POLICY RESEARCH WORKING PAPER 7164

Promoting Handwashing and Sanitation

Evidence from a Large-Scale Randomized Trial in Rural Tanzania

Bertha Briceño
Aidan Coville
Sebastian Martinez

WORLD BANK GROUP

Water Global Practice Group
  &
Development Research Group
Impact Evaluation Team
January 2015

Abstract

The association between hygiene, sanitation, and health is well documented, yet thousands of children die each year from exposure to contaminated fecal matter. At the same time, evidence on the effectiveness of at-scale behavior change interventions to improve sanitation and hygiene practices is limited. This paper presents the results of two large-scale, government-led handwashing and sanitation promotion campaigns in rural Tanzania. For the campaign, 181 wards were randomly assigned to receive sanitation promotion, handwashing promotion, both interventions together, or neither. One year after the end of the program, sanitation wards increased latrine construction rates from 38.6 to 51 percent and reduced regular open defecation from 23.1 to 11.1 percent. Households in handwashing wards show marginal improvements in handwashing behavior related to food preparation, but not at other critical junctures. Limited interaction is observed between handwashing and sanitation on intermediate outcomes: wards that received both handwashing and sanitation promotion are less likely to have feces visible around their latrine and more likely to have a handwashing station close to their latrine facility relative to individual treatment groups. Final health effects on child health measured through diarrhea, anemia, stunting, and wasting are absent in the single-intervention groups. The combined-treatment group produces statistically detectable, but biologically insignificant and inconsistent, health impacts. The results highlight the importance of focusing on intermediate outcomes of take-up and behavior change as a critical first step in large-scale programs before realizing the changes in health that sanitation and hygiene interventions aim to deliver.

Promoting Handwashing and Sanitation: Evidence from a Large-Scale Randomized Trial in Rural Tanzania

Bertha Briceño
Aidan Coville
Sebastian Martinez[1]

[1] *Briceño*: Inter American Development Bank, 1300 New York Avenue, NW, Washington DC 20577, berthab@iadb.org; *Coville*: World Bank, 1818 H St NW, Washington DC 20433, acoville@worldbank.org; *Martinez*: Inter American Development Bank, 1300 New York Avenue, NW, Washington DC 20577, smartinez@iadb.org.

Acknowledgements

1. **Introduction**

Understanding how to reduce enteric and diarrheal diseases has important implications for child morbidity, mortality and long-term growth. Diarrhea is the second largest killer of children under five (WHO-UNICEF, 2013), and poor nutrition at an early age can cause growth faltering and reduced cognitive development in the longer term (Victora *et al.*, 2008). Exposure to, and ingestion of, contaminated fecal matter is understood to be the main pathway through which children are exposed to diarrheal diseases and other afflictions that affect nutrient absorption and longer-term growth, such as soil-transmitted helminthes and environmental enteropathy (Humphrey, 2009). Poor children in developing countries are the ones most at risk of exposure.

Given their role in reducing fecal-oral pathogen transmission, hygiene and sanitation programs are often cited as important development interventions to reduce morbidity and mortality worldwide. It is estimated that unsafe water, and inadequate sanitation and hygiene contribute to over one million deaths per year, or 1.5% of all deaths (Prüss-Ustün *et al*, 2014); in Sub-Saharan Africa, these deaths are mostly of children under 5 years old. As a result, governments and the international community are increasingly investing in the provision of water, sanitation, and hygiene (WASH) for the poor. However, the existing evidence on the effectiveness of these investments when delivered at scale is limited and of varied quality.[2] Small-scale efficacy trials show large reductions in diarrhea from handwashing interventions (Ejimot-Nwadiaro *et al.*, 2008), but there are important concerns about potential measurement bias of self-reported measures (Schmidt, 2014). Apart from recent promising work in India (Biran *et al.*, 2014), most of the available research suggests that, when taken to scale, sustained behavior change remains elusive (Chase & Do, 2012; Galiani *et al.*, 2012). Currently, we do not have efficacy trials of sanitation on which to provide a benchmark, and the little evidence on effectiveness trials of rural sanitation interventions is far from conclusive.[3] There are a number of reasons identified in the literature that explain why we may see no, or even negative, impacts on health from sanitation interventions. Improved latrine coverage may do little to reduce exposure to other important contaminants such as animal feces (Ngure *et al.*, 2013). Due to the externalities associated with open defecation, health improvements may not be observed if there is limited take-up in the community (Andres *et al.*, 2013). Increasing latrine coverage may even have unintended negative consequences, by contaminating groundwater through seepage (Dzwairo *et al.*, 2006) or increasing exposure to feces through unhygienic facilities (Greene *et al.,* 2012). The latter points towards potentially important complementarities between sanitation and hygiene campaigns. While the importance of exposure to contaminated fecal matter on the burden of disease is not in dispute, the effectiveness of hygiene and sanitation promotion campaigns to

---

[2] See Clasen *et al.* (2010), Fewtrell *et al.* (2005), Kremer and Zwane (2007), Independent Evaluation Group (2008), Waddington *et al.* (2009), Cairncross *et al.* (2010), and DFID (2011).

[3] See Cameron et. Al (2013), Patil *et al* (2014), Clasen *et al.* (2014) and Hammer & Spears (2013).

adequately reduce this transmission mechanism, particularly when implementing large-scale programs, is still an open and important empirical question.

We evaluate the impact of two large-scale, government-implemented programs that promote handwashing with soap and improved sanitation in rural Tanzania. The programs are part of a multi-country effort to address poor hygiene and sanitation conditions for large rural populations in the developing world, implemented through local governments with assistance from the World Bank.[4] From mid-2009 to early 2011, the two interventions were rolled out in 10 districts[5] of Tanzania following a factorial experimental design. 181 rural wards included in the study were divided into 4 groups to receive the handwashing intervention alone, the sanitation intervention alone, both the handwashing and sanitation, or neither intervention (control). We conducted an endline survey in 2012, approximately one year after the conclusion of the program. The survey includes 3,619 households and 5,768 children under five, with an effective response rate of 97.8%. The trial was registered as NCT01465204 at clinicaltrial.gov.[6]

Handwashing wards were provided with a package of intensive social marketing interventions, including training of community activists, direct consumer contact through road shows, mass media campaigns and promotional activities, and technical assistance to build handwashing stations with local materials. Sanitation wards received a similar package of marketing efforts coupled with a community-led total sanitation triggering event geared towards increasing demand for improved sanitation facilities and promoting open defecation free (ODF) communities, followed by the creation of a village sanitation committee in charge of ensuring sustained behavior change. This was complemented with supply side interventions to train local masons in latrine construction and marketing. In both cases, sanitation marketing messages concentrated on positive aspirational messages rather than shame tactics. No subsidies were used.

Approximately one year after the end of the program, treatment households are significantly more likely to report being exposed to handwashing and sanitation messages of the sort used by the program. Households in sanitation wards are more likely to know of a mason who builds latrines and be aware of a sanitation committee operating in their village. Private latrine construction increased by 12.4 percentage points over the intervention period, a relative increase of 33% compared to the control. Increased latrine coverage leads to a 52% decline (23.1% vs 11.1%) in reported open defecation as the primary means of depositing feces in sanitation wards. Handwashing results are less impressive. Households in handwashing wards show marginal improvements in handwashing with soap prior to food preparation but not at other critical

---

[4] Since 2007, the Water and Sanitation Program (WSP) has provided technical assistance to local and national governments implementing large rural sanitation and handwashing promotion programs in India, Indonesia, Peru, Senegal, Tanzania, and Vietnam, under the umbrella of two related projects, Global Scaling Up Handwashing and Global Scaling Up Rural Sanitation. More information is available at www.wsp.org/scalingupsanitation.

[5] Out of 129 districts in Tanzania at the time of the 2002 census. These districts cover a population of 2,7 million people.

[6] http://clinicaltrials.gov/ct2/show/NCT01465204

junctures such as after defecation. We also find small but significant improvements from handwashing promotion along other indicators of cleanliness including the child's state of cleanliness, caregiver's hands, probability of covering food and cleaning of the household latrine. Intermediate outcomes in the combined handwashing and sanitation intervention are consistent with impacts found in the single intervention wards. We find no clear evidence of meaningful health impacts. On the one hand, using 34,045 child observations from the listing data, we find a marginally significant (12.5%) reduction in diarrhea,[7] particularly for children under three in the combined treatment wards, pointing towards potential complementarities. The household survey data of 5,768 children shows a statistically insignificant reduction. Counterintuitively, we also find small but significant reductions in hemoglobin levels measuring iron-deficiency anemia (1.5%) and weight-for-age in the combination ward, which could be driven by gut infections and soil-transmitted helminthes (STHs) reducing the body's ability to absorb nutrients. These anemia and weight results are statistically significant but biologically small. We show that this counterintuitive result is unlikely to be driven by selective mortality or poor quality latrine construction and the results remain robust to a number of different specifications.

The study design presents some limitations which we mention upfront. First, the 10 districts where we conduct the study were selected by government based on political priority rather than random selection, affecting the external validity of the study. This said, comparing our control group to representative surveys of rural Tanzania, including the 2010 Demographic and Health Survey and National Panel Survey, reveals that the two areas are more similar than not along basic observable characteristics (refer to Table 1). Since these are predominantly rural households, we also find that our sample is poorer on average than the national population. Following Filmer and Pritchett (2001), we use principal-components analysis to derive a wealth index based on 16 household assets captured both in the study questionnaire and the nationally representative 2010-2011 Living Standards Measurement Survey (LSMS). By plotting the wealth distributions of the impact evaluation and LSMS samples, Figure 1 shows that approximately 57% of the impact evaluation group falls below the 40th percentile of the national wealth distribution – the target group for the World Bank's goal of increasing shared prosperity.

The second important concern is that we lack a baseline of pre-intervention characteristics.[8] While we attempt to reconstruct baselines for key variables, these suffer from problems of recall. Differential migration and attrition could also be a concern. However, data from the complete census listings of selected enumerator areas provides evidence of limited migration and attrition. Less than 5% of households moved into the community within the three year intervention period,

---

[7] The household-level indicator for diarrhea including 5768 households does not find any significant effect on diarrhea, although the direction of the sign for the coefficient is consistent.

[8] Although a baseline data collection was intended, unanticipated problems with reliability of data resulted in the cancelation of field work in five out of the 10 districts originally planned and the impossibility of using the data to validate the randomized design, as it was originally planned.

and this does not differ across treatment groups. We mitigate any confounding that migration may cause by restricting the sample to households residing in the area since 2009. Finally, though this study represents one of the first attempts to experimentally test the impact of handwashing and sanitation interventions separately, the study design does not isolate the impact of multiple individual components that make up each of the handwashing and sanitation intervention packages.

Our study complements the literature by providing the first experimental evidence of the separate and combined effects of sanitation and hygiene campaigns at scale. The work adds to the growing body of evidence suggesting that, while handwashing with soap may be an effective way to reduce an important fecal-oral transmission mechanism, getting people to actually change their behavior is incredibly difficult (Chase & Do, 2012). The work contributes to a more limited literature on the causal effects of rural sanitation. Companion studies in Indonesia and India have shown that, while it is possible to increase latrine construction through similar promotion campaigns,[9] these changes fall short of ensuring universal latrine coverage, limiting the potential for positive health outcomes (Cameron *et al.,* 2013; Patil *et al.*, 2014).[10] More recent work by Clasen *et al.* (2014) finds a similar conclusion even with relatively large improvements in latrine coverage (52 percentage point increase). The results highlight the importance of first addressing take up and behavior change as a central goal of at-scale WASH interventions before embarking on the quest for health impacts (Coville & Orozco, 2014).

The paper is presented in the following way: Section 2 describes the causal chain linking sanitation and handwashing to improved health (presented visually in Figure 2), contextualizes the problem in Tanzania, and describes the program interventions and implementation in detail. Section 3 describes the study design, Section 4 provides an overview of the data and Section 5 outlines the econometric specification used to derive the results. Section 6 presents results, followed by robustness checks in Section 7. We discuss some possible mechanisms to explain the counterintuitive health outcomes in Section 8 and conclude in Section 9.

2. Handwashing and Sanitation Promotion
2.1. *Transmission Mechanism*

Better sanitation, through safe containment of human feces and reductions in open defecation, is expected to decrease the presence of fecal pathogens in the environment, which can otherwise be transmitted via soil, surface water, hands or flies. Similarly, handwashing at critical times can decrease the risk of infection by reducing the presence of enteric pathogens in the hands/fingers

---

[9] An important difference in India's Total Sanitation Campaign is that latrine promotion is accompanied by a subsidy which is not the case in Indonesia or Tanzania.

[10] Hammer & Spears (2013) find a large improvement of India's Total Sanitation Campaign on child stunting, and while the analysis is rigorous, the results are somewhat implausible from a biological viewpoint given the small changes in latrine construction, which suggests possible data quality concerns.

and objects or utensils of frequent use, thus reducing the chances of person-to-person transmission and food or water contamination. The variety of transmission channels has been summarized in the seminal "F-diagram" reproduced in Figure 3.

High levels of enteric pathogens manifest in three main ways that adversely affect nutritional outcomes, especially for young children: diarrheal diseases, gastrointestinal infections and intestinal parasitic infections (soil helminthes and protozoans).[11] The pathogens that cause diarrheal diseases are transmitted mainly via fecal-oral routes, by ingestion of contaminated food, water, or other beverages, by direct or indirect contact with contaminated hands, animal or human feces.[12] Gastrointestinal infections and environmental enteropathy, are caused by ingestion of fecal bacteria in large quantities. Intestinal parasitic infections such as ascariasis, trichuriasis and hookworm are transmitted through soil contaminated with feces; giardia is a waterborne infection but can also be transmitted through food and person-to-person contact.

Despite being less visible than diarrheal diseases, subclinical gastrointestinal infections and intestinal parasitic infections also affect nutrition in substantial ways. The absorption of nutrients is reduced by direct loss through diarrhea, especially by repeated episodes, but also by continuous and silent theft of nutrients by parasites (including blood by hookworm), and by diversion of energy and proteins to combat infections.[13] Environmental enteropathy reduces absorptive capacity while increasing permeability of the small intestine (Humphrey, 2009). The result is faltering weight and height for age and increased risk of anemia.[14]

It is commonly believed, on the basis of practitioners' claims and some observational studies, that water, sanitation and hygiene investments may result in additional benefits when provided together. Arnold *et al.* (2013) note that, even though some observational studies and theoretical modeling of transmission pathways support the use of combined approaches promoted by implementing agencies, rigorous evidence justifying this approach is scarce. A meta-analysis by Fewtrell *et al*. (2005) finds that the only evidence on combined interventions do not reduce diarrhea more than individual interventions. Our study, being the first large randomized trial exploring gains from sanitation promotion and handwashing implemented separately and in combination, provides important evidence for understanding the potential complementarities of WASH provision for at-scale interventions.[15]

---

[11] Other important tropical diseases that can be prevented through adequate water, sanitation and hygiene include trachoma and schistosomiasis. [Mara, 2010]

[12] See Clasen *et al*. (2010), Prüss *et al*. (2008)

[13] Chambers ,R (2012) Handout on fecally-transmitted infectious and undernutrition.

[14] Helminth infections are also frequent cause of anemia in pregnant women [Mara, 2010]

[15] In the near future, it is expected that other ongoing trials like the WASH benefits study will contribute to a more comprehensive understanding of the interactions among water quality, sanitation, handwashing, and nutrition interventions alone and combined designed as an efficacy trial for proof of concept. See Arnold *et. al* (2013) for the complete study protocol.

This study measures the effects of the sanitation and handwashing interventions on the health of children under five years old. These children represent the age group most susceptible to diarrheal disease, growth faltering and acute lower respiratory infections, which are major causes of childhood morbidity and mortality in developing countries.

2.2. Context: Rural Tanzania

Approximately 70% of Tanzania's population lives in rural areas characterized by subsistence livelihoods, poor access to health and education services, and high morbidity and mortality rates. The 2010 Demographic and Health Survey estimated that 9.2% of children in rural Tanzania died before their fifth birthday, 13.6% experienced diarrhea within a 2 week period, and 57.8% could be classified as anemic. Stunting and wasting, as measured by height-for-age and weight-for-age z-scores, is also commonplace, with 44.5% of children under 5 falling below two standard deviations of the reference population mean in height-for-age measures and 4.9% for weight-for-age measures. Less than half of the rural population has access to an improved water source, and more than half spends over 30 minutes collecting water. The country experienced a rapid acceleration of sanitation coverage following a cholera outbreak in the late 1970s, with widespread adoption of basic latrines in rural areas. Coverage in rural areas was estimated at 50 percent by 1980 and the 93/94 HRDS[16] survey found that 92.3% of rural Tanzanian households had a traditional or improved pit latrine. Over time, these (mostly shared) facilities gradually fell into disrepair and, currently, do not meet the JMP standards for "improved" sanitation (World Bank, 1996). As such, while most Tanzanian households have access to some form of latrine, the majority of these latrines can be more accurately described as 'fixed-point open defecation'. While approximately 80% of households have access to a pit latrine, only 8% have a slab to ensure that fecal matter is safely separated from human contact. Open defecation continues to be the most common form of toileting among 18% of households and child feces are disposed of in an unsafe manner 40% of the time.

The 92 villages in 46 wards included as our study control sample provide a nuanced snapshot of sanitation and hygiene conditions in our study area absent the intervention.[17] While 80% of households have a pit latrine with slab or VIP latrine, more than 40% of these slabs are in disrepair and a third of the facilities are shared with other households. Furthermore, only 17% of these latrines have a cover for the squat hole to contain the feces and reduce potential transmission by flies, and more than two-thirds of the households with squat hole covers did not have them in place when visited. Our data show that, although people are aware of the

---

[16] Tanzania Human Resources Development Survey, 1993/94

[17] Note that control areas were exposed to mass media campaigns on the radio.

importance of washing hands with soap, the presence of soap and handwashing devices,[18] together with actual handwashing behavior, remains low. We observe that 82% of caregivers know that handwashing with soap and water is the best method for cleaning hands, while only 8% of households have any soap visible near the place where hands are washed and only 1.2% have a fixed handwashing device. In-depth structured observations reveal that any form of handwashing at critical junctures, such as post-toileting or before eating, is low (27%), and washing hands with soap is substantially lower (4%). Households have toileting facilities, but they are in disrepair and often lack the ability to effectively store fecal matter safely and reduce the risk of contamination. Caregivers have basic knowledge of appropriate handwashing behavior, but fail to translate this into practice.

2.3. Rural Sanitation: Total Sanitation and Sanitation Marketing

The rural sanitation intervention – Total Sanitation and Sanitation Marketing (TSSM) – combines demand- and supply-side strengthening in an attempt to shift the sanitation equilibrium in targeted villages. TSSM uses Community Led Total Sanitation (CLTS) and sanitation marketing to increase demand for improved sanitation, while strengthening the supply of sanitation goods and services to local markets, with the aim of making these products more affordable and accessible. The objective of the program is to move all household members up the sanitation ladder, away from open defecation and towards the use of improved latrines, thereby creating ODF communities.

CLTS is an approach originally championed in South Asia to reduce open defecation by using community mobilization activities and behavior change strategies. CLTS recognizes that poor sanitation practices pose a fundamental collective action problem and uses a catalytic "triggering" event to highlight the problems created by open defecation. A CLTS leader triggers a community by bringing community members together and producing a defecation map, highlighting all of the areas where defecation occurs. Community members then visit these locations on a transect walk. Finally, the facilitator conducts a feces calculation to highlight the quantity of fecal matter produced in the community, and generates a mobility chart to show households how the fecal matter inevitably contaminates drinking water. In Tanzania, the CLTS model deviated from its usual focus on a shame-triggered message that urges behavior change, to a more positive one focusing on aspirations and pride (Perez *et al.*, 2012). The project trained local CLTS trainers at the district level who, in turn, trained CLTS facilitators in each ward. Facilitators then performed CLTS triggering in the communities, after which communities would create a CLTS committee consisting of 5 members who would be responsible for starting a

---

[18] Most handwashing devices in our sample are simple mobile buckets with water and soap. We differentiate between mobile devices such as these and more permanent "fixed" devices which include tippy taps (specifically promoted by the intervention), a sink with tap or fixed basin.

village latrine register and monitoring progress, while continually motivating households to move up the sanitation ladder.

CLTS activities were accompanied by intensive sanitation marketing and promotion campaigns directed towards the head of the household. Marketing approaches included direct consumer contact (DCC) which are large, community-based events, and mass media (print and radio). Promotional activities at the village level were delivered by village leaders, community health workers, as well as the CLTS champions. To maintain focus on aspirations, activities of the sanitation marketing campaign fell under the common slogan *Choo bora chawesekana* ("a good latrine is possible").

While the mass media radio campaign was broadcasted across the country, 182 villages in 91 wards were targeted to receive direct sanitation marketing efforts through promotional activities. TSSM demand-side strengthening activities consisted of the following:

1. A mass media campaign including: (i) radio advertisements that feature a short jingle about the importance of sanitation, and (ii) a soap opera during a primetime slot on two radio channels with 15 minute episodes covering promotional messages for both latrine upgrading and handwashing behavior.
2. Wall murals, posters and flyers with the *Choo bora chawesekana* slogan providing information on pricing, simplicity and the benefits of the "*sungura*" [19] slab constructed by masons trained in the program (see Figure 4 for an illustration).
3. A roadshow (DCC event) promoting improved sanitation which would include skits, competitions, music, dancing, and *sungura* slab sales promotions.
4. CLTS triggering and the creation of a sanitation committee.

Beyond lack of demand, formative research developed to inform the intervention design identified limited availability of low-cost, durable sanitation options as being prominent challenges to the success of the intervention in rural communities. Furthermore, masons exhibited limited business and marketing skills. To overcome these constraints, demand-side interventions were accompanied by supply-side strengthening to ensure local provision of latrines and sanitation products. The program trained masons, or *fundis,* in latrine construction and marketing skills. The training lasted a week and taught masons how to build *sungura* cement sanitation platforms to be sold for approximately $5 per slab. Masons were also taught more general techniques for upgrading latrines together with basic marketing and management skills to guide their operations. The aim of the training was to enable masons to capitalize on the demand for sanitation products generated by the promotion activities, and to spur future demand.

---

[19] Sungura means rabbit in Swahili and refers to the appearance of the slab

A year into the program, limited availability of materials was creating bottlenecks in the process, which led to the decision to introduce a supply chain strengthening component aimed at incentivizing local distributors (hardware stores) to make the materials required for cement slabs more readily available to masons to use as inputs.

2.4. Handwashing with Soap

The handwashing with soap (HWWS) intervention targeted rural mothers with children under 5. The program conducted extensive formative research on barriers to handwashing with soap, and used the results to inform the design of communication campaigns. Barriers identified include a lack of both time and cues (e.g. handwashing stations) that are helpful in reinforcing habits. In light of these barriers, HWWS promoted an enabling technology called the "tippy tap[20]" that provides an external cue to handwashing with soap. It comprises a simple handwashing station equipped with water and soap (made cheaply using local materials). The campaign specifically targeted mothers, who are often seen as the guardians of children's health and wellbeing, and recognized them for their contribution to the family. The campaign was built around an overarching communication concept embodied by the slogan *Mikono Yenye Fahari* ("Hands to be proud of"), which was designed to tap into mothers' aspiration for recognition for the work they do for families (Coombes & Paynter, 2010).

The campaign, which later evolved into *Asante Mama* ("Thank you mother"), formed the backbone of the district-level activities. Activities of the program included mass media interventions (radio adverts and soap opera); branded intervention material (posters, hats, clothing, etc.); DCC roadshow events (similar to the TSSM interventions); and interpersonal contact (IPC) through front-line activators (FLAs), or *Msabunashajis*,[21] who are trained to visit households and conduct handwashing promotion events with women on market days, during pre-natal clinic visits, and at village meetings FLAs would also distribute promotional material including flyers and comics on how to construct a tippy tap, and a wall calendar used to remind households of the times to watch the soap opera.

The program identified volunteers from each community and provided them with a 3-day-long training structured around pathways for contamination and advocating appropriate hygiene practices related to handwashing with soap (including the best methods to follow and critical junctures). FLAs were subsequently taught how to effectively engage with the community, relay campaign messages to caretakers, and successfully monitor hygienic practices in the village. They were trained in tippy tap construction, and were expected to impart this learning to village households, with caregivers of young children as their primary target.

---

[20] For more information on tippy-taps and other enabling technologies, see www2.wsp.org/scalinguphandwashing/enablingtechnologies.

[21] Roughly translated as "The Soaper"

The HWWS DCC events targeted mothers and stressed the importance of ensuring the cleanliness and health of children, which required washing hands with water and soap at critical junctures. People were shown how to construct a tippy tap and given a chance to try using one. The FLAs reinforced this message in their interaction with caretakers. The HWWS DCC event was otherwise implemented in the same manner as the TSSM DCC activities.

2.5. Program Implementation

WSP worked in collaboration with Tanzania's Ministries of Health and Social Welfare to develop the TSSM and HWWS campaign materials and trained district government counterparts who in turn conducted training of local FLAs, CLTS facilitators and masons and coordinated the distribution of marketing materials. DCC events were implemented by private companies (separately for the HWWS and TSSM campaigns). The objective was for the Tanzanian government to have overall oversight and control of the program implementation for sustainability purposes.

TSSM and HWWS promotional campaigns were phased in between February 2009 and December 2010 across 10 rural districts: Mpwapwa, Kondoa, Rufiji, Iringa, Sumbawanga, Kiteto, Masasi, Musoma, Karagwe and Igunga. The two interventions were implemented separately – in this case the combined intervention implied the overlapping delivery of both interventions, rather than a coordinated hybrid of the two. A timeline of activities is presented in Figure 5. The mass media campaign ran from February 2009 – December 2010 across the country and became a joint TSSM/HWWS intervention; 15 minute radio slots discussed issues related to both sanitation and hygiene. Distinct HWWS and TSSM DCC roadshows and other marketing materials were initially rolled out from June 2009 – June 2010. However, two separate DCC campaigns were conducted for the HWWS intervention, in contrast to only one for the TSSM intervention. The first campaign ran from January – June 2010 and the second ran from August – October 2010. Mason and FLA training took place early on in the campaigns (June 2009 – March 2010). Masons were then ready to meet the newly generated demand for sanitation following the implementation of CLTS triggering and DCC events in the latter half of 2009 and early 2010. Similarly, FLAs were in place to reinforce HWWS messages.

The mass media radio campaign was broadcasted across the country on two of the major radio channels. A total of 80 15 minute soap opera episodes and 1800 radio adverts of 45 seconds each were aired, reaching an estimated 10 million people. Since the radio campaign reached all wards, irrespective of treatment assignment, it was not directly evaluated and all results in this paper can be interpreted as the effect of the TSSM/HWWS interventions at the ward level in the presence of a radio campaign.[22]

---

[22] In a nested experiment we attempted to unpack the effect of the mass media campaign by conducting a randomized encouragement design which provided one third of HWWS households with a calendar or comic mentioning the date and time

Table 2 summarizes the reach of the various intervention components based on project monitoring data. For TSSM, 407 masons were trained across all treatment villages. A further 282 CLTS facilitators were trained and conducted CLTS triggering in a total of 79 wards.[23] All TSSM villages received murals and/or posters, and 81 out of the 90 originally assigned wards received the DCC event[24] reaching a recorded 180,000 people.

For the HWWS campaign, 433 FLAs were trained across all of the treatment villages. The DCC event missed 5 of the 91 assigned wards, reaching an estimated 220,000 people during the first round and 234,000 in the second round.[25]

3. Experimental Design

3.1. Identification Strategy

Tanzania is administratively separated into 30 regions, 169 districts and 3,643 wards, with the average ward holding approximately 12,000 people. To evaluate the impact of sanitation, handwashing and combined interventions, we implement a cluster-randomized evaluation with random assignment of interventions at the ward level, including a total of 135 intervention and 46 control wards. Wards were identified as the optimal operational unit of implementation for the project, and of sufficient geographic extension to minimize the risk of significant information spillovers between populations exposed to the localized messages, community events, and other forms of social promotion activities.

The sample was drawn from 10 districts spread throughout the country selected by the Ministry of Water (MoW) and Ministry of Health and Social Welfare (MoHSW) to provide geographic diversity at the national level (see map in Figure 6). These districts are not random, and were targeted because of operational feasibility for program implementation, taking into account the existence of ongoing MoW and MoHSW projects, including the Health Village Campaign (HVC) and water and sanitation interventions. Of the 245 wards in these 10 districts, initially 13 wards were dropped from the sample. Three of these wards were urban, and thus ineligible for the program, and the remaining 10 wards were pre-selected as pilot areas for the program and excluded from the evaluation. Among the remaining 232 wards the program selected the 190 largest wards by population size in order to maximize the population under treatment. These 190 wards were subsequently randomly assigned to one of four groups: (1) Handwashing intervention, (2) Sanitation intervention, (3) Handwashing and Sanitation intervention, and (4)

---

of the soap opera. We find this had a positive but non-significant increase in recall of the soap opera which was not a large enough effect to be able to measure subsequent impacts of the campaign through the encouragement.

[23] While 90 wards were assigned to receive CLTS, 11 were not treated due to logistical challenges of district facilitators training and visiting wards, with lack of funds for the district facilitator to travel cited as the reason for non-compliance.

[24] Imperfect compliance resulted from logistical constraints of reaching villages during the roadshow schedule

[25] It is not possible to compute how much overlap there was between the first and second round, but it is expected that these numbers mostly reflect repeat exposure since the events were conducted in the same villages in both rounds.

Control (no intervention). After the initial sample selection, the district of Massasi experienced a re-districting process through which 9 wards were reassigned to a neighboring district which was not part of the program and were dropped from the sample. The reassignment was balanced across treatment arms and included 3 wards from the sanitation only group and 2 wards from each of the other study arms resulting in a final experimental sample of 181 wards.

We implemented a block randomization procedure within districts with the objective of balancing the population sizes of treatment and control wards within each district. First, wards were ordered by population size within districts and blocked into groups of four, starting with the four most populous wards, second most populous, and so on. A random number was then assigned to each ward, and a group number between 1 and 4 was given to each ward based on the ordering of the random number within its respective block. Finally, group numbers were randomly assigned to represent one of the three treatment or control groups.[26]

3.2. Compliance

Table 3 presents treatment compliance at the ward level. Of the 181 wards selected for the sample, 45 were assigned to handwashing, 44 to sanitation, 46 to the combined intervention and the remainder to control. According to administrative records, the implementing agency accidentally conducted handwashing promotion in one of the sanitation wards, resulting in actual delivery of TSSM only to 43 wards and combined TSSM and HWWS to 47 wards. There were no reported deviations from the planned implementation of mason or FLA training, and no information was available to assess the actual delivery of village-level media (wall drawings, posters, etc.), although, based on information from the field managers, we expect that a majority of the print material dissemination was implemented as scheduled. Within each ward, the two largest villages were targeted for program implementation.

Administrative records of program implementation determined that some wards received only partial treatment. 2 of the 45 HWWS wards were not exposed to the DCC event, 7 of the 43 TSSM wards did not have a CLTS triggering, and 5 of these 7 wards did not receive the DCC roadshow event. Of the 47 wards that were assigned to both TSSM and HWWS, 4 wards did not receive a CLTS triggering or DCC exposure.

3.3. Balance

Table 4 presents balance statistics. Since we do not have a baseline survey, this consists of a combination of time-invariant indicators and retrospective responses asked in the endline dating to February 2009 – before the intervention had started. Of the 87 variables we are able to present for each of the three groups (resulting in 261 comparisons to the control), we find a statistically

---

26 The randomized selection procedure can be obtained by the authors on request

significant difference in 12 tests for balance at the 5% significance level. The expected number of "by-chance" imbalances from a random draw of 261 is 13, which suggests that this is well within the expected range. However, some concerns persist. Firstly, half (6) of the imbalances are found in the HWWS group, which is more likely to have a cement floor and piped water connection. The sanitation only households are more likely to have a pit latrine with slab or ventilated pit latrine (VIP) in 2009, and the combination ward households are more likely to listen to the radio and have slightly older household members than the control group. The most concerning imbalance for the study is the difference in sanitation coverage between TSSM and the control group. However, there are good reasons to be skeptical about the validity of these imbalances, given that the collected data is retrospective in nature and subject to recall bias. Nonetheless, we run all of the results and include all of the imbalanced variables as controls, with the tables presented in the online appendix. Overall, the TSSM impact on latrine construction and availability reduces, to line up more closely with results found in the combined intervention, but signs and significance remain consistent, suggesting that these imbalances are not driving any conclusions made in the paper.

4. Data and Sample

The sampling procedure, in following program operational guidelines which targeted the two largest villages in each treatment ward, selected the two largest villages, based on population size, in each of the 181 evaluation wards. For the 362 villages in the sample, the full list of census enumeration areas (EAs) was obtained from the Tanzanian National Bureau of Statistics (NBS). For each village, the sample included one EA, selected with probability proportional to size (PPS). A census listing exercise was then conducted in each EA to collect basic information to determine household eligibility for the survey. Survey eligibility criteria were (i) the household was present during the period of listing; (ii) had been living in the village since the beginning of 2009 or earlier; and (iii) had at least one child under the age of five. Ten eligible households were then selected from each EA at random for the sample.

Field work was conducted between May and December 2012.[27] Field teams visited each village for three days. The first day was used to conduct the EA census listing exercise.[28] The census collected data on 72,705 households in the 362 selected EAs. 31% of listed households were eligible for the survey, primarily based on the presence of a child under age five in the household. Field supervisors then ran an in-field, automated randomization procedure to select 10 households at random from each EA to participate in the household survey together with a replacement sample of 5 households to be used in the case of potential non-response. If a household refused to participate or was not available after two visits, the first household in the

---

[27] In 2009 a baseline survey was initiated but not completed in 5 of the 10 intervention districts; The endline survey was implemented under an independent sampling scheme, following the original sample design.

[28] EA sizes ranged from 51 to 746 households with an average of 221.

replacement sample was selected. A total of 105 such replacements were made resulting in 3,619 completed interviews from 3,724 attempted (97.2% response rate).

In addition, two of the ten sampled households were selected randomly for structured observations which took place on the second day (resulting in a sub-sample of 724 households). Structured observations consisted of a three hour visit to the household in the early morning (determined by the time the primary caregiver would wake up) where the enumerator would observe caregiver handwashing behavior. This was conducted before any household interviews were held in the village, with the intention of reducing any potential Hawthorne effects that could result from respondents being aware that the intention of the visit was to record handwashing behavior. Enumerators would record the various handwashing "critical junctures" experienced by the caregiver and target child, which include (i) before preparing, serving or eating food and (ii) post-toileting. They would then record whether critical junctures were accompanied by handwashing with soap and water.

Following the structured observations, household surveys were conducted on the second and third days of field work. The questionnaire[29] included modules on demographics, productive activities and assets, water and sanitation status (observed and self-reported), education, hygiene practices and knowledge, social capital, self-reported exposure to the program interventions and child health. Health results were collected for children between the ages of 6 months and 5 years. A description of the priority outcomes is as follows:

Access to an improved latrine: The variable construction follows the JMP improved sanitation definition based on observed latrine type and private ownership. Based on a specific interest of the Tanzanian Government, we also consider how access to an improved latrine varies when we include shared facilities, or expand the definition of sanitation quality as being fulfilled by only one squat hole with a slab that does not expose contents.[30] For households with latrines, we also ask when the latrine was constructed to determine whether this was done during the intervention period. While a small number of households may have flush/pour facilities, including septic tanks, we refer to "latrine" construction since 97.5% of households with a toilet facility are referring to some form of pit latrine.

Open defecation: This is a measure of households that report using "no facilities" as the usual defecation practice (as per JMP guidelines.) We present a second definition that asks households directly whether they practice open defecation always, sometimes, rarely or never. We then present these results as a measure of intensity.

---

[29] Based on the global WSP survey instrument applied for HHWS impact evaluations in Senegal, Peru and Vietnam and TSSSM evaluations in India and Indonesia.

[30] see http://www.wssinfo.org/definitions-methods/watsan-categories for the formal JMP definition

Caregiver handwashing practices: These were collected based on self-reported information and structured observations. Given the potential courtesy bias associated with self-reported measures, we consider self-reports as reflecting knowledge rather than practice. For structured observations we measure whether the caregiver uses water and soap to wash her own or her child's hands in conjunction with the following exposure events:

- **After fecal contact:** (i) after defecating; (ii) after toileting; (iii) after cleaning child post-toileting
- **Before handling food:** (i) before cutting or preparing food; (ii) before eating; (iii) before serving food; (iv) before breastfeeding.

Diarrhea: From the household data we use a caregiver-reported symptom-based measure defining 7 day diarrhea prevalence as having 3 or more loose/watery stools in a 24 hour period or having a stool with blood or mucus (Baqui *et al.*, 1991). To ensure consistency with the 2010 DHS, in the listing survey we ask caregivers of 34,045 children (under 5 years) a question on whether children have had diarrhea in the past 14 days.

Anemia: Anemia status (iron deficiency) was collected using a HemoCue™ Hb 201+ photometer, which measures blood hemoglobin levels in real time. These biomarkers were collected for all children between 6 months and 5 years old in the household. In line with WHO guidelines, we consider children to be anemic if their hemoglobin concentration is below 110 g/L, (WHO, 2011).

Anthropometry: Child height, weight and head circumference were collected by specially trained enumerators with a nursing background, who followed WHO anthropometric data collection protocols to assess malnutrition levels. Results are transformed into height-for-age, weight-for-age and head circumference-for-age standardized z-scores based on WHO international growth standards.

5. Econometric Specification

We estimate the intention-to-treat (ITT) estimator as the difference between average outcomes across treatment and control groups. The basic specification is:

$$Y_{ij} = \alpha + \sum_{k=1}^{3} \beta_k T_{jk} + \sum_{l=1}^{48} S_{jl} + \epsilon_{ij} \qquad (1)$$

Where, Y_{ij} is the outcome of interest for household or individual i in ward j, T_{jk} is a dummy variable equal to 1 in wards assigned to receive treatment k where $k = \{1,2,3\}$ for HWWS, TSSM and the combined intervention respectively. β_k is the estimate of the average effect of treatment k. S_{jl} is a dummy variable equal to one if Ward j is included in block l, representing the

block fixed effects for the 48[31] stratified ward blocks included to improve precision (Bruhn & McKenzie, 2008). For child-level outcomes we include age and month dummies as covariates. Standard errors are clustered at the ward level. We then check whether the impacts of the three treatment arms are statistically different from each other by presenting F-statistics to test whether we can reject the null hypotheses: $\beta_1 = \beta_2$; $\beta_1 = \beta_3$ and $\beta_2 = \beta_3$. The p-values associated with these tests are reported in all of the regression tables.

In addition to the basic specification, we estimate the following models: (i) include control variables to reduce residual variance and account for any baseline imbalance[32] and (ii) estimate the local average treatment effect for receiving the program, by instrumenting random assignment on actual implementation based on program monitoring data. These additional specifications are presented in the online appendix.

6. Results

For all household- and caregiver-level responses, the tables present the specification that includes block fixed effects only. For the child health indicators, the specification presented includes block fixed effects together with child age (month dummies) and gender controls. In addition we include the full set of models in the online appendix.

6.1. Program Exposure and Outputs

Table 5 shows impacts on self-reported program exposure to TSSM and HWWS channels, with large and significant differences in the probability of being exposed to one or more messages among the three treatment groups. Three channels are included for HWWS and TSSM. Both interventions include promotional materials (which can include wall drawings, posters and comics) and participation in respective DCC events. The third channel of exposure to HWWS includes being visited by an *msabunashaji*/front-line activator, while for TSSM it includes attending a CLTS event. 22.4% of households in the control group report being exposed to any TSSM messages, while this proportion increases by 14.1 percentage points in the TSSM-only group, by 25.1 percentage points in the handwashing-only group, and by 34.5 percentage points in the combined group. When measuring the number of events recalled, we find a nearly linear relationship with the number of DCC triggering activities in each group (1 in the TSSM group, 2 for the handwashing, and 3 in the combined group), suggesting that the frequency of program activities is strongly related to message exposure and recall.

---

[31] Not all districts have ward numbers divisible by 4, in which case some blocks contain only 2 or 3 Wards.

[32] We use the variables found to be unbalanced across intervention groups. This includes: (i) whether the household had a pit latrine or VIP at baseline; (ii) whether household members listen to the radio; (iii) household asset ownership; and (iv) whether the household has access to piped water.

While virtually no one in the control group reports exposure to all three channels, 10 percent of households in the combined treatment group do. Impacts on exposure to HWWS messages show a similar pattern to TSSM. We find increases of 37 and 44 percentage points respectively, off of a control mean of 13.5%, for households reporting exposure to one or more channels in the HWWS and combined wards, while around 6% of households report exposure to three channels.

We also ask households if they are aware of the presence of a CLTS committee in the village. Table 6 reports an approximately 13 percentage point increase in awareness among households in TSSM and combined wards, and a positive and significant 6 percentage point increase in handwashing only wards. Compared to control areas, where only 12 percent of households report knowing of a CLTS committee, awareness in sanitation areas is almost doubled as a consequence of the program. Similarly, when asked if they are aware of a mason in the community, households report an 18 percentage point increase above the 14% awareness level in control villages, implying a relative increase of 128%. Reported awareness of a mason in handwashing only wards also increases significantly, but by 9 percentage points. Lastly, when asked whether the household thinks everyone in the village knows someone who could help build a latrine, there is a significant increase of 7 to 8 percentage points in TSSM treatment areas but no impact in HWWS-only wards.

The presence of reported exposure to TSSM messaging in HWWS-only wards and vice versa can be expected given imperfect recall a year and a half following the implementation of the interventions. This is exacerbated further by the fact that the second HWWS DCC event was conducted more recently than the TSSM activities. Furthermore some messaging content in both the TSSM and HWWS shared similar features, for example with regards to hygiene behavior. That said, as we will see in the results that follow, impacts on final outcomes such as latrine construction are much more consistent with the intervention models, suggesting that this reporting may reflect imperfect recall rather than real contamination.

6.2. Impacts on Latrine Construction and Open Defecation

The primary objective of the sanitation intervention was to increase the coverage of improved latrines and reduce open defecation. This could happen through upgrading of existing latrines, or building new ones. Since it was believed that basic pit latrine coverage was high in Tanzania, the TSSM campaign focused on simple ways to upgrade latrines, for instance, by incorporating the concrete *sungura* slab. However, latrine construction in Tanzania is a fairly frequent activity, with 57% of control households reporting the construction of a new latrine in the last 3 years (the period since the start of the intervention) and while we find limited evidence of households trying to upgrade existing latrines, we find strong evidence of households deciding to build new, private facilities. As a baseline falsification test we estimate the impact of the program on the probability of latrine construction in the baseline period, prior to the start of the intervention, and find no association between the program and pre-intervention latrine construction (Table 7).

The TSSM intervention produces an 8.2 percentage point increase in the probability of building a new latrine in the TSSM only wards, and a 7.7 percentage point increase in the combined wards; these two effects are not significantly different from one another. The effect in handwashing only wards is not significantly different from the control group. New latrine construction is primarily from private rather than shared latrines. The probability of constructing a new private latrine increases between 10 and 12 percentage points above 38% of control households that build new private facilities over the intervention period. Consistent with this result, we find that the probability of sharing a latrine with another household falls by 9.2 and 7.6 percentage points in the TSSM and combined wards respectively.

As expected, we observe no impact of the HWWS intervention on the probability of constructing new private latrines. Unexpectedly, however, there is a significant reduction of about 4 percentage points in the construction of shared latrines in HWWS and TSSW only wards, and an insignificant reduction in combined wards. This may result from household decisions to shift from the construction of shared latrines, to the construction of private ones.

The increase in latrine construction in TSSM wards translates not only into more private latrines but also better quality ones. Table 8 shows that the probability of using improved sanitation increases in both TSSM and combined treatment wards, irrespective of the definition for improved sanitation employed. The use of *sungura* slabs is still low, but increases significantly in the intervention wards from 1.4% in the control areas to 7% and 4.6% in the TSSM and combined wards.

Through behavior change and increased presence of latrines, one of the program's primary sanitation objectives is the reduction of open defecation. Consistent with this objective, we observe large and significant reductions in open defecation in both the TSSM and combination groups. While 23.1% of households report open defecation as the primary form of feces disposal in the control group, this is reduced by 12 percentage points in TSSM only wards and by 7.4 percentage points in combination wards. As presented in Table 9, the majority of the reduction in open defecation is from regular to less frequent open defecation, rather than complete cessation. 51% of households report at least some open defection in the comparison group and the likelihood is not statistically different in the treatment arms. Given the important externalities associated with sanitation and defecation practices, it is useful to understand how this translates into community-wide practices. Community leaders were asked whether the village had been declared open defecation free (ODF). Consistent with the findings on latrine construction and household open defecation, we find a significant increase of 12.7 and 8.7 percentage points in TSSM and combination villages respectively from 5.4% of control villages claiming to be ODF. There is a positive but insignificant increase of 4.6 percentage points in the proportion of ODF HWWS villages. Using the fact that we have a sample of 10 households per village, we can also estimate village-level open defecation prevalence by aggregating household responses. Figure 7 presents the cumulative distribution functions for village-level regular open defecation in each

intervention group. Figure 8 shows the same distribution for households practicing at least some open defecation. Consistent with the previous findings we observe large improvements in village-level regular open defecation across the distribution for TSSM and combination wards, but very little (insignificant) differences for village households practicing open defecation some of the time. The distributions of these two indicators are also strikingly different. For instance, almost all villages have at least some households practicing some open defecation, which increases roughly linearly. In contrast, more than half of the sampled villages have nobody practicing *regular* open defecation. A further 25% of villages have between 0% and 20% regular open defecation prevalence. The results highlight the fact that the intervention is successful in shifting behavior, but this does not result in complete cessation of open defecation which remains pervasive.

We also ask whether the household perceives that people in the community practice open defecation and we find that this perception falls in all groups by just over 5 percentage points below the mean of 84% in comparison communities. Finally, we observe an increase in the probability of correct child feces disposal, as per the JMP definition, in both sanitation treatment groups. We are concerned that results from the TSSM group may be driven by possible baseline imbalance. Controlling for baseline latrine type, we find positive TSSM results reduce slightly and converge to be more closely aligned with combination ward results, with signs and significance tests remaining the same with and without controls. These results are found in the online appendix.

6.3. Impacts on Handwashing Knowledge and Behavior

We now turn to the impacts on handwashing related outcomes. We generate an index from 0 to 1, which signifies the proportion of 5 unprompted handwashing junctures of which the caregiver is aware– after going to the latrine; after washing baby's bottom; before preparing food; before eating; before feeding/breastfeeding. We find that caregivers in the handwashing and combined groups show small but significant improvements in knowledge (Table 10). When inquiring whether the caregiver knows that handwashing with soap and water is the best method we observe an increase of over 5 percentage points in TSSM and combined wards, and unexpectedly no significant impacts in the HWWS only group. However when we turn to self-reported handwashing with soap in the last 24 hours we do not see a significant change in any of the treatment areas relative to control. Furthermore, while the presence of soap by a handwashing station is 8% in the control group, this proportion is unchanged in the handwashing and combined wards and actually declines slightly (significant at the 10% level) in TSSM wards. Households in the TSSM and combined group report higher expenditures on soap in the last month.

The intervention does not have an impact on the probability of having any form of handwashing station, including mobile stations, regardless of whether soap is present (48% of households do),

but does have a relatively large impact on the probability of having a fixed handwashing station and a handwashing facility within 6 meters of the latrine in combined intervention wards. While only 1.2% of control households have a fixed handwashing station, this proportion increases by 1.7 and 2.8 percentage points respectively in HWWS and combination groups. Most of this change is driven by the presence of "tippy taps", which we include as fixed handwashing stations. While statistically significant, these numbers remain small – a total of 22 tippy taps were observed in the entire sample, which questions the sustainability of this intervention.

Only 3.7% of control households have a handwashing facility within 6 meters of their latrine, and this proportion increases by 6.3 percentage points in combination wards (but not the stand alone groups). Though in absolute terms the presence of latrines with handwashing facilities in combination wards is still relatively small (about 10% of households), this result may reflect an interaction of the TSSM and HWWS interventions whereby households exposed to both types of messages are more likely to install new latrines with a handwashing station nearby.

To measure handwashing practice we utilize both direct observation of handwashing at critical points in time in the household, as well as self-reported measures. The number of exposure events identified through direct observation is very similar across groups and not statistically different, with an average of 5.8 exposure events observed per household. Overall, we find no differences between treatment and control groups in self-reported or observed handwashing with soap after fecal contact (Table 11). Moreover, while 47% of respondents report washing hands after fecal contact in the last 24 hours, only 12% of individuals washed hands after fecal contact in the direct observation sub-sample. The likelihood of observed handwashing with soap after fecal contact actually declines by 5.6 percentage points in TSSM wards, though this is likely related to the increased opportunity for observing otherwise unobserved handwashing behavior in households with new latrines constructed following the TSSM intervention. That being said, there is no significant decrease in observed handwashing in combination wards.

While there are no significant impacts of the program on handwashing behavior after fecal contact, we do find small increases in self-reported and observed handwashing practice before preparation of foods. There is a 7.7 percentage point increase in self-reported handwashing before food preparation in the HWWS group over the 15% reported in the control group (and no significant impacts in the TSSM or combination groups). In the direct observation sample, there is a 1.6 percentage point increase in the likelihood of observing handwashing when handling food or feeding among members of the HWWS and combination groups, over a mere 1.3% of handwashing observed at this junction in the control group.

Beyond the small positive impacts on handwashing prior to food handling, results from the direct observation of handwashing practice are not encouraging. For any exposure event observed by the enumerator, 27% were followed by any form of handwashing in the control group with no significant differences compared to any treatment group. Only 3.8% of exposure events were

followed by handwashing with soap and water, and again there are no statistically significant differences between the control and treatment groups. These results report on the full sample of individuals and observed events are largely unchanged when focusing on the behavior of the child's caregiver or the average household behavior (see online appendix).

6.4. **Impact on Other Hygiene Practices**

In addition to handwashing behaviors we measure other indicators related to hygiene practices in the household, including the observed cleanliness of children and caregiver's hands, observed animal or human feces around the house and living area, smell of feces around the house, presence of garbage, and whether food is covered. Overall, children appear to be cleaner in the HWWS and combination wards. We combine three binary indicators: whether the child has dirty hands, fingernails or face, to generate an index from 0 (dirty) to 1 (clean), which increases from a base of 0.56 for children in the control group, by 0.077 and 0.069 points respectively in the handwashing and combination wards. The results are consistent for each individual indicator comprised in the index, amounting to an increase of 5 to 8 percentage points in the likelihood of being clean across each of the three dimensions. The caregiver hand cleanliness index also shows improvement in handwashing and combination wards. Enumerators rate cleanliness on a scale of 1 (visible dirt), 2 (unclean appearance), or 3 (clean) for nails, palms and fingerpads of caregivers separately. The scores are summed to get a value between 3 and 9. We observe an increase of about 0.4 and 0.45 points in the handwashing and combination wards respectively, above a mean score of 6.7 in the control group. No changes are observed in TSSM wards.

We observe no changes in the probability of observing human or animal feces around the home (9% of households), smell of feces around the house (12% of households) or observed loose garbage in the kitchen or house (38% of households). There is a large and significant increase, however, in the likelihood that food is completely covered. While food coverage is observed in 28.3% of control households, this proportion increases significantly in all three treatment groups, with magnitudes ranging from 11.1 percentage point increase in the handwashing only group to 6.8 percentage points in the combined treatment group (impacts are not significantly different between groups).

Finally, with regard to sanitation-related behavior we find that households with latrines in the TSSM and combined groups are more likely to report having ever cleaned the latrine they use, an increase of about 7 percentage points above 80% of households that report this behavior in the control group. Furthermore, there is a reduction in the probability of observed feces in the latrine or outside of the pit in combination wards. We find a reduction in feces observed in the latrine areas outside the pit, from 22% in control households to 17.7% in treatment ones. However, there are no significant differences in the reported presence of flies in the latrine area (53.7% of households report flies being present all the time).

6.5. Health Outcomes

The TSSM and HWWS interventions are aimed at provoking handwashing and sanitation related behavior change with the ultimate goal of improving the population's health, particularly health outcomes amongst children under 5. These results are presented in Table 13. We analyze the program's impact on the probability of diarrhea in the past 7 and 14 days as a short term measure of health from exposure to pathogens; hemoglobin levels/anemia and weight as medium term measures of health; and height and head circumference as longer term indicators of the cumulative effects of improved health.

For the analysis of diarrhea we use the large-scale household listing survey measuring the existence of one or more diarrheal episodes amongst children under 5 in the past 14 days (n= 34,045), as well as the caregiver report during the in-depth household survey on symptoms of diarrhea amongst children under 5 (n=5,768). The diarrhea incidence from the listing data includes a much larger sample size and a longer time horizon when compared to the in-depth household survey data, thereby providing more power for the measure of impacts on this variable. Using the listing data, we estimate a decline in diarrhea of 2.1 percentage points (significant at the 10% level) in the combined treatment group, which constitutes a relative reduction of 12.5% compared to the control group where 16.8% of children are reported as having had diarrhea in the past 14 days. The estimated coefficients in the TSSM and HWWS only groups are negative but insignificant. However, when we analyze the household self-reported outcomes on diarrhea symptoms in the past 7 days we observe no significant differences between treatment and control groups. Diarrhea symptoms are reported for 8.6% of children in the control group, and while the coefficients on the three treatment groups are negative, they are small and statistically insignificant.

Hemoglobin (Hb) levels test for the presence of iron-deficiency anemia, which proxies for the child's nutrient intake and absorption. Nutrient absorption can be affected by intestinal pathogens that reduce the absorptive capacity of the child's intestinal tract. We observe no differences in either the level of measured Hb or the probability of anemia for children in the TSSM and HWWS only groups. However, in the combined treatment group, children show a small but significant decline of 1.65 g/L in hemoglobin (a 1.5% relative decline compared to the average of 111.4 g/L in the control group) and a significant increase in the probability of being anemic (defined as Hb<110 g/L), 6 percentage points higher than the 41.4% incidence observed in the control group. A second indicator that does not change in the singular treatment arms is weight-for-age (z-score), an indicator of children's short term health. The average weight-for-age z-score in the combined intervention group declines by 0.075 standard deviations off of an average weight-for-age z-score of -1.03. Although biologically insignificant, the unexpected direction of the effects of the program on children's health, as measured through anemia and weight, particularly in light of reduced presence of diarrhea in the same population, is a puzzle that we explore further in the next section.

Finally, our indicators for long run child health are the height-for-age z-score and head circumference-for-age z-score. On average we find no effects of the program on the height of children in any of the treatment groups relative to control and estimated coefficients are close to zero. For head circumference, estimated coefficients are positive but not significant (0.2 standard deviations in the handwashing groups). Thus, we conclude that the HWWS and TSSM interventions did not have a detectable impact on the long-term health status of beneficiary children.

6.6. Subgroup Analysis: Child Age

The time in a child's life during which he or she is exposed to a WASH intervention is likely to be an important determinant of the potential effects the intervention may have. For instance, stunting is the long-term effect of reduced nutrition at the earliest years of life, where growth faltering is most common. Younger children with less developed immune systems may also be more susceptible to disease in early life. When we separate our analysis into "younger" (ages 0, 1 and 2) and "older" (ages 3 and 4) age cohorts, we find little difference in height-for-age measures (Table 14). Absolute reductions in diarrhea for the combination wards are largest in the youngest age group, where we find a 2.6% reduction off of a control mean of 20.4%, compared to an insignificant 1.3% reduction off of a 10.6% control mean for ages 3 and 4, although the relative change is somewhat similar at 12.75% and 12.25% respectively. None of these results are statistically significant for either cohort in the TSSM or HWWS wards.

In the same age cohorts we find a larger reduction in hemoglobin levels for the younger group (-2.35 g/L) and a much smaller and statistically insignificant reduction in the older group (-0.47g/L) in combination wards. Similarly, weight-for-age reductions are stronger in the younger cohort. None of the results are statistically significant for either cohort in TSSM and HWWS wards.

It is important to note that this sub-group analysis was not included as part of a formal pre-analysis plan, and we did not stratify by age cohort in the randomization procedure, although we do find baseline balance in child and household characteristics between control and treatment wards in these subgroups. The results suggest that most of the health effects are being driven by the younger cohort.

7. Validity Checks

We may be concerned that the counterintuitive health results could be driven by data quality issues. Here we explore these concerns in relation to diarrhea (7 and 14 day prevalence), hemoglobin level,[33] height-for-age, weight-for-age and head circumference z-scores.

7.1. Courtesy Bias

While the nature of the intervention made it impossible to blind participants to their treatment status, we ensured that interviewers were blinded to the intervention status of each village in an attempt to reduce potential courtesy bias.

Objective biomarker measures help to overcome the potential of courtesy bias, which has been documented in previous literature related to WASH interventions (e.g. Schmidt, 2014; DFID, 2013). To test for potential courtesy bias in self-reported measures, we include a falsification question in the child health calendar survey instrument on whether the child has had abrasions, scrapes or bruising in the past 7 days, which we would expect to be uncorrelated with the handwashing or sanitation interventions. Results are presented in Table 15 and we find a small and borderline significant decrease in the combination ward, but no effect in the TSSM or HWWS wards suggesting potential but limited concern for reporting bias in the combined group.

7.2. Nonrandom Covariate Imbalances: Enumerator Bias and Differential Timing

Biometric measures are able to remove any potential subjective reporting bias, but we may still be concerned about non-random measurement bias given the level of skill required to measure height, weight, head circumference and anemia levels. We may also be concerned that roll-out of the survey (nine months in total) may have resulted in control and treatment households being visited at different times in the year. This would be problematic in cases where outcomes are seasonal, such as for anemia, which manifests as a symptom of malaria. We run all of the regressions for health outcomes (Table 15), and include interviewer and interview month fixed effects to account for these potential biases. We find that results remain consistent and significant for diarrhea, weight-for-age and hemoglobin level, while remaining small and insignificant for height-for-age. In fact, impacts increase slightly in all cases, and the standard errors reduce, yielding more precise estimates. Head circumference results vary somewhat across specifications, bringing into question the accuracy of this measure.

7.3. Outliers

---

[33] Being anemic is based on the cutoff of having less than 110 g/L of hemoglobin in the blood. We find that, although significant, the result is highly influenced by the cutoff level chosen, and any cutoff between 73 and 103 g/L would fail to reject a difference in means between the control and combination group (Figure 9). In this case, we prefer to conduct robustness on the raw hemoglobin levels.

Removing observations larger than 3 standard deviations from the mean for anthropometric measures discards 41, 33, 61 and 74 observations from the hemoglobin level, weight-for-age, height-for-age and head circumference z-scores respectively. Rerunning these regressions with this limited set (along with other similar specifications) yields results virtually unchanged from the original, except in the case for Hb level, where the impact is reduced and is no longer significant (Table 15). This is consistent with Figure 9 which indicates that most of the anemia change is coming from a small reduction in Hb levels close to the cutoff of 110 g/L, but there is also a left-hand tail in the combination distribution of severely anemic children (which are removed as outliers in this specification). Care should be taken when interpreting this result since removal of these outliers in the case of Hb levels may exclude important information regarding the treatment impact on the distribution of anemia.

8. Potential Mechanisms

While the generally accepted theory of change (and the one on which this study was originally based) provides clear justification for why we may expect to find positive health outcomes, it is less clear what may drive negative health outcomes. Here we consider three possible explanations: (i) differential mortality; (ii) contamination of groundwater; and (iii) poor latrine quality.

8.1. Differential Mortality

Lee *et al.* (1997) present a framework to describe how we may find negative health effects of water and sanitation interventions in high mortality settings. If the intervention reduces mortality, this may have the perverse effect of increasing the proportion of sick children, who would have otherwise died absent the intervention. If this differential selection is not accounted for, our health impacts will be biased downwards. We test for this possibility with the listing data available. If differential mortality is affecting health outcomes, we would expect a lower mortality rate in the treatment wards. While we do not have appropriate mortality measures, we do have data on the number of children under 5 in each of the 50,885 households asked this inclusion question during the listing exercise. We find no difference across any of the groups (Table 16). Child mortality may induce a fertility response, with mothers being more likely to have another child if their child passes away, thus equalizing under 5 ratios across groups. If this were the case, however, we would expect to see differences in the age distribution across groups (e.g. control households having a higher proportion of 0, 1, or 2 year olds). We run the same regression on each age group and find no evidence of different age distributions across groups, suggesting that selective mortality is not a concern in our data.

8.2. Contamination of Groundwater

Non-experimental research suggests a possible link between pit latrines and contamination of groundwater through seepage (Dzwairo *et al.*, 2006). Increasing latrine coverage could then have the potential negative effect of increasing exposure to pathogens through drinking water. In this

case, we would expect to find a differential impact of the program depending on whether households treat their water before drinking it. Running sub-group analysis on our health outcomes based on whether the household treats their water (40% of households do), we find no observable differences in health outcomes, suggesting that this transmission mechanism is unlikely (Table 17).

8.3. **Poor Latrine Quality**

An important current topic of debate is whether encouraging households to take their first step onto the "sanitation" ladder by promoting fixed point defecation could be more harmful than helpful, by way of localizing fecal content, bringing it closer to home and providing more opportunity for flies to breed and spread disease. The concern is that, if we encourage households to build latrines, and these latrines are of a very low quality, we may exacerbate contamination opportunities. Our survey instrument has a rich set of measures related to observed and reported latrine quality. Running regressions on a wide range of these variables suggests that there is no clear difference in latrine quality across groups and, if anything, TSSM and combination ward latrines may be slightly better quality than latrines in the control wards, with both groups being more likely to have a squat hole cover in place during the enumerator observation and the combination group having less visible feces surrounding the latrine (Table 17).

9. Discussion

In summary, the results from this study allow us to identify a few important facts to help reflect on potential implications for intervention implementation and future research in this area.

1. **TSSM is able to change behavior but not by enough to significantly influence the level of observed fecal matter:** We see significant increases in improved latrine coverage. However, this does not necessarily translate into a more hygienic environment, and it is not clear that this increase in coverage is enough to make a difference to the daily fecal exposure that children face. While open defecation reduces substantially, the reduction is mostly driven by people limiting regular open defecation, while occasional open defecation does not change significantly and remains high at 51.3% in the control wards. In the TSSM intervention areas, we find fewer than 20% of villages reporting to be ODF. Furthermore, 75% of households have animals and 55% allow these animals into their house. New evidence suggests that children in contact with animal feces face exposure to high levels of bacteria such as *E. coli* (Ngure *et al.*, 2013), which suggests that a more holistic approach to fecal removal may be required to effectively cut the transmission mechanism. Unsurprisingly, then, observed feces around the dwelling does not change across treatment groups and having a latrine is not correlated with whether feces is observed.

2. **At-scale handwashing campaigns produce significantly lower effects on health outcomes than efficacy trials. This is likely to result from the limited effectiveness of being able to change handwashing behaviors:** The HWWS intervention presents results consistent with the current literature on large effectiveness trials and suggests that significantly changing hygiene behavior through promotional activities remains a somewhat challenging task. We do find reasonably consistent and positive impacts on a range of hygiene-related indicators, but the magnitude is small and diminishes as we go further along the causal chain. It is clear that there is a large wedge between knowledge and behavior, with 82% of caregivers knowing the importance of washing hands with soap, but only 3.8% of exposure events actually being accompanied by this practice. Current large-scale behavior change campaigns seem unable to effectively close this gap between knowledge and practice.

3. **The tippy taps designed for Tanzania are not sustainable:** Since change in knowledge is not enough to shift behaviors, the presence of visual cues and simplifying the burden of handwashing were promoted through the campaign. This was the rationale for promoting the construction and use of tippy taps. However, while anecdotal evidence suggests that there was initially high take up, we find that only 22 households of a possible 1,829 (1.2%) exposed to the HWWS campaign were using them at the time of follow up.

4. **The messages promoted by the campaigns do not translate into predictable behavior change:** The TSSM intervention promoted the upgrading, rather than construction of new latrines, with the *sungura* slab being produced by masons as an add-on to current household facilities. In reality households chose to "upgrade" by building their own private latrines rather than sharing with others. HWWS focused on washing hands with soap at critical junctures. There seems to be little evidence to suggest that this resulted in handwashing behavior change, and yet we do find some (weak) evidence on broader measures of hygiene improvements such as latrine cleanliness. This may result from two things: (1) message creep – regardless of how nuanced the design of the message may be, at scale it becomes more difficult to control the delivery of the message (Coombes & Paynter, 2010); and (2) personal interpretation – while designers and implementers may have a clear hypothesis of how we expect people to react, the same information may be interpreted, and acted upon, in different ways.

5. **Combining handwashing and sanitation interventions does not produce clear health benefits:** No detectable health effects are found in the individual interventions. When turning to the combined treatment, diarrhea levels improve marginally, but anemia and weight-for-age outcomes worsen. Putting this in perspective, the anemia and weight-for-age results, while statistically significant, do not seem to be biologically important – they amount to weight reductions of less than

100 grams per child on average, and hemoglobin reductions of 1.6 g/L. However, a diarrhea reduction of 2.1 percentage points, when taken at face value, can be considered biologically important, although this measure is only marginally significant (statistically) and does come with reasonable concerns linked to self-reporting, making it less reliable than the biometric markers.

The implications for policy and program implementation require some extrapolation from the data and should therefore be considered only as suggestive. If the primary objective of the policymaker is to reduce child morbidity and mortality, then intervention intensity needs to be considered carefully since there is a tradeoff between going to scale and delivering interventions effective enough to generate health impacts through TSSM and HWWS behavior change campaigns. Exactly how to increase intensity is beyond the scope of this study. Broadly this could focus on doing more of the same interventions more intensively (e.g. multiple follow ups within the same villages) or including a broader range of tools (e.g. sanitation subsidies, animal fecal containment practices).

If, on the other hand, the primary objective for policy makers is increased coverage of improved practices and materials in their own right (reducing open defecation or increasing handwashing with soap), then TSSM seems to be an effective tool to deliver on these objectives, whereas changing behavior through HWWS remains a challenge. In summary, the evidence would suggest that, whichever objective the policymaker seeks to achieve (increased coverage or health improvements), HWWS should be a complementary rather than stand-alone activity.

10. Conclusion

Three systematic reviews over the past decade have shown that handwashing with soap consistently reduces diarrhea by between 39% and 47% (Fewtrell *et al*, 2003; Curtis & Cairncross*, 2005; Ejimot-Nwadiaro *et al*, 2008). This evidence, however, comes mostly from small-scale efficacy trials exploring proof of concept, or matched studies with high levels of handwashing compliance. While useful, it is not clear how scalable these interventions are, and thus how much national policy can and should draw from this evidence when designing at-scale programs where handwashing compliance is more limited. One outlying piece of recent evidence using emotional drivers in India has found sustained changes in handwashing practices at scale but does not measure resulting health impacts (Biran *et al.*, 2014). Evidence on rural sanitation is less comprehensive, and a series of ongoing evaluations will provide evidence on sanitation efficacy in the coming years. Moving beyond efficacy trials, this study explores, for the first time, what can be achieved *at scale* through independent and combined sanitation and handwashing campaigns and builds on evidence from the global WSP program looking at the impacts of government-led, at-scale interventions in Vietnam and Peru (HWWS), and India and Indonesia (TSSM). Consistent with the evidence from these evaluations, we find significant improvements in intermediate outcomes resulting from the TSSM intervention (latrine

construction and reduction in open defecation) and very limited improvements in intermediate outcomes resulting from the HWWS intervention (building handwashing stations and washing hands with soap at critical junctures). Also consistent with the programmatic evidence, neither intervention on its own is able to measurably improve health outcomes for children under 5. It is only in the combined intervention where health impacts are observed; however the inconsistency in these outcomes and limited biological significance of the point estimates suggest no clear health benefits from the approach.

The results from this study highlight the importance of focusing on intermediate outcomes of take up and behavior change as a critical first step before realizing the changes in health that WASH interventions aim to deliver. Finding the balance between intensity, the right incentives, holistic coordination and scale becomes an important policy question. The biological reasoning behind promoting WASH interventions is theoretically sound, but identifying ways to close the gap between objectives, intervention design and delivery, particularly when working at scale, should be the priority for researchers, policy makers and implementers alike.

11. **References**

Andres, L. A., Briceno, B., Chase, C., & Echenique, J. A. (2014). *Sanitation and externalities: evidence from early childhood health in rural India* (No. 6737). The World Bank.

Arnold BF, Null C, Luby SP, *et al*. (2013) Cluster-randomized controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. BMJ Open 2013; 3(8):e003476. doi:10.1136/bmjopen-2013-003476

Baqui AH, Black RE, Yunus M, Hoque AR, Chowdhury HR, *et al*. (1991) Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. Int J Epidemiol 20: 1057–1063.

Bartram J, Cairncross S (2010). Hygiene, sanitation and water: Forgotten foundation of heatlh. PLoS Med 7(11): e1000367. Doi:10.371/journal.pmed.1000367

Biran, A., Schmidt, W. P., Varadharajan, K. S., Rajaraman, D., Kumar, R., Greenland, K., Gopalan, B., Aunger, R. & Curtis, V. (2014). Effect of a behaviour-change intervention on handwashing with soap in India (SuperAmma): a cluster-randomised trial. The Lancet Global Health, 2(3), e145-e154.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. American economic journal: applied economics, 200-232.

Cairncross, S., J. Bartram, O. Cumming, and C. Brocklehurst (2010). "Hygiene, sanitation, and water: what needs to be done?," PLoS Med Nov 16;7(11).

Cameron, L., Shah, M., & Olivia, S. (2013). Impact evaluation of a large-scale rural sanitation project in Indonesia.

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

Clasen TF, Bostoen K, Schmidt WP, Boisson S, Fung ICH, Jenkins MW, Scott B, Sugden S, Cairncross S. (2010) Interventions to improve disposal of human excreta for preventing diarrhoea. Cochrane Database of Systematic Reviews 2010, Issue 6. Art. No.: CD007180. DOI: 10.1002/14651858.CD007180.pub2.

Clasen T, Schmidt WP, Rabie T, Roberts I and Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. BMJ 334(7597): 782.

Coombes, Y and Paynter, N (2010). Scaling Up Handwashing Tanzania: A Handwashing Behavior Change Journey. The World Bank Water and Sanitation Program, Washington, D.C.

Coville, A., & Orozco, V. (2014). Moving from efficacy to effectiveness: using behavioural economics to improve the impact of WASH interventions.Waterlines, 33(1), 26-34.

Curtis V and Cairncross S (2003) Effect of washing hands with soap on diarrhea risk in the community: a systematic review. Lancet Infect Dis 3(5): 275–81.

Do, Q. T., & Chase, C. (2012). Handwashing behavior change at scale: evidence from a randomized evaluation in Vietnam.

DFID (2011) "Evidence Paper. Water, Sanitation and Hygiene." DFID Literature Review. September 9, 2011.

DFID (2013) "Evidence Paper. Water, Sanitation and Hygiene." May, 2013.

Dzwairo, B., Hoko, Z., Love, D., & Guzha, E. (2006). Assessment of the impacts of pit latrines on groundwater quality in rural areas: A case study from Marondera district, Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(15), 779-788.

Ejemot-Nwadiaro RI, Ehiri JE, Meremikwu MM, Critchley JA. (2008) Hand washing for preventing diarrhoea. Cochrane Database of Systematic Reviews 2008, Issue 1. Art. No.: CD004265. DOI: 10.1002/14651858.CD004265.pub2.

Fewtrell, L., R. B. Kaufmann, D. Kay, W. Enanoria, L. Haller, and J.M.C. and Colford (2005) "Water, sanitation, and hygiene interventions to reduce diarrhea in less developed countries: a systematic review and meta analysis." Lancet Infectious Diseases; Vol. 5. No. 1: 42-52.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure Data—Or tears: An application to educational enrollments in states of india*. Demography, 38(1), 115-132.

Galiani, S., Gertler, P., & Orsola-Vidal, A. (2012). *Promoting handwashing behavior in Peru: The effect of large-scale mass-media and community level interventions* (No. 6257). The World Bank.

Greene, L. E., Freeman, M. C., Akoko, D., Saboori, S., Moe, C., & Rheingans, R. (2012). Impact of a school-based hygiene promotion and sanitation intervention on pupil hand contamination in Western Kenya: a cluster randomized trial. *The American journal of tropical medicine and hygiene*, *87*(3), 385.

Hammer, J., & Spears, D. (2013). Village sanitation and children's human capital: evidence from a randomized experiment by the Maharashtra government.

Humphrey, J. H. (2009). Child undernutrition, tropical enteropathy, toilets, and handwashing. *The Lancet*, *374*(9694), 1032-1035.

International Initiative for Impact Evaluation (2009). "Running Water, working toilets and safe hygiene practices: Essentials services to save lives." 3ie Enduring Questions Brief No. 10. August, 2009.

Independent Evaluation Group (2008). "What Works in Water Supply and Sanitation? Lessons from Impact Evaluations." IEG Word Bank Fast Track Brief. July 3, 2008

Independent Evaluation Group (2008). "What Works in Water Supply and Sanitation: Lessons from Impact Evaluation." World Bank, Washington D.C.

Kremer M. and A. P. Zwane (2007). "What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review." World Bank Research Observer; 22(1): 1 - 24.

Lee, L. F., Rosenzweig, M. R., & Pitt, M. M. (1997). The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations. *Journal of Econometrics*, *77*(1), 209-235.

Luby SP, Agboatwalla M, Painter J, Altaf A, Billhimer W, Keswick B, Hoekstra RM (2006) Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. Trop Med Int Health. 2006 Apr; 11(4):479-89.

Murray *et al*. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012; 380: 2197–223

Ngure, F. M., Humphrey, J. H., Mbuya, M. N., Majo, F., Mutasa, K., Govha, M., ... & Stoltzfus, R. J. (2013). Formative research on hygiene behaviors and geophagy among infants and young children and implications of exposure to fecal bacteria. *The American journal of tropical medicine and hygiene*, *89*(4), 709-716.

Patil, S. R., Arnold, B. F., Salvatore, A. L., Briceno, B., Ganguly, S., Colford Jr, J. M., & Gertler, P. J. (2014). The Effect of India's Total Sanitation Campaign on Defecation Behaviors and Child Health in Rural Madhya Pradesh: A Cluster Randomized Controlled Trial. PLoS medicine, 11(8), e1001709.

Perez, E, Cardosi, J, Coombes, Y, Devine, J, Grossman, A, Kullmann, C, Kumar, CA, Mukherjee, N, Prakash, M, Robiarto, A, Setiawan, D, Singh, U, and Wartono D (2012). What does it take to scale up rural sanitation, WSP working paper, June 2012.

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J. M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A. D., De France, J., Fewtrell, L., Freeman, M. C., Gordon, B., Hunter, P. R., Johnston, R. B., Mathers, C., Mäusezahl, D., Medlicott, K., Neira, M., Stocks, M., Wolf, J. and Cairncross, S. (2014), Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. Tropical Medicine & International Health. doi: 10.1111/tmi.12329

Rabie, T. and Curtis, V. (2006), Handwashing and risk of respiratory infections: a quantitative systematic review. Tropical Medicine & International Health, 11: 258–267. doi: 10.1111/j.1365-3156.2006.01568.x

Schmidt, W. P. (2014). The elusive effect of water and sanitation on the global burden of disease. *Tropical Medicine & International Health*, *19*(5), 522-527.

Victora CG, Adair L, Fall C, *et al*. Maternal and child undernutrition: consequences for adult health and human capital. Lancet 2008;371:340–57

Waddington H, Snilstveit B, White H and Fewtrell L (2009) Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries. Synthetic review 1. New Delhi: 3ie.

WHO (2011) Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Vitamin and Mineral Nutrition Information System. Geneva, World Health Organization.

WHO-Unicef (2013a) Levels and Trends in Child Mortality. Report 2013. World Health Organization

WHO-Unicef (2013b). Progress on sanitation and drinking-water - 2013 update. World Health Organization

World Bank (1996) Tanzania: Social Sector Review. Report No. 14039-TA. The World Bank. Washington, D.C.

World Health Organization and UNICEF (2008). "Progress on Drinking Water and Sanitation – Special Focus on Sanitation." At http://www.who.int/water_sanitation_health/monitoring/jmp2008/en/index.html

Table 1: Comparison of IE data to nationally representative surveys

Variable	Source	Source data	IE data
Demographics			
% HH Head that are Female	DHS	23.4%	13.6%
Avg HH size	DHS	5.40	4.90
Housing Characteristics			
Flooring material			
Earth	DHS	80.9%	82.5%
Concrete	DHS	17.5%	14.9%
Cooking fuel			
Paraffin/Kerosene	DHS	0.5%	0.5%
Charcoal	DHS	8.5%	8.5%
Firewood	DHS	90.1%	91.0%
Access			
HH has access to improved water source	DHS	46.2%	33.3%
Assets			
Electric Generator	DHS	4.3%	1.4%
Radio	DHS	57.1%	67.7%
TV	DHS	5.3%	2.9%
Mobile Phone	DHS	51.3%	58.9%
Non-Mobile Phone	DHS	0.2%	0.4%
Iron	DHS	19.6%	13.5%
Refrigerator	DHS	1.2%	1.3%
Bike	DHS	50.9%	62.7%
Motorcycle	DHS	4.9%	5.1%
Car or Truck	DHS	1.1%	0.6%
Net School Enrollment Rate			
Pre-Primary	National Panel	21.0%	20.4%
Primary	National Panel	79.0%	67.9%
Secondary	National Panel	20.0%	16.2%
Higher	National Panel	1.0%	0.0%
Health			
Anemia			
% of children under 5 with Hemoglobin less than 8 grams per deciliter	DHS	5.4%	2.1%
Anthropometrics			
Height for Age 2SD Below	DHS	44.5%	47.5%
Weight for Height 2SD Below	DHS	4.8%	2.5%
Weight for Age 2SD Below	DHS	16.9%	15.4%
Height for Age 3SD Below	DHS	17.7%	17.1%
Weight for Height 3SD Below	DHS	1.3%	0.4%
Weight for Age 3SD Below	DHS	4.1%	3.2%
Height for Age Mean	DHS	-1.80	-1.95
Weight for Height Mean	DHS	0.00	0.02
Weight for Age Mean	DHS	-1.00	-1.03

Notes: DHS = 2010 Demographic and Health Survey; National Panel = Tanzania National Panel Survey (LSMS) 2010 - 2011

Table 2: Program outputs

District	Population size	Wards sampled	TSSM / Combination Wards	Masons trained	Villages with trained masons	Number of CLTS facilitators trained	Number of communities triggered	People reached in TSSM DCC	HWWS / Combination Wards	Frontline activators trained	People reached in DCC (1)	People reached in DCC (2)
Igunga	399,727	24	12	42	42	32	282	36,500	12	40	35,509	29,093
Karagwe	332,020	24	12	50	49	40	465	21,700	12	46	16,893	39,138
Musoma	178,356	24	12	47	48	40	289	NA	12	50	25,441	37,229
Rufiji	247,993	16	8	49	46	45	185	34,950	8	45	8,783	14,483
Masasi	217,274	15	7	36	34	50	217	28,100	8	42	25,283	13,935
Iringa	254,032	12	6	28	30	18	168	15,580	6	28	14,471	15,071
Sumbawa	305,846	16	8	40	144	24	211	21,350	8	32	29,219	29,077
Kiteto	244,669	8	4	13	30	18	50	5,700	4	53	3,936	10,403
Kondoa	269,704	30	15	76	149	5	183	7,600	15	67	24,418	32,016
Mpwapwa	305,056	12	6	26	58	10	63	9,200	6	30	36,618	13,631
Total	**2,754,677**	181	90	407	630	282	2,113	180,680	91	433	220,571	234,076

Table 3: Program contamination (monitoring data)

	Planned	Lost Wards (Masasi)	Final Assignment	Contamination	Actually treated	Not triggered by CLTS	Didn't receive TSSM DCC	Didn't receive HWWS DCC
Control	48	2	46	-	46	-	-	-
HWWS	47	2	45	-	45	-	-	2
TSSM	47	3	44	-1	43	7	5	-
HWWS + TSSM	48	2	46	+1	47	4	4	3

Notes: no reported problems with (i) mason training or (ii) training of front-line activators (FLAs); no available information for delivery of village-level media (wall drawings, posters, etc.)

Variable	Overall N	Overall Mean	Control N	Control Mean	HWWS N	HWWS Mean	HWWS p-value	TSSM N	TSSM Mean	TSSM p-value	HWWS+TSSM N	HWWS+TSSM Mean	HWWS+TSSM p-value
Dwelling characteristics													
A household member owns the dwelling	3618	0.90	918	0.91	900	0.87	0.11	880	0.89	0.47	920	0.91	0.88
Household uses a clean form of energy for lighting (electricity, solar, gas or batteries)	3619	0.24	919	0.29	900	0.27	0.72	880	0.20	0.09	920	0.22	0.22
Household uses electricity as main energy source for lighting	3619	0.03	919	0.03	900	0.05	0.32	880	0.03	0.95	920	0.03	1.00
Household uses paraffin lamps as main lghiting source	3619	0.74	919	0.70	900	0.72	0.68	880	0.78	0.12	920	0.76	0.23
Main fuel used for cooking is charcoal	3619	0.08	919	0.08	900	0.12	0.34	880	0.05	0.15	920	0.08	0.77
Main fuel used for cooking is firewood	3619	0.91	919	0.91	900	0.87	0.26	880	0.94	0.23	920	0.91	0.94
Floor of main living area is made of cement	3613	0.17	919	0.15	897	0.23	**0.04**	877	0.15	0.99	920	0.16	0.69
Floor of main living area is made of earth/clay	3613	0.80	919	0.82	897	0.74	**0.03**	877	0.84	0.54	920	0.81	0.73
Roof of main living area is made of mud/thatch/grass	3619	0.43	919	0.48	900	0.40	0.16	880	0.42	0.28	920	0.40	0.19
Roof of main living area is made of tin/zinc	3619	0.57	919	0.52	900	0.60	0.18	880	0.58	0.31	920	0.59	0.20
Walls of main living area are made of brick/cement	3603	0.71	917	0.67	889	0.75	0.22	879	0.73	0.39	918	0.72	0.49
Walls of main living area are made of mud and poles	3603	0.28	917	0.33	889	0.24	0.21	879	0.27	0.36	918	0.27	0.38
HH Head Characteristics													
Male	3614	0.88	919	0.86	898	0.87	0.87	878	0.89	0.25	919	0.88	0.27
Age	3614	40.61	919	39.87	898	41.23	0.10	878	40.19	0.68	919	41.13	0.17
Age Squared	3614	1814.22	919	1750.67	898	1865.33	0.14	878	1768.85	0.80	919	1871.16	0.19
Ever attended school	3612	0.77	919	0.76	897	0.79	0.34	878	0.79	0.25	918	0.76	0.91
Can read and write	3614	0.74	919	0.71	898	0.76	0.13	878	0.75	0.16	919	0.73	0.55
Years of Education (if attended school)	2796	6.48	695	6.47	706	6.57	0.43	697	6.44	0.87	698	6.43	0.79
Muslim	3618	0.31	919	0.29	899	0.29	0.98	880	0.31	0.78	920	0.35	0.44
Christian	3618	0.60	919	0.60	899	0.62	0.79	880	0.62	0.85	920	0.57	0.68
Born in this village	3614	0.67	919	0.68	898	0.68	0.90	878	0.64	0.35	919	0.68	0.93
Number of years living in this village	1191	18.26	295	17.56	284	18.23	0.64	320	18.61	0.40	292	18.64	0.43
Married	3614	0.75	919	0.72	898	0.73	0.77	878	0.76	0.43	919	0.78	0.22
Divorced/widowed/separated	3614	0.11	919	0.12	898	0.12	0.97	878	0.10	0.14	919	0.10	0.22
Lives with partner	3614	0.13	919	0.15	898	0.13	0.76	878	0.14	0.87	919	0.11	0.45
HH Member Demographics													
HH size at Baseline	3609	4.94	917	4.89	898	4.94	0.82	877	5.02	0.55	917	4.93	0.86
Age	22159	17.78	5593	17.47	5482	18.03	0.08	5470	17.52	0.85	5614	18.11	**0.03**
Age Squared	22159	597.34	5593	580.72	5482	615.94	0.14	5470	576.69	0.85	5614	615.86	0.14
Member is Male	22161	0.49	5593	0.50	5482	0.49	0.91	5472	0.50	0.96	5614	0.49	0.62
Age of caregiver	3801	30.65	965	30.39	946	30.60	0.65	925	30.79	0.35	965	30.82	0.32
Age of caregiver Squared	3801	1026.28	965	1018.55	946	1022.03	0.92	925	1030.06	0.74	965	1034.55	0.65
Caregiver is male	3801	0.01	965	0.01	946	0.01	0.09	925	0.02	0.45	965	0.01	0.31

	Overall		Control		HWWS			TSSM			HWWS+TSSM		
Variable	N	Mean	N	Mean	N	Mean	p-value	N	Mean	p-value	N	Mean	p-value
HH Income Sources at Baseline													
Most Important income: Employment (Paid Employee)	3619	0.03	919	0.03	900	0.03	0.57	880	0.04	0.15	920	0.02	0.32
Most Important income: Employment (Self-Empoyment with employees)	3619	0.06	919	0.06	900	0.06	0.91	880	0.05	0.55	920	0.06	0.85
Most Important income: Not-employed (Remmitances)	3619	0.00	919	0.00	900	0.00	0.55	880	0.00	0.98	920	0.00	0.31
Most Important income: Self Employed Agricultural	3619	0.74	919	0.74	900	0.72	0.68	880	0.75	0.72	920	0.76	0.64
Latrine Type at Baseline													
Flush/pour latrine	3618	0.03	919	0.02	900	0.05	0.05	879	0.02	0.82	920	0.04	0.34
No toilet facilities (open defecation)	3618	0.12	919	0.17	900	0.13	0.48	879	0.06	**0.01**	920	0.09	0.09
Pit latrine with slab or VIP	3618	0.84	919	0.80	900	0.81	0.84	879	0.91	**0.01**	920	0.86	0.23
Axcces to Water													
Main source of drinking water comes from piped water	3619	0.14	919	0.08	900	0.22	**0.01**	880	0.12	0.23	920	0.14	0.12
Main source of drinking water comes from a well or borehole	3619	0.32	919	0.37	900	0.26	0.06	880	0.31	0.39	920	0.34	0.62
Main source of drinking water comes from surface water	3619	0.40	919	0.42	900	0.38	0.54	880	0.38	0.54	920	0.42	0.98
Is the water source covered	3127	0.22	849	0.27	706	0.25	0.88	777	0.19	0.18	795	0.16	0.08
Number of person trips HH makes to fetch water per day	3477	2.77	886	2.75	838	2.66	0.65	857	2.85	0.59	896	2.81	0.73
Number of times out of the last 10 attempts water has not been available	3617	0.30	919	0.25	899	0.33	0.39	880	0.27	0.81	919	0.35	0.28
HH stores water	3619	0.93	919	0.91	900	0.94	0.18	880	0.95	0.12	920	0.93	0.55
HH treats their water	3619	0.40	919	0.37	900	0.39	0.80	880	0.41	0.46	920	0.44	0.22
Assets owned by HH at Baseline													
Wealth index at Baseline	3619	0.00	919	0.09	900	0.13	0.13	880	-0.34	0.35	920	0.10	0.60
Household listens to the radio regularly	3619	0.61	919	0.55	900	0.60	0.30	880	0.64	0.07	920	0.64	**0.04**
Total number of assets	3619	44.18	919	44.44	900	43.95	**0.03**	880	44.09	0.17	920	44.22	0.35
Another House	3615	0.15	918	0.14	900	0.16	0.36	878	0.17	0.19	919	0.14	0.86
Radio/CD/Cassette Player	3616	0.68	919	0.65	900	0.67	0.55	878	0.71	0.07	919	0.67	0.51
TV	3617	0.03	919	0.03	900	0.04	0.49	878	0.03	0.99	920	0.02	0.46
Bicycle	3617	0.63	919	0.62	900	0.60	0.59	878	0.62	0.87	920	0.67	0.24
Motorcycle	3616	0.05	919	0.04	900	0.07	0.13	878	0.04	0.94	919	0.06	0.25

Variable	Overall		Control		HWWS			TSSM			HWWS+TSSM		
	N	Mean	N	Mean	N	Mean	p-value	N	Mean	p-value	N	Mean	p-value
Assets owned by HH at Baseline (cont)													
Car	3617	0.01	919	0.00	900	0.01	0.09	878	0.00	0.73	920	0.01	0.51
Elctric/Gas Stove	3617	0.00	919	0.00	900	0.01	0.70	878	0.00	0.95	920	0.00	0.17
Other Stove	3617	0.25	919	0.20	900	0.30	**0.01**	878	0.26	0.06	920	0.25	0.13
Refridgerator	3617	0.01	919	0.01	900	0.01	0.31	878	0.01	0.79	920	0.02	0.70
Mattress	3617	0.72	919	0.66	900	0.75	0.05	878	0.72	0.23	920	0.73	0.14
Sewing Machine	3616	0.05	918	0.04	900	0.06	0.14	878	0.05	0.93	920	0.05	0.93
Mosquito Net	3617	0.88	919	0.88	900	0.89	0.74	878	0.86	0.43	920	0.87	0.59
Mobile Phone	3617	0.48	919	0.45	900	0.51	0.08	878	0.47	0.54	920	0.49	0.27
Fixed-line phone	3617	0.00	919	0.00	900	0.00	0.31	878	0.00	0.65	920	0.01	0.18
Iron	3616	0.13	919	0.11	899	0.14	0.22	878	0.15	0.10	920	0.12	0.71
Bed frame	3617	0.70	919	0.67	900	0.73	0.08	878	0.71	0.29	920	0.69	0.59
Jewellery	3616	0.01	919	0.01	899	0.01	0.59	878	0.00	0.16	920	0.01	0.44
Land or Field	3616	0.66	919	0.67	900	0.69	0.61	877	0.67	0.94	920	0.63	0.36
Agricultural Equipment	3617	0.16	919	0.16	900	0.16	0.90	878	0.15	0.91	920	0.17	0.76
Electricity Generator	3617	0.01	919	0.02	900	0.01	0.09	878	0.01	0.06	920	0.01	0.08
Solar Panel	3617	0.01	919	0.02	900	0.01	**0.03**	878	0.01	0.47	920	0.01	0.37
Sponged Sofa	3617	0.07	919	0.06	900	0.09	0.17	878	0.06	0.87	920	0.06	0.96
Sofa (non-sponged)	3617	0.09	919	0.08	900	0.11	0.40	878	0.08	0.89	920	0.09	0.89
Fan	3617	0.01	919	0.00	900	0.01	0.39	878	0.00	0.74	920	0.01	0.83
Camera	3617	0.00	919	0.00	900	0.00	0.46	878	0.01	0.21	920	0.00	0.70
Number of Animals Owned by HH at Baseline													
Total Number of Animals Owned by HH	3619	16.12	919	17.11	900	16.57	0.86	880	15.09	0.42	920	15.68	0.59
Cows	3619	2.38	919	3.32	900	2.45	0.52	880	1.66	0.16	920	2.04	0.29
Bulls	3619	1.22	919	1.90	900	1.01	0.17	880	0.89	0.10	920	1.09	0.20
Donkeys	3619	0.14	919	0.21	900	0.13	0.41	880	0.10	0.20	920	0.12	0.28
Goats	3619	3.31	919	3.28	900	3.94	0.43	880	2.77	0.43	920	3.23	0.94
Sheep	3619	0.68	919	0.85	900	0.59	0.36	880	0.38	0.07	920	0.87	0.96
Pigs	3619	0.30	919	0.30	900	0.26	0.68	880	0.34	0.73	920	0.32	0.85
Chickens	3619	7.39	919	6.77	900	7.43	0.40	880	8.22	**0.04**	920	7.18	0.53
Ducks	3619	0.46	919	0.27	900	0.53	0.08	880	0.36	0.45	920	0.66	**0.03**
Geese	3619	0.00	919	0.00	900	0.01	0.31	880	0.01	0.31	920	0.00	-
Rabbits	3619	0.04	919	0.06	900	0.07	0.84	880	0.01	0.15	920	0.01	0.11
Guinea Fowls	3619	0.06	919	0.05	900	0.04	0.70	880	0.10	0.44	920	0.06	0.91
Other Animals	3619	0.14	919	0.10	900	0.10	0.99	880	0.24	0.41	920	0.11	0.93

Notes: p-values are based on a simple t-test comparison between control and relevant treatment group with standard errors clustered at the ward level. P-values < 0.05 are in bold

VARIABLES	HWWS				TSSM			
	High (3 channels) (1)	Medium (>= 2 channels) (2)	Low (>= 1 channel) (3)	Number of channels (4)	High (3 channels) (5)	Medium (>= 2 channels) (6)	Low (>= 1 channel) (7)	Number of channels (8)
HWWS (β1)	0.071***	0.212***	0.373***	0.656***	0.068***	0.193***	0.251***	0.512***
	(0.012)	(0.024)	(0.033)	(0.061)	(0.012)	(0.024)	(0.031)	(0.057)
TSSM (β2)	0.006	0.036*	0.100***	0.142***	0.026**	0.080***	0.141***	0.247***
	(0.008)	(0.021)	(0.025)	(0.047)	(0.010)	(0.019)	(0.029)	(0.050)
HWWS + TSSM (β3)	0.061***	0.254***	0.442***	0.758***	0.103***	0.266***	0.345***	0.714***
	(0.011)	(0.026)	(0.026)	(0.055)	(0.014)	(0.024)	(0.027)	(0.057)
p-values $\beta_1 = \beta_2$	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000
for F-test $\beta_1 = \beta_3$	0.500	0.180	0.061	0.160	0.030	0.015	0.006	0.004
$\beta_2 = \beta_3$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	3,619	3,619	3,619	3,619	3,619	3,619	3,619	3,619
R-squared	0.071	0.168	0.238	0.245	0.077	0.141	0.186	0.191
Control Mean	0.001	0.014	0.135	0.150	0.004	0.040	0.224	0.269

Notes: Robust standard errors in parentheses clustered at the ward level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects.

Table 6: TSSM awareness

VARIABLES	Aware of a mason in the community (1)	Aware of a CLTS committee in the village (2)	Agrees with statement: Everybody in this community knows somebody that can be paid to build or improve a latrine (3)
HWWS (β1)	0.094***	0.060**	0.027
	(0.032)	(0.023)	(0.037)
TSSM (β2)	0.183***	0.131***	0.081**
	(0.036)	(0.026)	(0.035)
HWWS + TSSM (β3)	0.188***	0.130***	0.069*
	(0.031)	(0.028)	(0.035)
p-values $\beta_1 = \beta_2$	0.020	0.005	0.138
for F-test $\beta_1 = \beta_3$	0.004	0.012	0.251
$\beta_2 = \beta_3$	0.871	0.972	0.738
Observations	3,619	3,619	3,619
R-squared	0.133	0.288	0.072
Control Mean	0.143	0.119	0.297

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects.

Table 7: Latrine construction and availability

| VARIABLES | Household has a […] latrine that was constructed within the last 3 years | | | Household has a […] latrine that was constructed more than 3 years ago | | | Household uses a shared latrine |
| | any | private | shared | any | private | shared | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HWWS (β1)	-0.015	0.026	-0.042**	0.027	0.008	0.019	-0.031
	(0.035)	(0.030)	(0.020)	(0.024)	(0.017)	(0.013)	(0.030)
TSSM (β2)	0.082**	0.124***	-0.042**	0.035	0.036**	-0.001	-0.092***
	(0.034)	(0.032)	(0.019)	(0.022)	(0.018)	(0.011)	(0.027)
HWWS + TSSM (β3)	0.077**	0.100***	-0.022	-0.009	-0.002	-0.007	-0.076***
	(0.031)	(0.028)	(0.019)	(0.022)	(0.018)	(0.010)	(0.026)
p-values β1 = β2	0.006	0.003	0.972	0.731	0.083	0.158	0.028
for F-test β1 = β3	0.006	0.013	0.302	0.150	0.592	0.052	0.098
β2 = β3	0.892	0.434	0.267	0.067	0.04	0.602	0.465
Observations	3,469	3,469	3,469	3,469	3,469	3,469	2,974
R-squared	0.121	0.108	0.041	0.044	0.039	0.032	0.074
Control Mean	0.571	0.386	0.185	0.181	0.134	0.046	0.332

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects.

Table 8: Presence of an improved latrine

VARIABLES	Improved latrine (including shared facilities)		Improved latrine (JMP definition)		Improved latrine (strict definition)	
	Reported	Observed	Reported	Observed	Reported	Observed
	(1)	(2)	(3)	(4)	(5)	(6)
HWWS (β1)	0.038	0.035	0.049	0.046	0.021	0.028
	(0.040)	(0.039)	(0.034)	(0.034)	(0.029)	(0.030)
TSSM (β2)	0.124***	0.117***	0.157***	0.151***	0.122***	0.086***
	(0.039)	(0.039)	(0.034)	(0.035)	(0.029)	(0.028)
HWWS + TSSM (β3)	0.074*	0.077**	0.103***	0.106***	0.076***	0.057**
	(0.038)	(0.038)	(0.031)	(0.031)	(0.029)	(0.026)
p-values $\beta_1 = \beta_2$	0.023	0.030	0.001	0.003	0.000	0.076
for F-test $\beta_1 = \beta_3$	0.337	0.262	0.091	0.058	0.055	0.339
$\beta_2 = \beta_3$	0.176	0.283	0.088	0.161	0.111	0.327
Observations	3,619	3,619	3,619	3,619	3,619	3,619
R-squared	0.230	0.214	0.156	0.152	0.151	0.093
Control Mean	0.761	0.749	0.507	0.497	0.373	0.270

Notes: Robust standard errors in parentheses clustered at the ward level.*** p<0.01, ** p<0.05, * p<0.1.

Coefficients estimated using a linear probability model including block fixed effects.

Definition of improved latrine is based on JMP standard definition. Broad definition considers only the latrine type, while the strict definition requires that the slab is suitable for separating humans from excreta (there is only one whole in the slab and it is in good repair).

Table 9: Household and community open defecation practices

VARIABLES	HH members usually defecate in fields/bushes/rivers	HH members practice open defecation…		HH is aware of community members practicing OD	Child feces are safely removed
		Always/regularly	Sometimes		
	(1)	(2)	(3)	(4)	(5)
HWWS (β_1)	-0.032	-0.059	-0.057	-0.052*	0.043
	(0.040)	(0.041)	(0.044)	(0.029)	(0.036)
TSSM (β_2)	-0.120***	-0.098**	-0.062	-0.055*	0.117***
	(0.039)	(0.041)	(0.043)	(0.029)	(0.034)
HWWS + TSSM (β_3)	-0.074*	-0.085**	-0.042	-0.066**	0.084**
	(0.038)	(0.039)	(0.046)	(0.032)	(0.033)
p-values $\beta_1 = \beta_2$	0.020	0.342	0.884	0.925	0.036
for F-test $\beta_1 = \beta_3$	0.264	0.506	0.735	0.649	0.332
$\beta_2 = \beta_3$	0.202	0.746	0.632	0.719	0.332
Observations	3,616	3,610	3,610	3,592	3,619
R-squared	0.229	0.135	0.108	0.106	0.186
Control Mean	0.231	0.299	0.513	0.839	0.716

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects.

Table 10: Caregiver handwashing knowledge and availability of handwashing material

VARIABLES	Knows when to wash hands (index) (1)	Knows best method to wash hands (2)	Washed hands with soap in last 24 hours (3)	HH soap expenditure in last month (TZS) (4)	HH has a handwashing device (5)	HH has a handwashing station with soap (6)	HH has a fixed handwashing device (7)	Handwashing device within 6m of toilet (8)
HWWS (β1)	0.045**	0.021	0.003	144.695	-0.015	-0.013	0.017**	0.012
	(0.018)	(0.019)	(0.021)	(116.570)	(0.039)	(0.017)	(0.007)	(0.017)
TSSM (β2)	0.007	0.053**	0.018	269.014**	0.025	-0.030*	-0.001	0.024
	(0.017)	(0.022)	(0.020)	(120.058)	(0.040)	(0.018)	(0.006)	(0.018)
HWWS + TSSM (β3)	0.040**	0.057***	0.020	212.111*	-0.004	-0.018	0.028***	0.063***
	(0.017)	(0.021)	(0.021)	(122.888)	(0.039)	(0.016)	(0.008)	(0.020)
p-values β1 = β2	0.031	0.089	0.462	0.294	0.285	0.317	0.028	0.509
for F-test β1 = β3	0.776	0.054	0.440	0.565	0.776	0.734	0.259	0.013
β2 = β3	0.049	0.844	0.953	0.650	0.439	0.440	0.001	0.075
Observations	3,614	3,614	3,599	3,453	3,419	3,295	3,419	2,800
R-squared	0.072	0.051	0.059	0.050	0.352	0.149	0.040	0.105
Control Mean	0.302	0.822	0.823	4131.177	0.484	0.080	0.012	0.037

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects.

Table 11: Caregiver handwashing practices and cleanliness

| VARIABLES | HWWS after fecal contact | | HWWS before handling food | | Any exposure event is accompanied by | |
| | Observed | Reported | Observed | Reported | water only | soap and water |
	(1)	(2)	(3)	(4)	(5)	(6)
HWWS (β1)	-0.028	0.042	0.016*	0.077***	0.011	0.006
	(0.030)	(0.033)	(0.010)	(0.020)	(0.018)	(0.009)
TSSM (β2)	-0.056*	0.033	0.009	-0.023	-0.005	-0.010
	(0.031)	(0.029)	(0.008)	(0.017)	(0.020)	(0.008)
HWWS + TSSM (β3)	0.003	0.025	0.016*	0.022	0.017	0.011
	(0.029)	(0.033)	(0.008)	(0.019)	(0.020)	(0.008)
p-values β1 = β2	0.297	0.787	0.490	0.000	0.369	0.062
for F-test β1 = β3	0.224	0.640	0.976	0.010	0.741	0.614
β2 = β3	0.031	0.813	0.456	0.014	0.275	0.014
Observations	961	3,307	2,238	3,307	4,126	4,126
R-squared	0.074	0.106	0.034	0.119	0.035	0.027
Control Mean	0.127	0.474	0.013	0.150	0.273	0.038

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects.

Table 12: General hygiene

VARIABLES	Caregiver hand cleanliness index (1)	Child cleanliness index (2)	Feces observed in house (3)	Smell of feces in dwelling (4)	Garbage is visible in dwelling (5)	Food is covered (6)	HH has ever cleaned latrine (7)	Flies visible around latrine (8)	Feces are visible outside the latrine (9)
HWWS (β1)	0.403**	0.077**	-0.000	-0.001	-0.037	0.111***	0.077***	-0.031	-0.007
	(0.187)	(0.030)	(0.017)	(0.020)	(0.031)	(0.026)	(0.022)	(0.036)	(0.022)
TSSM (β2)	0.037	0.025	0.007	0.002	0.035	0.089***	0.015	-0.018	0.014
	(0.181)	(0.028)	(0.016)	(0.019)	(0.032)	(0.027)	(0.023)	(0.035)	(0.021)
HWWS + TSSM (β3)	0.455**	0.069**	0.008	0.011	-0.037	0.068**	0.072***	0.019	-0.043**
	(0.179)	(0.030)	(0.018)	(0.021)	(0.031)	(0.029)	(0.024)	(0.037)	(0.021)
p-values β1 = β2	0.020	0.061	0.602	0.826	0.013	0.398	0.005	0.673	0.341
for F-test β1 = β3	0.740	0.790	0.617	0.504	0.993	0.124	0.831	0.162	0.115
β2 = β3	0.008	0.118	0.951	0.618	0.015	0.473	0.017	0.280	0.009
Observations	3,606	5,585	3,583	3,601	3,524	2,780	2,921	2,974	2,896
R-squared	0.091	0.210	0.063	0.061	0.045	0.049	0.094	0.087	0.055
Control Mean	6.765	0.566	0.089	0.124	0.383	0.283	0.800	0.537	0.222

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects. Column (2) also includes child age (month) and gender dummies.

Table 13: Under 5 child health

VARIABLES	Diarrhea in past 7 days (1)	Diarrhea in past 14 days (Listing Data) (2)	Hemoglobin level (g/L) (3)	Anemic (Hb < 110 g/L) (4)	Anthropometric z-scores			
					weight-for-age (5)	height-for-age (6)	weight-for-height (7)	head circumference (8)
HWWS (β1)	-0.004	-0.013	0.089	0.007	0.015	0.030	-0.006	0.229
	(0.012)	(0.011)	(0.754)	(0.023)	(0.043)	(0.057)	(0.048)	(0.141)
TSSM (β2)	-0.001	-0.010	-0.772	0.024	-0.044	-0.006	-0.061	0.092
	(0.012)	(0.012)	(0.713)	(0.022)	(0.035)	(0.059)	(0.048)	(0.129)
HWWS + TSSM (β3)	-0.011	-0.021*	-1.652**	0.060**	-0.075**	-0.008	-0.097**	0.227
	(0.013)	(0.013)	(0.772)	(0.023)	(0.038)	(0.057)	(0.045)	(0.154)
p-values β1 = β2	0.744	0.813	0.208	0.398	0.137	0.515	0.247	0.263
for F-test β1 = β3	0.594	0.428	0.023	0.020	0.039	0.476	0.050	0.991
β2 = β3	0.399	0.360	0.214	0.091	0.396	0.961	0.445	0.351
Observations	5,768	34,045	5,203	5,203	5,203	5,208	5,202	5,208
R-squared	0.053	0.051	0.194	0.159	0.062	0.084	0.064	0.223
Control Mean	0.086	0.168	111.441	0.414	-1.033	-1.946	0.055	-0.511

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects, child gender and age (month) dummies. Column (1) reports on the symptom-based diarrhea measure capured in the household survey. Column (2) reports on direct diarrhea reports captured in the listing survey.

VARIABLES	(1) Diarrhea (14 days) AGE 0-2	(2) Diarrhea (14 days) AGE 3-4	(3) Hemoglobin level (g/L) AGE 0-2	(4) Hemoglobin level (g/L) AGE 3-4	(5) Weight-for-age AGE 0-2	(6) Weight-for-age AGE 3-4	(7) Height-for-age AGE 0-2	(8) Height-for-age AGE 3-4	(9) Head circumference-for-age AGE 0-2	(10) Head circumference-for-age AGE 3-4
HWWS (β1)	-0.012	-0.014	-0.306	0.951	0.026	-0.010	0.039	0.001	0.281*	0.151
	(0.354)	(0.161)	(0.689)	(0.349)	(0.641)	(0.849)	(0.598)	(0.990)	(0.053)	(0.317)
TSSM (β2)	-0.008	-0.015	-0.199	-1.674	-0.057	-0.043	0.009	-0.059	0.106	0.060
	(0.548)	(0.173)	(0.787)	(0.103)	(0.219)	(0.330)	(0.906)	(0.424)	(0.408)	(0.683)
HWWS + TSSM (β3)	-0.026*	-0.013	-2.359***	-0.469	-0.094**	-0.060	0.013	-0.056	0.288*	0.120
	(0.078)	(0.239)	(0.005)	(0.646)	(0.047)	(0.245)	(0.854)	(0.430)	(0.073)	(0.446)
Observations	21,337	12,707	3,239	1,964	3,236	1,967	3,241	1,967	3,241	1,967
R-squared	0.043	0.021	0.164	0.149	0.072	0.080	0.109	0.081	0.225	0.229
Control Mean:	0.204	0.106	109.038	115.594	-1.019	-1.058	-2.004	-1.847	-0.467	-0.586

Notes: Robust standard errors in parentheses clustered at the ward level.*** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated using a linear probability model including block fixed effects, child gender and age (month) dummies. Columns (1) and (2) use listing data while columns (3) - (10) use household survey data.

Table 15: Robustness checks

	Variables	Specification	HWWS		TSSM		HWWS + TSSM		Observations	R-squared	Control Mean
(i)	Diarrhea (past 14 days)	Original	-0.013	(0.250)	-0.010	(0.373)	-0.021*	(0.096)	34,045	0.051	0.168
		Interviewer dummies	-0.016	(0.115)	-0.008	(0.421)	-0.023**	(0.024)	34,045	0.093	0.168
		Month dummies	-0.014	(0.210)	-0.007	(0.498)	-0.019*	(0.086)	34,045	0.094	0.168
(ii)	Height-for-age	Original	0.030	(0.057)	-0.006	(0.059)	-0.008	(0.057)	5,208	0.084	-1.946
		Removing outliers	0.031	(0.054)	0.002	(0.053)	-0.004	(0.052)	5,147	0.085	-1.953
		Interviewer dummies	0.035	(0.054)	-0.008	(0.054)	-0.020	(0.055)	5,208	0.097	-1.946
		Month dummies	0.034	(0.054)	-0.008	(0.053)	-0.021	(0.052)	5,208	0.100	-1.946
(iii)	Weight-for-age	Original	0.015	(0.043)	-0.044	(0.035)	-0.075**	(0.038)	5,203	0.062	-1.033
		Removing outliers	0.024	(0.040)	-0.042	(0.033)	-0.073**	(0.036)	5,170	0.063	-1.017
		Interviewer dummies	0.031	(0.042)	-0.050	(0.033)	-0.079**	(0.037)	5,203	0.068	-1.033
		Month dummies	0.040	(0.040)	-0.047	(0.031)	-0.081**	(0.035)	5,203	0.071	-1.033
(iv)	Head circumference-for-age	Original	0.229	(0.141)	0.092	(0.129)	0.227	(0.154)	5,208	0.223	-0.511
		Removing outliers	0.229*	(0.123)	0.086	(0.107)	0.181	(0.131)	5,134	0.187	-0.543
		Interviewer dummies	0.100	(0.095)	0.075	(0.095)	0.222**	(0.107)	5,208	0.331	-0.511
		Month dummies	0.067	(0.100)	0.041	(0.093)	0.149	(0.106)	5,208	0.343	-0.511
(v)	Hemoglobin level (g/L)	Original	0.089	(0.754)	-0.772	(0.713)	-1.652**	(0.772)	5,203	0.194	111.441
		Removing outliers	0.199	(0.714)	-0.453	(0.690)	-1.101	(0.733)	5,162	0.196	111.519
		Interviewer dummies	-0.181	(0.796)	-0.916	(0.724)	-1.608**	(0.760)	5,203	0.200	111.441
		Month dummies	-0.656	(0.803)	-1.340*	(0.723)	-2.217***	(0.785)	5,203	0.206	111.441
(vi)	Abrasions / bruising in past 7 days		0.001	(0.008)	0.001	(0.008)	-0.015*	(0.008)	5,764	0.042	0.044

Notes: Robust standard errors in parentheses clustered at the ward level.*** p<0.01, ** p<0.05, * p<0.1. Coefficients estimated using a linear probability model including block fixed effects, child gender and age (month) dummies. Row (i) uses listing data; rows (ii) - (vi) use household survey data. For rows (i) - (v) outcomes are subjected to robustness tests. "Original" provides the result presented in the main tables. "Removing outliers" excludes observations that are more than three standard deviations away from the mean. "Interviewer dummies" includes control dummies for all survey interviewers. "Month dummies" specification includes both interviewer and month-of-interview dummies. Row (vi) provides evidence on the falsification test.

Table 16: Mechanism testing: differential mortality rates

VARIABLES	(1) Any children under 5	(2) Number of children under 5	(3) Child age	(4) Age 0 - 1	(5) Age 1 - 2	(6) Age 2 - 3	(7) Age 3 - 4	(8) Age 4 - 5
HWWS	-0.014	-0.022	0.009	-0.003	0.002	-0.001	-0.000	0.002
	(0.014)	(0.028)	(0.018)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
TSSM	0.011	0.020	-0.015	-0.004	0.007	0.001	0.006	-0.010*
	(0.014)	(0.030)	(0.017)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
HWWS + TSSM	-0.013	-0.034	0.010	-0.005	0.002	-0.001	0.009	-0.004
	(0.014)	(0.029)	(0.018)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Observations	50,885	50,885	34,092	34,092	34,092	34,092	34,092	34,092
R-squared	0.068	0.100	0.003	0.002	0.004	0.004	0.002	0.006
Control Mean:	0.444	0.666	1.928	0.205	0.209	0.214	0.196	0.176

Notes: Robust standard errors in parentheses clustered at the ward level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects. All results based on listing data.

Table 17: Mechanism testing: latrine seepage and quality

VARIABLES	(1) Slab is cleanable	(2) Floor is cleanable	(3) Drophole is only hole	(4) Solid slab	(5) Flooded by water	(6) Strong odor	(7) Feces visible outside pit	(8) Flies always present	(9) Has a squat hole cover in place
HWWS	0.026	0.040**	0.026	0.018	0.011	-0.023	-0.008	-0.031	0.005
	(0.020)	(0.019)	(0.033)	(0.027)	(0.011)	(0.019)	(0.022)	(0.036)	(0.014)
TSSM	0.032*	0.037**	-0.025	-0.001	0.009	-0.002	0.013	-0.018	0.052***
	(0.019)	(0.018)	(0.029)	(0.024)	(0.010)	(0.018)	(0.021)	(0.035)	(0.015)
HWWS + TSSM	0.014	0.022	-0.023	0.014	0.005	-0.012	-0.044**	0.019	0.039***
	(0.020)	(0.021)	(0.030)	(0.024)	(0.007)	(0.020)	(0.021)	(0.037)	(0.014)
Observations	2,896	2,896	2,896	2,883	2,887	2,892	2,892	2,974	2,791
R-squared	0.196	0.224	0.127	0.269	0.053	0.035	0.055	0.087	0.082
Control Mean	0.853	0.863	0.580	0.758	0.013	0.176	0.222	0.537	0.038

Notes: Robust standard errors in parentheses clustered at the ward level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients estimated using a linear probability model including block fixed effects. Results are conditional on household having a latrine.

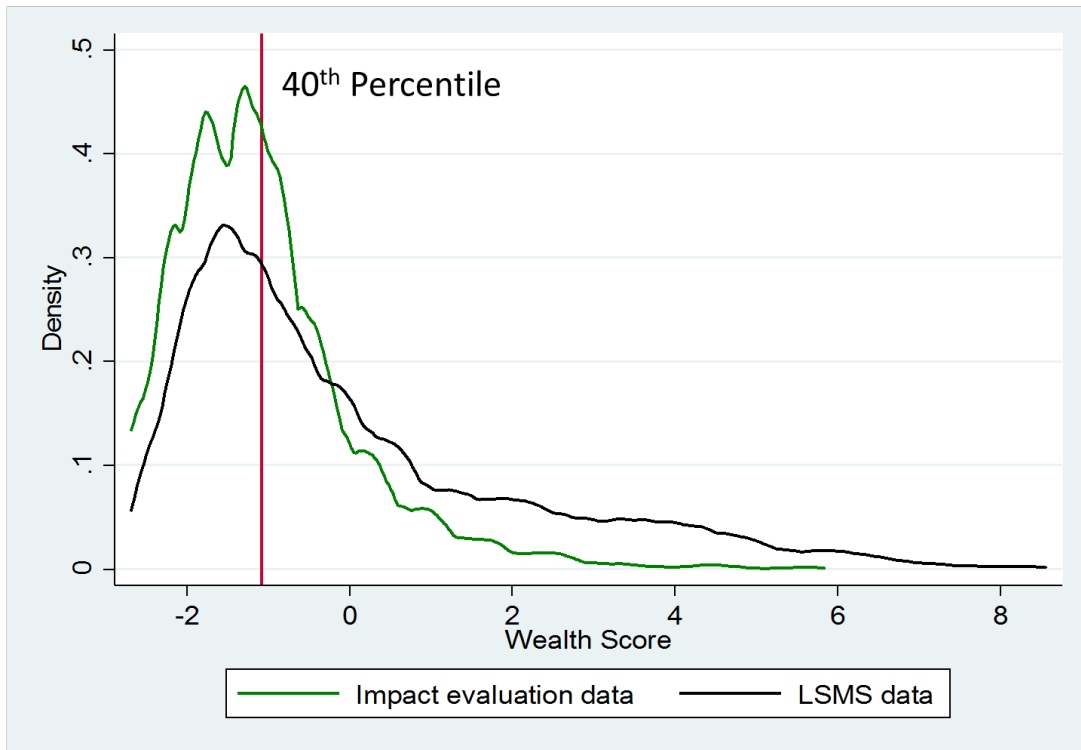Figure 1: Wealth distributions for LSMS and IE data
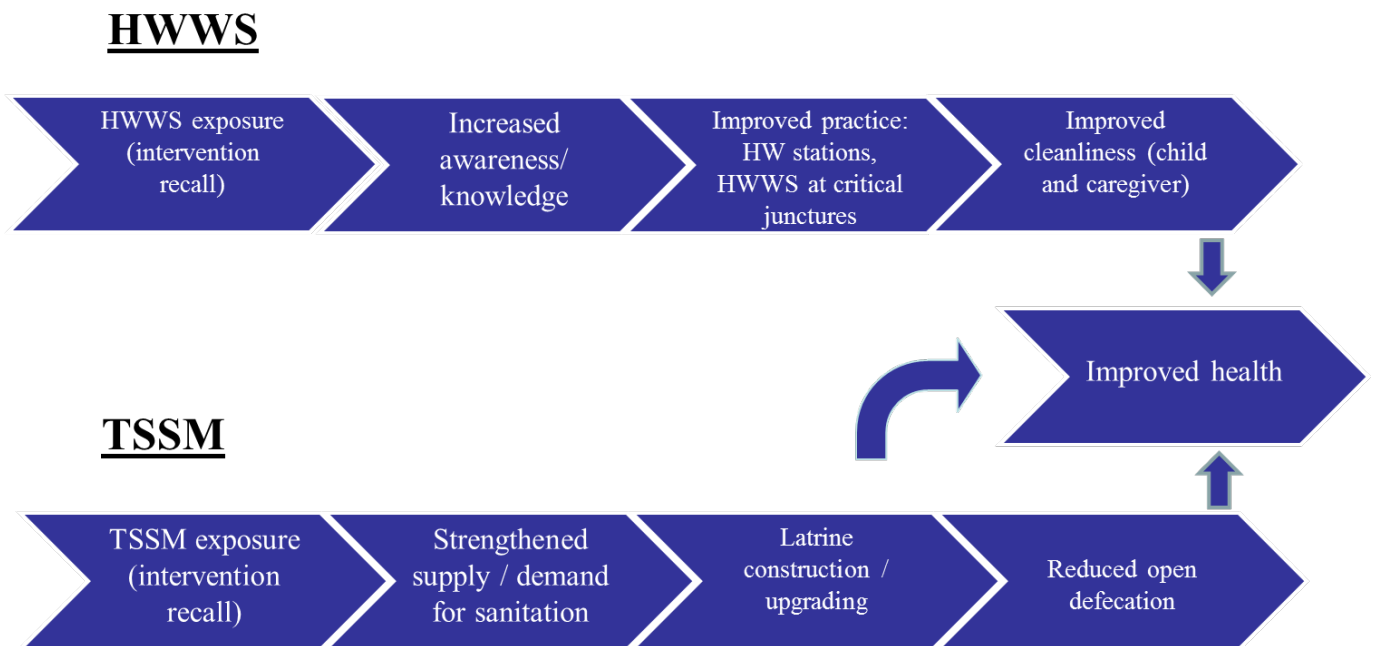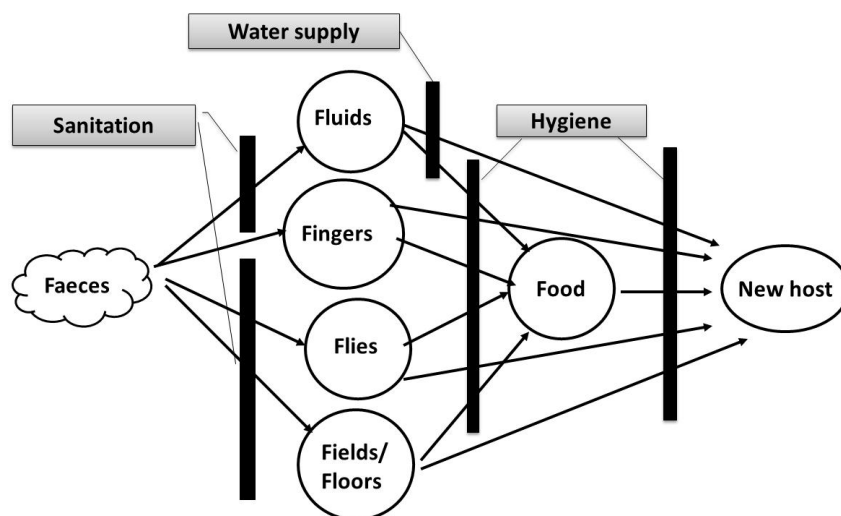

Figure 2: Theory of change

Figure 3: The "F-diagram": fecal-oral transmission pathways and interventions to break them

Source: Perez (2012). Adapted from Mara (2010) and originally from Wagner (1958).

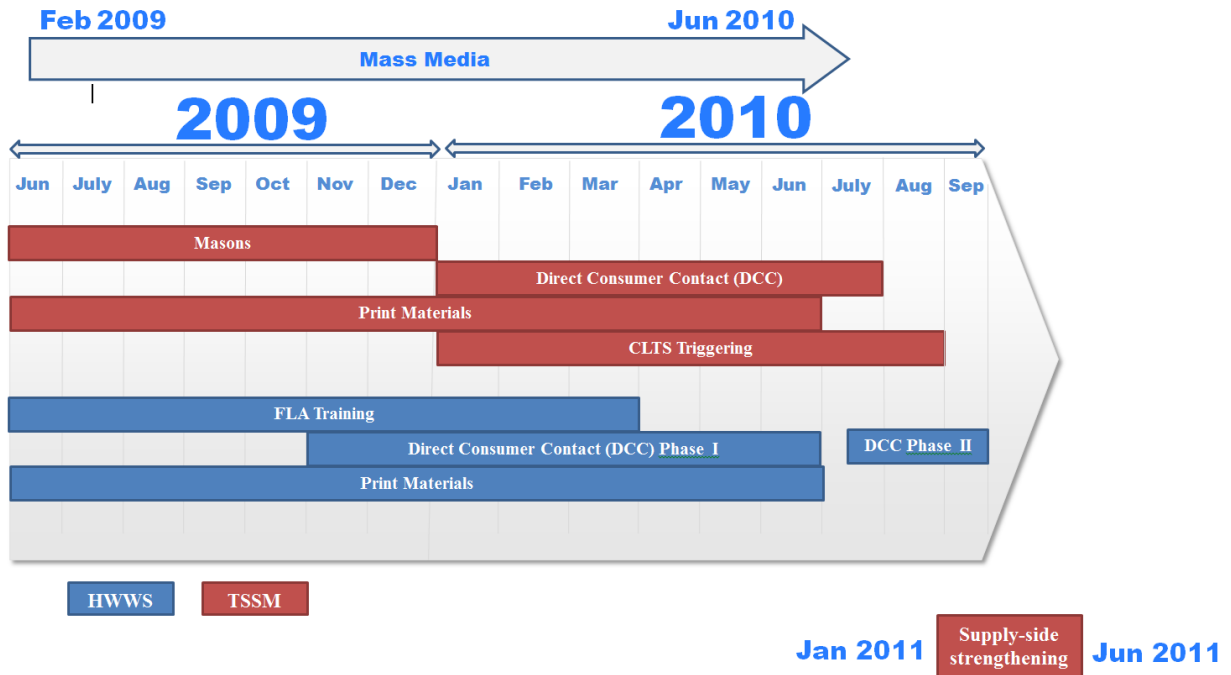Figure 4: Marketing materials

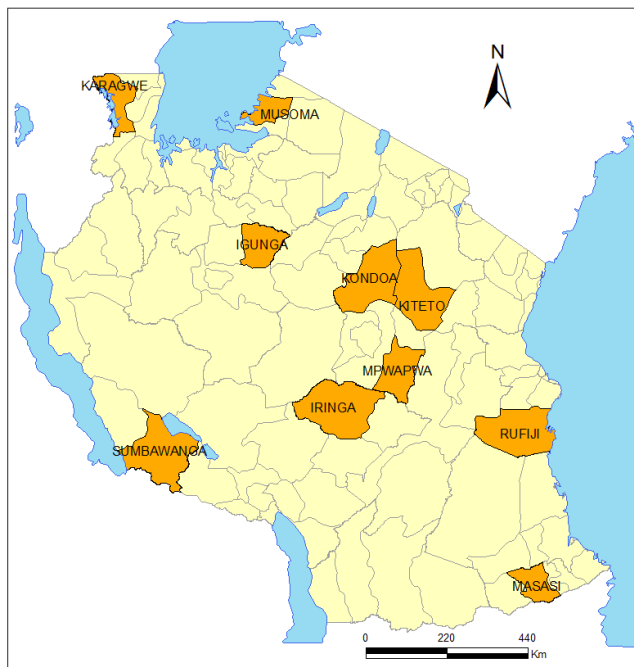Figure 5: Intervention timeline

Figure 6: Map of selected districts

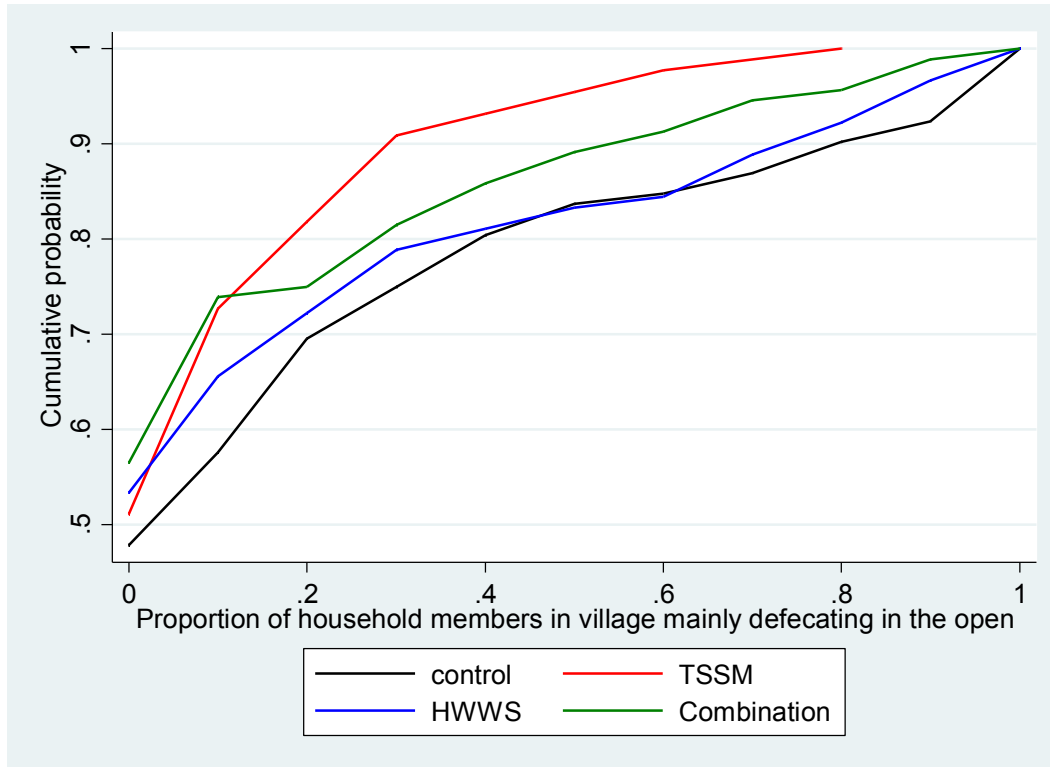Figure 7: CDF – village-level open defecation (households practice regular open defecation)

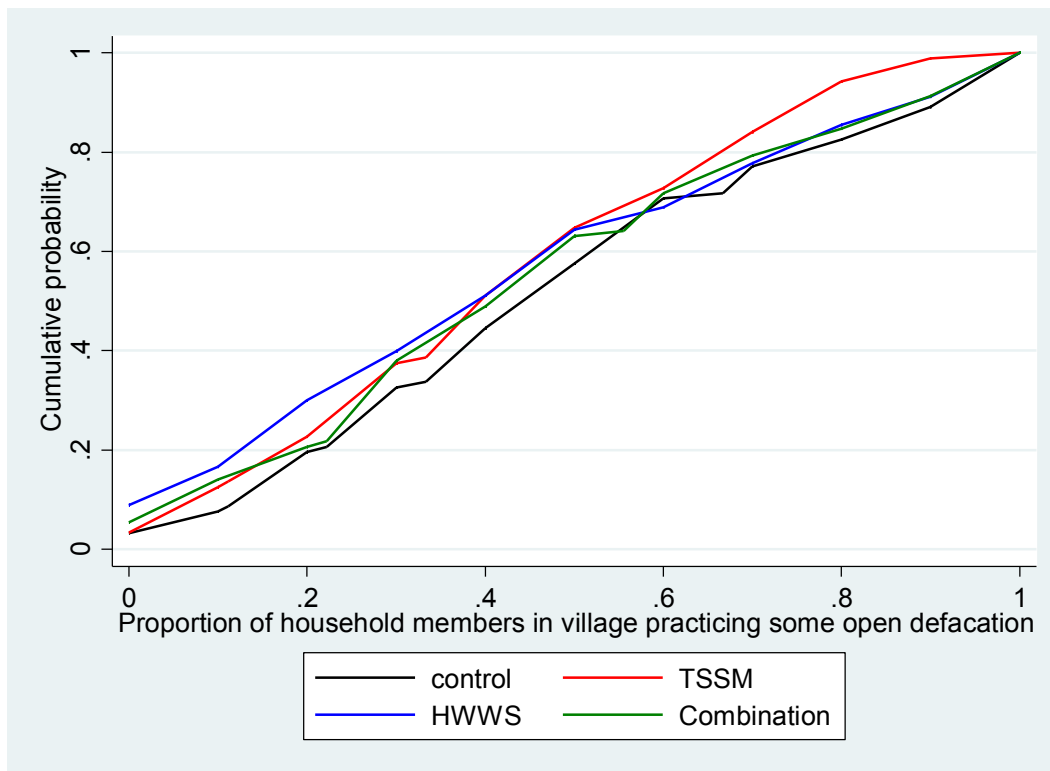Figure 8: CDF – village-level open defecation (households practice at least some open defecation)

Figure 9: Anemia significance by cutoff level