ratio, are cluster-randomized designs that effectively randomize individuals' distance to treated units (designs such as Egger et al. (2022)). Concerns about the inability to detect true treatment effects due to cross-treatment spillovers have led cluster-randomized designs to be highly prevalent in economics randomized controlled trials, especially within development economics.

H.2 Methods

We downloaded the AEA RCT Registry data on November 11, 2024 (*AEA RCT Registry Dataverse*, 2024). Since there is no single variable indicating the randomization procedure, we utilized ChatGPT 4o to help conduct a text analysis of text-response variables.

Cluster-randomized versus individually-randomized trials: We identify individually-randomized trials from cluster-randomized trials using the "randomization unit", "sample size number of clusters", and "sample size number of observations" variables. The assignment rules were as follows.

ClusteredTrial = 0 if:

1. "Sample size number clusters" is equal to any of the following (ignoring cases and punctuation): "no cluster", "no clusters", "NA", "none", "same as observations", "no", "individual", or is blank
2. "Sample size number clusters" includes any of the following phrases (ignoring cases and punctuation): "treatment is not clustered", "no clustering", "there are no clusters", "equals the total number of participants", "not clustered", "not applicable", "treatment will not be clustered", "there are no planned clusters", "this is not cluster-randomized", "this is not a cluster-randomized", "NA ", "N/A ", "does not have clustered randomization", "does not cluster randomize", "there is no need for clustering", or "no cluster"
3. "Sample size number of clusters" = "Sample size number observations"

4. "Randomization unit" is equal to any of the following (ignoring cases and punctuation): "individual", "we will randomize individuals", "individual is the unit of randomization", "individual level", "individual randomization for treatments", "treatments are randomized within-session at the individual level", "randomization takes part on the participant level", "individual survey participant", "individual worker", "individual participant", "the randomization was done using stratified random sampling at the individual level", or "randomization will be done at the individual level"

5. "Randomization unit" includes any of the following (ignoring cases and punctuation): "without clustering"

6. The only difference between the values in "Randomization Unit" and "Sample size number of observations" is that "Sample size number of observations" has a number that is not present in "Randomization Unit"

7. The number in "Sample size number of clusters" is the same as the number in "Sample size number of observations" and "Randomization unit" includes the word "individual" and "Randomization Unit" does not include the word "cluster", and "Sample size number of clusters" does not include any words (only numbers)

8. The number in "Sample size number of clusters" is the same as the number in "Sample size number of observations" and "Randomization unit" does not include the word "cluster", and "Sample size number of observations" does not include any words

Among cases where ClusteredTrial $\neq 0$, ClusteredTrial $= 1$ if :

1. "Sample size number of clusters" includes both a number and a unit (for example, "300 schools", "100 households" or "50 villages")

2. "Experimental design" or "Experimental design details" include the words "medium treatment saturation", "high treatment saturation", or "low treatment saturation"

3. "Experimental design" or "Experimental design details" include the words "% saturation" or "% treatment saturation"

4. There is a number included in both the "Sample size number of observations" variable and the "Sample size number of clusters" variable, and these numbers are not the same

5. "Sample size number of clusters" and "Sample size number of observations" are both just numbers, and "Sample size number of clusters" is a smaller number than "Sample size number of observations"

After implementing these rules, we are able to classify 96.8% of trials (the treatment design of 304 out of 9494 trials remain unidentified).

Full- vs. partial-treatment saturation: We then classify a trial as cluster-randomized with varying treatment saturation across clusters if the the trial is classified as cluster-randomized, and if

the registration ever uses terms related to "treatment saturation" or "two-stage randomization" across any variable.

KeywordSat = 1 for observations that mention any word stemming with "saturat" across any variable. Otherwise, KeywordSat = 0.

KeywordTwoStage = 1 if any variable includes the phrase "two stage randomization", "two-stage randomization", "2-stage randomization" or "2 stage randomization". Otherwise, KeywordTwoStage = 0.

We consider a trial to be cluster-randomized with varying treatment saturation if ClusteredTrial = 1 and either KeywordSat =1 or KeywordTwoStage = 1.

Research Topics: For each trial, researchers select "keywords" associated with their trial. First, we make minor cleaning adjustments for cohesion across variations in specific keywords that people use:

1. "behavior" includes: "behavioral", "behavioral economics"
2. "environment" includes: "climate change", "environment and energy"
3. "health" includes: "covid-19", "mental health", "nutrition"
4. "crime violence and conflict" includes: "post-conflict"
5. "experiment" includes: "rct", "online experiment", "survey experiment", "field experiment", "lab"
6. "firms" includes: "firms and productivity"
7. "productivity" includes: "firms and productivity"

Next, we identify all of the keywords that are associated with at least fifty trials (we do not analyze keywords that are excessively niche). This leaves us with thirty-nine keywords to analyze.

H.3 Results: Prevalence of Clustered Randomization

Among all projects registered on the American Economics Association RCT Registry, 35% of trials in high-income countries (HICs) featured cluster-randomized designs, whereas 62% of trials in low- and middle-income countries (LMICs) featured cluster-randomized designs. Of trials that did not specify or publish the country, 48% were cluster-randomized. We used the World Bank classification of country income.

Of all the cluster-randomized trials registered to the American Economics Association RCT Registry, only 2.3% were classified as cluster-randomized with varying treatment saturation (98 trials total). An additional 26 trials met the criteria based on the keywords "treatment saturation" or "two-stage randomized", but were individually randomized trials (a case of type

1 error). Of the cluster-randomized trials with varying treatment saturation, 38% were within LMICs, 9% were within HICs, and 53% did not specify or publish the country.

H.4 Results: By Topic

For which questions do researchers implement cluster-randomized designs, especially those with varying treatment saturation? To analyze the prevalence of research topics across experimental designs, we use the following regression for each keyword K:

$$K_t = \beta_0 + \beta_1 C_t + \beta_2 S_t + \beta_3 C_t \times S_t + \gamma_t + \epsilon_t \tag{3}$$

where $K_t = 1$ if trial t was registered with keyword K, and 0 otherwise; $C_t = 1$ if the trial is cluster-randomized; $S_t = 1$ if the trial included the terms "treatment saturation" or "two-stage randomization" anywhere in the registration; and γ_t is a world-region fixed effect (HIC, LMIC, or unspecified). Then β_1 is the probability that a fully-saturated cluster-randomized trial is associated with keyword K relative to individually-randomized trials; and $\beta_2 + \beta_3$ is the probability that a cluster-randomized trial with varying treatment saturation is associated with keyword K relative to fully-saturated cluster-randomized trials.

Keywords that are *more* likely to be associated with fully-saturated cluster-randomized trials, relative to individually-randomized trials, include "agriculture", "cash transfers", "communication", "early childhood development", "education", "gender", "health", "incentives", "poverty", and "technology adoption". Keywords that are *less* likely to be associated with fully-saturated cluster-randomized trials, relative to individually-randomized trials, include "altruism", "behavior", "beliefs", "electoral", "fairness", "inequality", "labor", "migration", "redistribution", "social media", "trust", and "other". There are no statistically significant differences in the probability of individual- or clustered-randomization among the remaining keywords.[46]

Keywords that are more likely to be associated with cluster-randomized trials with varying treatment saturation, relative to fully saturated cluster-randomized trials, are "agriculture" and "migration" (there are no statistically distinguishable differences across any other keywords).

Keywords Associated with Our Trial

We analyze the prevalence of treatment designs across trials that use top keywords that we group together as "behavioral", "health", "technology adoption", or "communication". Behavioral trials are most likely individually randomized, and, when cluster-randomized, almost always with full treatment saturation. Health, technology adoption, and communication trials are very likely to be cluster-randomized, but usually with full treatment saturation.

Cluster-randomized trials with full saturation are 11pp $(24\%, p < 0.001)$ *less* likely to be associ-

[46] Keywords with no differences in treatment design: "cooperation", "crime violence and conflict", "discrimination", "entrepreneurship", "environment", "experiment", "finance", "financial literacy", "firms", "governance", "information", "productivity", "savings", "social norms", "taxation", "training", and "welfare".

ated with a top behavioral economics keyword ("behavior", "beliefs", "fairness", or "altruism") than individually-randomized trials. Cluster-randomized trials with varying treatment saturation are another 8pp ($25\%, p = 0.099$) less likely to include a top behavioral economics keyword, relative to fully saturated cluster-randomized trials.

Cluster-randomization with full saturation – the design that allows for the strongest within-treatment spillovers *but conceals them* – is the most common experimental design researchers implement for economics trials focused on technology adoption, health, or communication. Technology adoption trials are 0.7pp ($163\%, p < 0.001$) more likely to be cluster-randomized than individually randomized; health trials are 1.9pp ($12\%, p = 0.020$) more likely to be cluster-randomized than individually randomized; and communication trials are 0.5pp ($55\%, p = 0.024$) more likely to be cluster-randomized than individually randomized.

Appendix I Detailed Literature Review

I.1 Experiential Learning in Health

This paper speaks to a small literature on the value of experiential and social learning in adopting health technology. There exists a large literature investigating how experts themselves learn about health from their own experiences, but this literature is more focused on physician skill development than on belief formation (Halm et al. (2002); Facchini (2022)). While health experts have the information that is most likely to be accurate on average, drawing from clinical trials and personal experiences, there are large inequities in access to experts (Dussault and Franceschini, 2006). Conditional on access to experts, there are still large inequities in how much trust people have in experts. Oftentimes, mistrust is a result of past wrongs committed by the medical community on groups with whom patients identify (Alsan and Wanamaker (2018); Lowes and Montero (2021); Martinez-Bravo and Stegmann (2022)), but there is evidence that knowing that doctors have financial incentives linked with the medical care they provide is enough to sow mistrust (Banerjee et al. (2023)).

Disseminating health information through laypeople or social networks may help solve the problems of trust and access associated with relying on experts to relay health information (Alsan and Eichmeyer (2024),Banerjee et al. (2019)). However, the most socially isolated people may be excluded from information transmission through the social network. Furthermore, the types of information that individuals have access to may differ in quality. Calónico et al. (2023) find that a clinically unsupported treatment for COVID-19 spread through Argentina in a pattern that follows rational learning from neighbors, suggesting that networks can facilitate the spread of medically dubious information. Similarly, Chen et al. (2022) find that access to well-informed networks can generate health inequities. Individuals in Sweden who have a doctor in the family invest more in preventive health, and consequently enjoy healthier and longer lives.

There is scant evidence of experiential learning as a method through which people form beliefs about the efficacy of health inputs and behaviors. Bennett et al. (2018) find that a hygiene course in Pakistan leads to more hygienic behavior when the course includes showing participants microbes under a microscope, suggesting that "seeing is believing." Corno (2014) finds that individuals in rural Tanzania are more likely to seek clinical care if they previously healed after utilizing clinical care, or if they previously *did not* heal after *foregoing* clinical care. Akram and Mendelsohn (2021), whose treatment arm we replicate in our learning arm, find that the Info-tool leads to an increase in water chlorination in the long run. We draw from a large literature in development economics showing that individuals learn about agricultural technologies through their own experiences and their neighbors' experiences (Foster and Rosenzweig, 1994; Hanna et al., 2014; Conley and Udry, 2010). Choosing an agricultural input is similar to choosing a health input in that it is a high-stakes, high-dimensional problem, subject to random shocks and with potential for misattribution.

Globally, 663 million people lack access to an improved source of drinking water while fecal contamination affects 1.8 billion people UNICEF and WHO (2015). Annually, contaminated water contributes to 1.7 billion cases of diarrheal disease and 1.6 million deaths, including half a million under-five children, with most of the disease burden concentrated in developing countries (WHO, 2017). Safe drinking water results in better health, particularly among young children (Kremer et al., 2023; Haushofer et al., 2021), and yields long-term health and cognitive improvements (Scharf et al., 2014). Point-of-use decontamination technologies such as chlorine tablets can drastically reduce the burden of diarrheal disease (Reller et al. (2003), Quick et al. (2002), Quick et al. (1999)).

I.2 Habit Formation versus Learning

Recent literature on long-term behavior change recognizes habit formation, or the generation of complementarities in use across time, as playing a key role in sustained change (e.g., Becker and Murphy (1988), Wellsjo (2021), Celhay et al. (2015), Allcott and Rogers (2014), Royer et al. (2015), Aggarwal et al. (2020)). In both the case of habit formation and learning, higher initial use implies higher future use: in learning, due to initial exposure generating knowledge of the positive returns to a product and increasing the likelihood of future use; and in habit formation, due to intertemporal complementarities in consumption wherein greater initial consumption stock raises future desire to consume. With a few exceptions, however, these studies make no distinction between these two mechanisms, with persistent behavior change in the post-intervention periods often being explained by one mechanism without consideration of the other.

Recognition that these mechanisms may act in concert or be conflated with one another is rather recent in the literature. Caro-Burnett et al. (2021) examine sanitary latrine use in Kenyan slums and isolate the impact of habit formation interventions above and beyond that of learning by holding short-run use constant across treatment arms and examining only long-run behavior change. They find no detectable difference in long-run behavior across arms, though this is likely due to the study context (with sanitary latrine use less amenable to habit formation) and the nature of the interventions themselves (with a time-constrained subsidy intervention intended to embed the behavior in a habit loop, but challenging to do in practice given the potential unpredictability of defecation). Hussam et al. (2022) consider these mechanisms in the context of handwashing in rural West Bengal and attempt to distinguish habit formation from learning about returns by comparing households who experienced the same level of short-run financial incentives, but varied levels of health returns to their behavior. They find that those who experience larger improvements in health (whether between weeks or in aggregate) are no more likely to persist in their handwashing, suggesting that long-run behavior change is not driven by households independently engaging in learning about the value of handwashing from their health experiences. Alpízar et al. (2022) offer a conceptual framework for how these mechanisms may result in long-term adoption (theoretically distinguishing learning how to use a good and learning the returns to use from changes in taste via habit formation), but are

unable to disentangle these channels empirically in an experiment which generates long run use of water-saving technologies through short run incentives to engage.

Yet the distinction between learning and habit formation is critical to policy, as each mechanism implies a substantively different behavior change process and therefore intervention design. Should learning about returns be more effective at generating sustained behavior change or technology adoption, policy design efforts should focus on developing information campaigns that make returns to a behavior explicit. Should habit formation be more effective in motivating long term change, resources may be better spent incentivizing high short-run engagement, yielding long term intertemporal complementarities, and embedding behaviors within habit loops (a la Duhigg (2012) and Neal et al. (2015)).

Preventive health behaviors are often mundane acts that require repetition in order to generate meaningful health impacts - a setting in which habit-formation may be especially relevant. Our experiment allows us to distinguish between the role of learning through salient signals and learning through early adoption (either through habit formation or through the accumulation of more signals) in the long-term adoption of chlorine tablets.

Appendix J Audit Visits

Our main outcome, presence of chlorine residual in drinking water, is a good approximation for consistent use of chlorine tablets *unless* participants are more motivated to use chlorine prior to visits from enumerators, in anticipation of the enumerator's visit. There are several reasons we do not expect this to happen. T3 participants received monetary incentives based on the number of empty chlorine tablet wrappers they could present, rather than based on the presence of residual chlorine in the water. If T3 respondents did not like using chlorine tablets, they could throw away the tablets, present the wrappers, and still receive the prize. Thus, there is no pecuniary motivation for any group to make sure their water presents chlorine residual on the day of the enumerator's visit.

However, there may be social desirability motivations for participants to use chlorine in anticipation of a visit from an enumerator. It may still be difficult for participants to plan perfectly. Chlorine residual is only present for 24 hours, so participants would have to plan to the exact day. Furthermore, enumerators tested for chlorine residual once per month, rather than at every visit, and did not explain the results to participants except in the case of over-chlorination. When possible, enumerators were instructed to take a cup of water outside and test for the presence of chlorine residual without the caregiver observing, to further eliminate the possibility that caregivers expected a reward for water quality.

To be sure that caregivers did not plan chlorine use around enumerator visits, we conducted audit visits throughout the behavioral intervention period (Phase 1 Round 2). In these audit visits, an enumerator arrived unexpectedly to test for the presence of chlorine residual. Audits were conducted by a different enumerator than the enumerator who usually visited the house-

hold. During Phase 1, households were visited once every two weeks but chlorine was only tested once per month. We conducted audit visits during the visit period where water was not tested for chlorine residual, and visited selected households several days before or after their assigned survey visit date.

In each month, we randomly selected sixty-five households with whom to conduct audit visits, or just under 5% of treatment group households per month. We were most concerned about anticipatory behavior among households who frequently showed presence of chlorine residual at regularly scheduled visits, but did not want to oversample so heavily from this group that there would be a treatment effect from extra visits. As such, when selecting households for audit visits, we stratified by the frequency of testing positive for the presence of chlorine residual and ensured that no household was audited more than twice.

Appendix K Diarrhea

In each visit, we asked participants to recount the number of days in the last week that each child had diarrhea ("experienced motions"). Ultimately, we consider child anthropometrics as our key measure of child health because it is objective. However, since we have panel data on diarrhea, we can use our diarrhea data to understand how child health changed over time with the treatments. We analyze the aggregate of child-diarrhea-days across all children in the household, controlling for the number of children. Household total child-days of diarrhea is the same measure that Info-tool households recorded as a part of the Info-tool treatment.

Although we recorded data on child-days of diarrhea during the three months prior to the distribution of chlorine, we do not analyze data from this time period because we consider this a training period for the Info-tool group. Indeed, we see much higher rates of recorded diarrhea in the Info-tool group during this period, which we believe is likely over-reporting in response to the Info-tool treatment. To ensure uniformity across treatment groups in how respondents interpret and report loose stools, we used the Bristol stool chart to help caregivers identify loose stools at baseline. This chart provides illustrations of different potential stool consistencies. At endline, we showed households the chart again and asked them to identify which illustrations they would consider to be motions, to see if the treatments had changed participants' interpretation of what constitutes a loose stool. Although we expected the Info-tool treatment to be the most likely treatment to change participants' interpretation of loose stools (because the treatment would lead participants to more closely attend to children's stools), it was actually the Incentives group whose interpretation of loose stools appears to have changed. The Incentives group was more likely to consider illustrations of solid stools to be loose, and therefore might have overestimated their children's rate of diarrhea. There were no differences in the number of illustrations that Pure Control, Chlorine Only, and Info-tool considered to be loose stools. Consequently, we think that child-days of diarrhea is a good proxy for child health when comparing the Info-tool, Chlorine Only, and Control groups. Incentives households may have over-estimated their children's diarrhea rates, so we interpret comparisons

between the diarrhea rates in the Incentives group and other groups cautiously.

Table C.2 presents the results. Across all treatment groups, the impact of chlorine tablet dispensation is substantial: diarrhea rates drop by approximately 30% ($p < 0.005$) for Chlorine only, Incentives, and Info-tool households in the short run. These effects are largely sustained over the medium run, although Incentives households fall short, consistent with their chlorine residual patterns. Treated households broadly continue to demonstrate large and significant reductions in child diarrhea into the long-run, again with the exception of Incentive households.

Average treatment effects suggest that chlorine provision effectively and substantially reduces child diarrhea rates, these effects persist over the course of a year, and those in the Info-tool arm experience the largest gains, followed by those in Chlorine Only, and lastly the Incentives arm. While these patterns are merely suggestive given the self-reported nature of the outcome, we do observe the same pattern of results in our objective anthropometric measures of child health that we collect at endline, which we turn to next. Our findings suggest that continuous chlorine use, even during the season where diarrhea poses the lowest risk (our medium-run period, during which Incentives households most drastically reduce their chlorine use), is important for building young children's stock of health. This pattern also speaks to the importance of tools to help participants attend to subtle health signals. Diarrhea is rare in this season, so differences in health with the introduction of chlorine will be more difficult to observe than during the season when diarrhea rates are higher.