TI 2020-017/V
Tinbergen Institute Discussion Paper

The complementarity of community-based water and sanitation interventions: evidence from Mozambique

Melinda Vigh[1]
Chris Elbers[1]
Jan Willem Gunning[1]

[1] Vrije Universiteit Amsterdam and Tinbergen Institute Amsterdam, The Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

The complementarity of community-based water and sanitation interventions: evidence from Mozambique*

Melinda Vigh[†], Chris Elbers and Jan Willem Gunning

Vrije Universiteit Amsterdam and Tinbergen Institute

Amsterdam, The Netherlands

April 2, 2020

*Data collection was financed by the Evaluation Department (IOB) of the Netherlands Ministry of Foreign Affairs in the context of an evaluation study described in detail in UNICEF and IOB (2011). We are grateful to the UNICEF staff in Maputo, Rita Tesselaar of the Netherlands Ministry of Foreign Affairs and Menno Pradhan for very helpful comments. We thank the participants at the internal seminar at the Vrije Universiteit Amsterdam in 2018, the DeNeB PhD seminar in 2018 in Berlin, the Nordic Conference in Development Economics 2018 in Helsinki for valuable comments and helpful insights on earlier versions of this paper.

[†]Corresponding author, E-mail: m.vigh@vu.nl

Abstract

We use data on a large-scale water and sanitation program in rural Mozambique, implemented between 2008 and 2013, to investigate the complementarities between a behavior-change-based community-led total sanitation intervention and a community water supply intervention. Our findings indicate that the sanitation intervention increased the adoption of handwashing with soap or ash by 11 percentage points, latrine ownership and use by 8 percentage points, and the use of improved water points by 15 percentage points (conditional on access). Combining the water supply and sanitation interventions increased the treatment effects on all three outcomes. However, we find that the effect on toilet ownership was in large part driven by the selective intervention allocation of the implementing NGOs. These effects are measured up to 4 years after the intervention.

Keywords: impact evaluation, sanitation, WASH, CLTS, correlated random effects
JEL classification: D04, I15, O12

1 Introduction

Efforts to provide access to safe water sources and sanitation have been ongoing for several decades worldwide. While gains have been made globally, 32% of the rural population in low income countries still practiced open defecation as recently as in 2015, and 44% of them had no access to improved water sources.[1] These figures suggest that despite successes in the last decades, there is still a lot to do to achieve access to safe water and sanitation for all by 2030, number 6 of the Sustainable Development Goals of the United Nations.

Approaches to eradicating open defecation mainly differ between providing subsidized latrines (Pattanayak et al., 2009, Patil et al., 2014, Guiteras et al., 2015) or providing information/education and fostering community engagement (Whaley and Webster, 2011, Crocker et al., 2017, Cameron et al., 2019, Alzúa et al., 2020). We focus on the latter behavior-change-based sanitation interventions, which mainly follow the Community-Led Total Sanitation (CLTS) approach of Kar and Chambers (2008). CLTS uses a confrontational approach to raise awareness to the importance of safe sanitation practices for health, and relies on a community effort to build latrines without using subsidies. The meta-analysis of Garn et al. (2017) reports that CLTS interventions increased latrine coverage by 12% on average (95% CI: -2%, 27%). Reporting on CLTS programs in India, Indonesia, Mali and Tanzania, Gertler et al. (2015) finds that the effectiveness of CLTS in reducing open defecation varies with the intensity of the intervention: the more intensive the CLTS promotion, the larger its effect was (3–33 percentage point decrease in open defecation).

This paper contributes to this literature on the effectiveness of CLTS interventions by investigating possible effectiveness gains from implementing CLTS together with a water supply intervention in the framework of a large-scale water and sanitation program in rural Mozambique. We estimate medium- to long-term treatment effects with follow-up surveys conducted up to 4 years after the CLTS intervention. A further contribution of the paper is that we employ a novel method proposed by Vigh and Elbers (2017) to estimate population average treatment effects, which is robust to selective

[1] Figures obtained from the World Bank Database (https://data.worldbank.org/indicator).

intervention allocation motivated by selection on gains.

The program in Mozambique followed the CLTS methodology for household toilet facility construction with additional focus on handwashing with soap or ash after toilet use. It was implemented in underdeveloped rural communities that often had no access to safe water points and had a high prevalence of open defecation. Combining the water supply and sanitation interventions was envisioned to break all the main transmission pathways of fecal contaminants, thereby reducing the disease burden on the population (see for example Waddington and Snilstveit, 2009). Hence, the two interventions were expected to be complementary in terms of improving health outcomes.

However, complementarity of the water supply and sanitation interventions could also manifest itself in higher adoption of desired hygienic behaviors when the two interventions are combined. For example, providing water may encourage latrine use as it makes it easier to clean them. Building on this logic, Duflo et al. (2015) discusses a piped water supply and sanitation program in rural India, where installing a pour-flush toilet in every household in the village was a precondition for activating the village piped-water system. The logic behind the intervention design was that households will upgrade their sanitation in order to get running tap water in their homes. In this paper, we investigate the behavioral complementarity between water supply and sanitation programs from the opposite angle: providing a safe community water source may increase households' willingness to invest in sanitation. We focus exclusively on this second source of complementarity, and do not discuss health outcomes.

This form of complementarity may arise, for example, due to a more intensive and beneficial working relationship between the intervention implementers and the communities. Specifically, the program facilitators helped the communities apply for a new community water point and trained the water committee on sustainable management of the water point after the water point was installed. The sanitation intervention was usually implemented after the water supply intervention. Communities were confronted with their sanitary circumstances and were triggered to build simple toilet facilities (mostly pit latrines) using locally available material.

We find that implementing the sanitation intervention together with the WP intervention increases the effectiveness of the former on the ownership and use of toilet facilities and handwashing with soap or ash compared to implementing the sanitation component alone. At the same time, the sanitation intervention increases the effectiveness of the WP intervention on the use of improved WPs.

The implementation of the program does not lend itself to an impact assessment following the randomized controlled trial methodology. The NGOs carrying out the CLTS intervention were also contracted to carry out an initial assessment of the water and sanitation situation in the communities. The location of the new improved WPs was later decided by the local government based on need. However, the NGOs selected the communities for the CLTS intervention. In addition to the NGO's internal motivation to succeed, they were also financially incentivized to implement the interventions at locations that had a higher probability of becoming open defecation free. Therefore, it was in their interest to select those locations for CLTS where they expected the highest treatment effect among the available pool of communities.

Targeting the interventions based on need or expected impact suggests that the treatment effects among the potential beneficiaries were expected to vary. As a result of targeting or selective intervention placement, the treatment effect on the actual beneficiaries is likely to overestimate the average treatment effect among all potential beneficiaries. Heckman et al. (2006) called this phenomena "essential heterogeneity" or "selection on gains".

We rely on this incentive structure to differentiate between the treatment effect on the beneficiaries and the population average of the treatment effect. We expect that the intervention communities were selected in the order of the size of the expected treatment effect. Using a novel method proposed by Vigh and Elbers (2017) for such settings, we are able to filter out the effect of this selectivity and estimate the population average of the treatment effect.

We find that the selective placement of the CLTS intervention increased its effect on latrine ownership. The program increased latrine ownership by 8 percentage points (pp) among the program beneficiaries in addition to a 9pp general increase in the same period. The treatment effect is driven by communities that received both the WP and CLTS interventions, where the treatment effect on the treated is 12pp compared to an 7pp but insignificant increase in latrine ownership for the CLTS intervention alone. However, after filtering out the effects of selective intervention placement, the average treatment effect shows that the CLTS intervention would not be effective if implemented at randomly selected communities, irrespective of being implemented together with the WP intervention. The results on the use of toilet facilities and open defecation follow closely the results for latrine ownership.

The treatment effect of CLTS on handwashing with soap or ash after defecation was 12pp on average. We find no evidence of selectivity for this outcome, suggesting that upscaling the progam would be similarly effective for handwashing with soap or ash after defecation. The treatment effect on this outcome is also larger in communities with both the WP and CLTS interventions (16pp) compared to the CLTS intervention alone (10pp).

Uniquely, the program design also allows us the assess the effect of CLTS on the use of improved WPs. We find that conditional access to an improved WP, the CLTS intervention increased the use of improved WPs by 15 pp on average. Again, this effect is mainly attributable to communities that also received a WP intervention, where the combined effect of the interventions was 36 pp conditional on the availability of improved WPs. These effects are robust to the effects of selective intervention placement. However, when the distance to the improved WP is more than 2 km, the effect of CLTS disappear. This is in line with our finding that about three-quarter of the households fetch water from a source within 1 km from their dwelling.

Hence, we find that the synergy effects of the WP and sanitation interventions run both ways. The increased effectiveness of CLTS on latrine ownership and handwashing is not correlated with the distance to improved WPs.

These results should be interpreted as medium- to long-term effects given that the follow-up surveys were conducted at least 1-2 years after the intervention. In general, we find little reversion in latrine ownership (correlation coefficient of 0.8). This suggests real behavior change given the fact that the traditional pit latrines used by households have to be emptied or moved almost every year. We also observe that the program effects mainly come from new adopters in the case of latrine ownership. In the case of handwashing with soap after defecation, the program's effect can be attributed both to new adopters and a higher retention rate of earlier adopters.

In the next sections, we first introduce the program, the data collection and the outcome variables in section 2, then discuss the effects of the sanitation intervention with and without the WP intervention on latrine ownership and handwashing in section 3. In this section we also discuss our approach to filtering out the effects of selective intervention placement, and report on the average treatment effect on the treated and in the general population. In section 4, we turn to discussing the program's effects on the use of improved WPs, and then in secton 5 we discuss potential explanations for the observed complementarity between the water and sanitation interventions. Finally, section 6 concludes.

2 Data description

2.1 The program

The One Million Initiative (OMI) was implemented between 2006-2013 in 18 districts of Manica, Sofala and Tete provinces of central Mozambique. The program was a cooperation between the Governments of Mozambique and the Netherlands and coordinated by UNICEF Mozambique. It aimed to support the Government of Mozambique in achieving the Millennium Development Goals on the access to safe drinking water and basic sanitation by reaching one million people in poor rural areas (40% of the population in the program area) and providing them with means to use safe and sustainable drinking water and hygienic sanitation facilities.

At the time of the program, in 2010, only 35 percent of the rural population in Mozambique used improved water sources (such as a borehole or a protected well),[2] and 9.5 million people or almost 40 percent of the population were practicing open defecation.[3] The situation was even more extreme in the OMI program districts, where data from the Demographic and Health Survey from 2009 showed that only 27% of the rural population had access to an improved water point (WP), and 44% were using a toilet facility.

[2] Source: http://mdgs.un.org/unsd/mdg/data.aspx

[3] Source: http://www.who.int/water_sanitation_health/monitoring/jmp2012/fast_facts/en/

The program carried out community water point (WP) interventions by creating new boreholes and training water committees on maintenance (2008-2013), and com-munity-based sanitation and hygiene education through the Community Approaches to Total Sanitation, which combined the Community-Led Total Sanitation (CLTS) approach with an award scheme for Open Defecation Free (ODF) communities (2008-2013). Community participation was an essential component of the interventions.

In every district, different so-called "Community Participation and Education" activist NGOs were contracted as the focal point between the communities and the program. These NGOs were responsible for communicating and recruiting communities for the program, providing information over the needs of communities to local authorities, as well as, implementing the hygiene, sanitation (CLTS) and water committee trainings in the communities. In the program, the communities had to apply for an improved WP with the assistance of the NGOs, indicating the preferred location of the WP and community members responsible for the future management of the WP (water committee). Based on information received from the NGOs, the location of the water point interventions (WPI) was decided by local authorities. After the installation of the WP, the NGOs trained the water committee members on the sustainable management of the water point, including collecting user fees to cover later reparation costs.

The NGOs were free to choose the communities for implementing the CLTS trainings. However, the communities had to agree to the trainings and the NGOs had to reach their annual target of triggered communities. The implementation of the program is summarized in Figure B2 in Appendix B.

The aim of the Community-Led Total Sanitation methodology is to eradicate open defecation using a confrontational approach which included the mapping of locations used as "toilets", visiting these locations during the "walk of shame", and the triggering moment of demonstrating the fecal contamination of food.[4] The intervention also put an emphasis on proper handwashing after defecation. After the awareness was raised,

[4] See Kar and Chambers (2008) for more information about the implementation of Community Led Total Sanitation.

the solutions in the form of latrine construction based on local materials was discussed, and the community pledged to build latrines. No subsidies were provided to promote latrine construction. After the training, the NGOs could follow-up with the communities until they became ODF, meaning that all households owned and used a latrine. These were almost always traditional latrines in the program area. The program encouraged households to build safe latrines using locally available materials.

The triggered communities could apply to certify their ODF status in the annual ODF evaluation campaign, where it was verified that there are no fecal contaminants lying around in the community, and indeed all households use a latrine. The NGOs were financially rewarded for the number of ODF communities in their district, therefore they had an incentive to introduce CLTS and follow-up on the trainings at communities that had a high likelihood of success. In addition, leaders of the ODF certified communities also received a non-financial reward (for example, a bicycle for the community leader).[5] Based on Kar and Chambers (2008) the success of CLTS is related to factors including health problems, leadership, size of community (smaller is easier) and geographical factors (remoteness).

As the above description suggests, the program interventions were purposefully not randomized across communities in the program area. Emphasis was put on reaching communities that do not have or have only limited access to safe water sources and sanitary facilities, while the implementing NGOs were incentivised to implement CLTS in communities where they expected the highest impacts of the program (to become ODF). Overall, about 45% of the communities with a CLTS intervention were awarded ODF status by the end of 2013.

2.2 Data collection

For the evaluation of the program, survey data were collected in three rounds: August-October 2008 (baseline), August-October 2010 (midline) and July-August 2013 (endline). In each round, data were collected from 1600 households in 80 communities.[6] In each sampled community, three types of surveys were carried out: a

[5] Prizes for communities were abolished in 2011.

[6] Communities usually represent about 100 households. Villages can have multiple communities.

household survey, a community survey administered to local community leaders and a water point survey administered at all WPs used by the community. In addition, for a sub-sample in each community, water samples were taken at the water point and the point of use for microbiological analysis.

The surveys were designed to measure the effects of the program interventions on the use of improved WPs, ownership and use of (improved) latrines, practice of open defecation, hand washing practices and prevalence of water related diseases. The household surveys also collected data on hygiene awareness and attitude, socio-economic characteristics of the households, and the age, gender and education of all household members. Except for observations on sanitary facilities, most variables are self-reported. The community survey collected information on general community characteristics, water, sanitation and hygiene interventions and practices in the community.

We focus on the program's impact on three outcome variables: latrine ownership,[7] handwashing with soap or ash after defecation and the use of improved WPs. All three outcome variables are binary and defined at the household level.[8]

The survey communities were sampled based on their likelihood of receiving an intervention soon after the baseline survey. Communities were classified as having a high likelihood of intervention if they had already been visited by a program NGO.[9] All other communities were classified as having a low likelihood of intervention. In total, 302 high likelihood communities were listed based on information from the NGOs, and 2200 low likelihood communities were listed based on census data (INE, 1997). From these lists of eligible communities, 40 were selected for both the Intended Target (IT) and the Intended Control (IC) group using a systematic random selection procedure.[10]

[7] Results on open defecation are also discussed. However, open defecation highly correlates with latrine ownership. We use latrine ownership as the primary sanitation outcome because it can be verified by the interviewers and it is easier to report on.

[8] The ultimate objective of all water, sanitation and hygiene interventions is to improve the health outcomes of the population, especially for young children. However, we do not discuss results on health outcomes in this paper.

[9] During these initial visits the NGOs assessed the situation in the communities. In some cases interventions were already planned in these communities.

[10] The randomization procedure was stratified by districts, so that the total number of selected communities

The location of the program interventions was reported to UNICEF Mozambique by the NGOs for the CLTS trainings, and by the construction companies for the newly drilled functioning boreholes (community WPs). These were matched to the surveyed communities. Communities could receive either the WP intervention, the CLTS sanitation intervention or both interventions.[11]

Table 1 summarizes the number of communities that received the interventions before the given survey period (in rows) for each of the intervention arms. 22 communities without any program interventions serve as the comparison group. In the analysis, we rely on the period of the intervention implementation. 'Wave 1' and 'Wave 2' intervention groups refer to communities that received the interventions between the baseline (2008) and midline (2010) surveys, and between the midline and endline (2013) surveys, respectively. Hence, the 2010 row of the panels in Table 1 show the Wave 1 intervention groups, while the 2013 row sums the Wave 1 and Wave 2 groups together.

The Intended Target communities were more likely to receive the interventions, particularly between the baseline (2008) and midline (2010) surveys (Wave 1). Most of the WP interventions took place during this period. The CLTS interventions were more evenly spread out over the evaluation period (Wave 1 and 2), but also here we observe that the IT communities received the interventions earlier. Overall, 95% of the IT and 50% of the IC communities received at least one of the program interventions (see Table C1 in the Appendix C).

However, when analyzing the data, we only use the realized intervention status of the communities, taking into account the selective allocation of the interventions. We first report on the CLTS and WP interventions not taking account the possible joint occurance of the two interventions, which we will refer to as 'CLTS (total)' and 'WPI (total)' or the total effects of the interventions (Table 1a).

In order to investigate the synergies between the CLTS and WP interventions, we

in each district was proportional to its population size.

[11] In addition to the CLTS and WP interventions, a hygiene promotion was implemented in every community in the sample by the NGOs. The impact of the hygiene promotion cannot be evaluated using the survey as it was done in all communities prior to the baseline.

Table 1: Cumulative distribution of the program interventions in the sample

(a) Total CLTS and WP interventions

Year	CLTS (total)	WPI (total)
2008	0	0
2010	23	35
2013	41	47

(b) Total CLTS and WP interventions

Year	CLTS (alone)	CLTSxWPI	WPI (alone)
2008	0	0	0
2010	8	15	20
2013	20	21	26

Notes: Number of communities that received the intervention prior to the survey periods 2008, 2010 and 2013.

also report the results for the three individual intervention arms: 'CLTS (alone)', 'WPI (alone)' and CLTS&WPI (Table 1b). The CLTS&WPI intervention arm contains those communities that received both the CLTS and WP interventions in the same period. In these communities the WP intervention was usually followed by the CLTS intervention.[12]

The 'CLTS (alone)' intervention arm contains those communities that received the CLTS intervention without the WP intervention or the WP intervention was implemented in a different wave. Similarly, the 'WPI (alone)' intervention arm contains those communities that received the WP intervention without the CLTS intervention or the CLTS intervention was implemented in a different period.[13] Appendix C provides more detail on the disaggregation of the intervention groups.

Turning to the data collected in the sampled communities, in every community, 20 households were randomly selected to be interviewed at the baseline (2008), resulting in a sample of 1600 households.[14] The same households were revisited for the midline

[12] Only in 2 out of 21 communities did the CLTS precede the WP intervention.

[13] In Table 1b, 9 communities appear both in the CLTS and WPI intervention arms. Because in these communities one of the interventions happened before the midline and the other after the midline survey, we are able to measure the effect of both interventions separately (unlike for the communities in the CLTS&WPI intervention arm).

[14] In the IT communities, in order to ensure that survey households are indeed living close to a potential new borehole, communities were first asked which location they would prefer for a new WP. In the baseline households were then sampled in the neighborhood of that area.

(2010) and endline (2013) surveys. If a household could not be interviewed, it was replaced with a randomly selected household. At the endline survey, the interviewers first tried to revisit households that participated in the baseline survey. If they were unsuccessful, they tried to revisit the household's replacement at the midline (if applicable), and if unsuccessful, finally replaced the household. Overall, 1161 households (72.6%) were interviewed in all three rounds, and 1524 households (95.2%) that participated in at least two survey rounds (see Appendix E).[15] In the main regression analysis, we use the latter sample but the results are robust to using the balanced sample (results available upon request).

The tables in Appendix A summarize the household characteristics. We observe some differences in the main household characteristics among the intervention arms and the comparison group. Specifically, at the baseline households in communities with the CLTS only intervention were more likely to own and use a toilet facility compared to the comparison group. Baseline handwashing practices were similar in all intervention arms, and the use of improved WP was lower in communities with a WP intervention. Communities receiving the WP intervention appear to be better off at the baseline. In addition, households in the CLTS&WPI intervention arm have more often some formal education and larger household sizes. In the regression analysis, we control for these factors.

Looking at the patterns of attrition, we find that larger households are more likely to remain in the sample (see Appendix E). In addition, we observe that households in Tete province are more likely to remain in the sample. Households in Tete are also more likely to own a latrine. However, after controlling for the province, latrine ownership does not predict the probability of attrition.

[15] In 606 households the same respondent was interviewed in all three rounds, and in 1356 households the same respondent was interviewed in at least two survey rounds. Close to 60 percent of the respondents were female, and over 90 percent of the respondents were either the head of the household or the spouse.

2.3 Selection of the intervention communities

Given that the program interventions were not randomly assigned, it is important to understand the selection of the intervention communities.[16] Using data from the Demographic and Health Survey (DHS) in Mozambique from 2009, we find that the program targeted districts where the use of improved WPs was lower than average within the program provinces (30% vs. 45% of households, Table D1 in Appendix D). Comparing the figures from the DHS survey with those from the baseline survey, we observe that the surveyed program sample over-represents communities not using an improved WP even within the surveyed program districts (30% vs. 14% of households, Tables D1b and D3 in Appendix D).

The DHS data also reveal that the use of toilet facilities (mostly traditional latrines) showed a substantial variation between and within the provinces (21-61% of households in districts, Tables D1a and D1b in Appendix D). There is no indication that the program selected districts based on latrine use. We observe a similar pattern for the presence of a handwashing facility at the households' compounds using data from the 2011 round of the DHS survey (Tables D2a and D2b in Appendix D). Hence, we can conclude that the surveyed program sample primarily targeted communities in need of an improved WP.

Regarding the actual allocation of the interventions in the surveyed communities, we investigate whether observed community characteristics predict the interventions.[17] As discussed in section 2.1, the CLTS and WP interventions follow different allocation rules. We assume that the WPI interventions were allocated independently from the CLTS interventions, however, the allocation of the CLTS interventions takes into account the location of the WP intervention. Therefore, we first discuss the predictors of the WP intervention, and then the CLTS intervention.

On the one hand, we find that communities without an improved WP at baseline were significantly more likely to receive the WP intervention. Other than availability, villages with less income inequality (in terms of the standard deviation of the wealth

[16] Appendix D provides a detailed discussion about the predictors of the program intervention placement.

[17] Given the nature or the intervention, we cannot rule out that unobserved community characterstics also affect the allocation of the interventions.

index) and medium-small size (population of 500-1000) are found to be correlated to receiving the WP intervention. Notably, we find no correlation between the placement of WP intervention and the community mean of other outcome variables (latrine ownership and handwashing practices) and a health measure (water related disease prevalence). The allocation of the WP interventions followed similar patterns in Wave 1 and Wave 2 (Figure D1 in Appendix D).

On the other hand, the WP intervention had a significant effect on the placement of the CLTS intervention within the same period (Figure D2 in Appendix D). Overall, only 5 out of 30 communities with both interventions received the CLTS intervention prior to the water point intervention.

We observe also some differences in the observable baseline community characteristics between the Wave 1 and Wave 2 CLTS interventions. The Wave 1 interventions were more often targeted at small villages (population of less than 500 people), and at communities that already had higher rates of latrine ownership at the baseline relative to other communities in their district. Given that building a latrine requires non-negligible investment, it is important to point out that the CLTS interventions are not correlated with the relative wealth level of the communities nor with wealth inequality within the communities.

2.4 Outcome variables

The main outcome variables of interest are (i) latrine ownership, (ii) handwashing with soap or ash after defecation and (iii) the use of improved water points. The choice of these outcomes is motivated by the program objectives, as these three variables represent three separate channels for reducing the burden of water related diseases (**?**). The program survey data indicate surprisingly low correlation between the three outcomes (Table A5 in Appendix A).

We decided to focus on latrine *ownership* rather than open defecation or the *use* of a latrine facility because latrine ownership was encouraged by the CLTS intervention, and sharing of latrines was not very common in the program area.[18] In addition, it is

[18] The correlation coefficient between latrine ownership and use is r=0.91. See also Table D3 in Appendix

possible to verify the existence of a latrine. However, results for latrine use and open defecation are reported in Appendix F.

The choice of the outcome on handwashing with soap or ash after defecation is also motivated by the emphasis of the CLTS intervention on this outcome. While only 18% of the households reported washing their hands with soap or ash after defecation at the baseline, almost all of them reported to wash their hands without soap or ash both after defecation and before eating. Handwashing with soap or ash before eating remained close to 20% during the evaluation period. In the following we refer to the outcome on handwashing with soap or ash after defecation simply as 'handwashing'.

Bearing in mind the selective intervention allocation based on the before intervention values of the outcome variables, we look at factors that predicted the outcome variables before the interventions. We find that both latrine ownership and handwashing were correlated with education and wealth. The availability of improved WP[19] was the strongest predictor of using improved WPs but its coefficient of 0.45 indicates that a substantial share of the households preferred to use a traditional WP even when an improved WP was available in the community. The availability of an improved WP also increased the probability of latrine ownership.

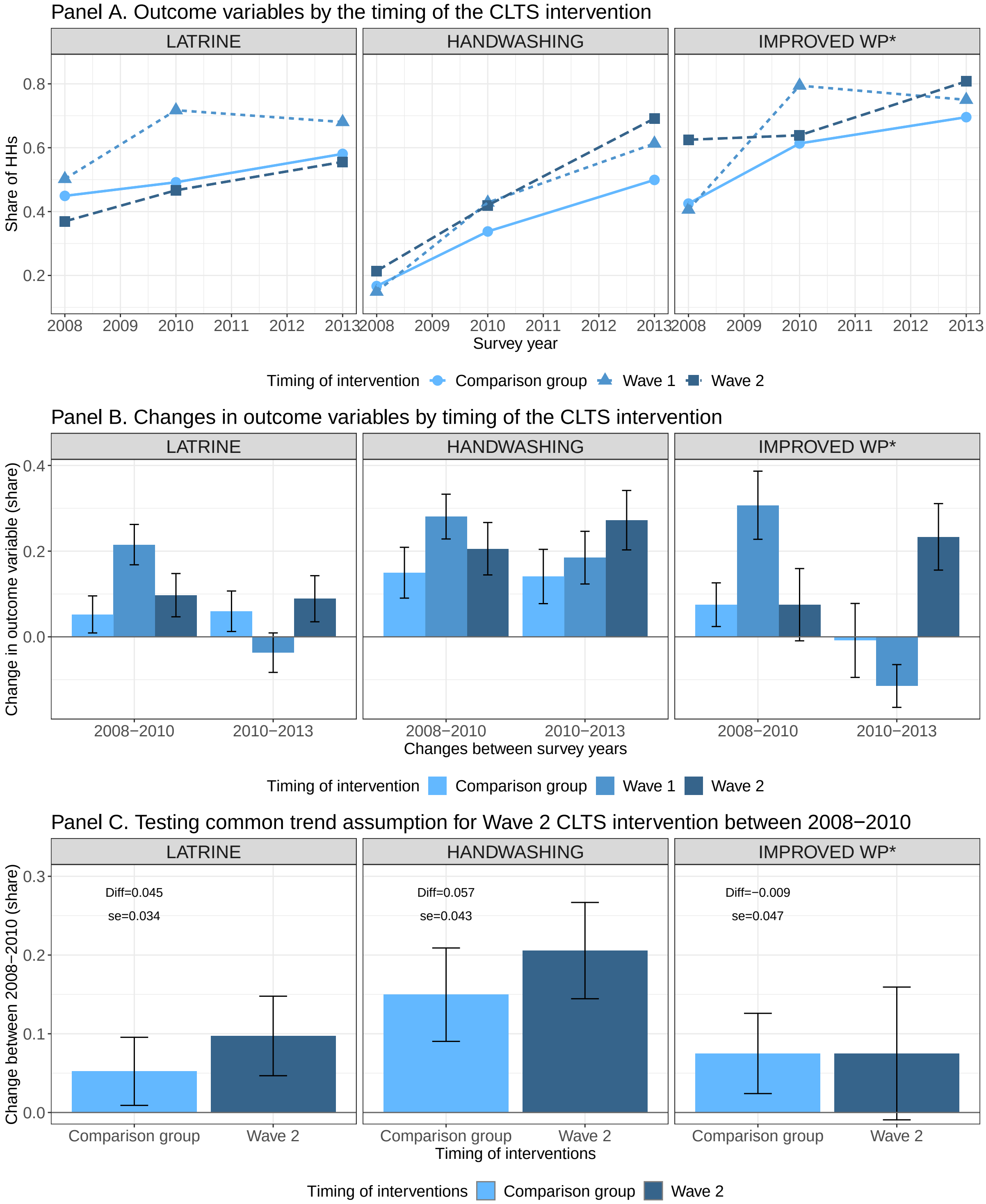
Figure 1 shows the changes in the outcome variables depending on the period of the CLTS intervention. Panel A shows the development in the share of households owning a latrine (left), handwashing (middle) and using an improved WP conditional on its availability (right). The outcomes are shown separately for the Wave 1 and Wave 2 CLTS intervention groups and the comparison group.[20] Panel B shows the corresponding changes in the outcome variables between the survey periods for the three groups. For an intervention with a positive treatment effect, the outcome is expected to increase (more than in the comparison group) from 2008 to 2010 for the Wave 1 intervention group, and from 2010 to 2013 for the Wave 2 intervention group. The outcome on handwashing follows this pattern the most closely. Note that there is

D.

19 Availability of an improved WP is defined here as a surveyed functioning improved (community) WP in the community.

20 The comparison group includes communities that did not receive any CLTS or WP interventions. We omitted data on communities that only received the WP intervention.

Figure 1: Changes in outcome variables by the period of the CLTS (total) intervention



Note: 95 percent confidence interval included around mean

** Use of improved WP is conditional on their availability in the community.*

Difference calculated controlling for HH size and wealth index. Std.error clustered at community level.

also a positive and significant trend in the comparison group for all three outcome variables.[21]

The changes in latrine ownership are significantly smaller (diff=-12.6pp, p=0.0005) after the CLTS intervention for the Wave 2 group compared to the Wave 1 intervention group. We argue that this pattern is the one we would expect if the implementing NGOs selected the location of the CLTS interventions based on where they expected the largest treatment effect in terms of the communities becoming open defecation free, as they were incentivized to do.

Panel C of Figure 1, looks at whether the changes in the outcome variables follow the same trend in the comparison and Wave 2 intervention groups between 2008 and 2010, before the intervention. In the estimation of the treatment effects, we need to make the assumption that the comparison and intervention groups follow the same trend. We can only test this assumption for the Wave 2 intervention group before the intervention. The figure shows that the Wave 2 intervention group had a somewhat higher trend than the comparison group for latrine ownership and handwashing. However, the difference is not significantly different from zero (p=0.185 and p=0.191, respectively) when controlling for observable baseline household characteristics (household size, wealth index and education).

Looking at the persistence of the outcome variables over time, we find that the unconditional retention rate of latrine ownership and the use of improved WP was around 80-90%. This means that of the households that owned a latrine or used an improved water point at one of the survey rounds, 80-90% still did so at the time of the next survey round. At the same time, we observe that the retention rate for handwashing was only 52-65% (see Table A5 in Appendix A). In the analysis, we investigate whether households facing an adoption or a retention decision are affected differently by the program interventions and their household characteristics.

The focus of the paper is on the effectiveness of the CLTS intervention with and without WP interventions. In section 3, we estimate the effect of CLTS on latrine ownership and handwashing. In section 4, we investigate the relationship between the

[21] The positive trend in the comparison group can partially be attributed to other simultaneously running water and sanitation interventions and to increasing wealth levels.

improved WPs and the CLTS intervention. Finally, in section 5 we investigate possible explanations how the CLTS and WP interventions could reinforce one another.

3 Effects of CLTS on latrine ownership and handwashing

The explicit objective of the CLTS intervention was to encourage latrine use and ownership and handwashing with soap or ash especially after defecation. In this section, we investigate to what extent the CLTS intervention was effective in achieving these objectives. In addition to the treatment effect on the beneficiaries, we are also interested to know whether the effects on the beneficiaries would be the same if CLTS had been implemented in randomly selected communities. This is especially important because of the strategic selection of the CLTS recipient communities. We follow Vigh and Elbers (2017) when estimating the average treatment effect in the population.

We also address whether the CLTS intervention was more effective in communities also receiving the WP intervention, and whether CLTS had a different effect on new adopters compared to its effect on the retention of hygienic practices of earlier adopters. The results should be interpreted as mid- to long-term effects of the CLTS intervention. Before discussing the results, we outline the econometric approaches used to generate the results.

3.1 Identification strategy

In order to assess the effectiveness of the program interventions, we consider the average treatment effect on the actual intervention communities (Average Treatment Effect on the Treated, ATT) and in the population (Average Treatment Effect, ATE). We estimate these treatment effects using a regression framework.

Due to the participatory nature of the CLTS intervention, its effects are likely to be heterogeneous among the communities. Therefore, we include community-specific treatment effects (β_c) in the model for estimating the treatment effects on the outcome variables (Y_{it}):

$$Y_{it} = \alpha_t + D_{c(i)t}\beta_{c(i)} + X_{it}\theta + \eta_i + \varepsilon_{it} \quad \text{for } t = 1, ..., T \tag{1}$$

where α_t denotes the time-varying but common trend for all individuals, X_{it} stands for household characteristics (household size, education and wealth index) with constant coefficients (θ), η_i is a household fixed effect and ε_{it} stands for idiosyncratic shocks. All the outcome variables of interest are binary variables. In the main analysis we use the linear probability model to estimate the treatment effects.[22]

We define the intervention variables D such that $D_{ct} = 1$ if an intervention was implemented in community c before time t, and $D_{ct} = 0$ otherwise.[23] D represents a vector of the intervention arms for the CLTS and WP interventions as shown in the two panels of Table 1.

We expect the he intervention arms (D) containing CLTS to be correlated with the community-specific treatment effect (β_c) because the implementing NGOs were incentivized to carry out the CLTS trainings in communities where they expected to have the largest treatment effect based on assessment visits at the potential intervention locations prior to delivering the intervention. We do not expect the treatment effect of the WP intervention to be community-specific, however, we allow for it in the regression specificiation.

We estimate the ATE and ATT following Vigh and Elbers (2017), who discuss the modeling and estimation of treatment effects in selective intervention placement models in cases such as the current setting. The identification of the ATE relies on the assignment mechanism that allocates the intervention to communities with the highest expected treatment effect given the number of communities to be served. The estimation procedure is summarized below.

[22] The average partial effects using a probit regression follow closely the results of the linear probability model (see Appendix G).

[23] This is also referred to as staggered intervention design, with the intervention roll-out implemented in stages.

Average Treatment Effect

The key insight for the estimation of the ATE is that, when the interventions are allocated to communities in order of the expected treatment effect, the period of the intervention implementation carries additional information about the treatment effect that researchers can exploit when estimating the ATE. This information is conveyed by the decrease in the treatment effect over time as the intervention is rolled out. To observe this, we need at least three rounds of observations, which are available from the program surveys (see Figure 1).

Following the correlated random slopes model of Wooldridge (2019), we can approximate the correlation between the community-specific treatment effect and the intervention placement by the community means of the intervention variable over time. Using the interaction term of the mean intervention and the intervention itself, we are able to filter out the component of the treatment effect heterogeneity which is correlated with the intervention.[24]

Adopting the described procedure to (1), we estimate ATE by the average treatment effect (β) in the following correlated random slopes (CRS) model

$$Y_{it} = \alpha_t + D_{ct}\beta + X_{it}\theta + D_{ct} \otimes \overline{\overline{D}}_c \xi + \eta_i + u_{it} \tag{2}$$

where $\overline{\overline{D}}_c = \overline{D}_c - \mu_{\overline{D}}$ is the (demeaned) community mean of the intervention variable over time, $\overline{D}_c = 1/(TN_c)\sum_{t=1}^{T}\sum_{i=1}^{N_c} D_{it}$ is the community mean of the intervention D_{ct} over time, and $\mu_{\overline{D}} = 1/C\sum_{c=1}^{C}\overline{D}_c$ is the mean of the community means. $D \otimes \overline{\overline{D}}$ denotes all interaction terms between the multidimensional D and $\overline{\overline{D}}$.[25]

Vigh and Elbers (2017) argue that the most important assumptions for estimating ATE using (2) are that the heterogeneity in the treatment effect is constant over time, that no selection occurred on time-varying unobservable variables ($E(u_{it}|D, X, \eta_i) = 0$), and that the correlation between the treatment effect heterogeneity and the intervention variable is indeed due to the allocation mechanism

[24] These community mean of the intervention variable also has to be demeaned over the whole sample in order for the level of these variables not to effect the estimate of the average treatment effect.

[25] It is possible to include other time-independent interaction terms ($D \otimes \overline{\overline{W}}$) in equation (2) that contribute to the treatment effect heterogeneity.

described above.[26]

In addition, we also must assume that the common trend assumption holds, and we have to use a population representative sample. Given these assumptions, β in (2) provides an unbiased estimate of the ATE.

We estimate (2) using the within transformation (fixed effects regression). Because residual heterogeneity in the treatment effect (uncorrelated to the intervention) is flushed into the error term u_{it}, we adjust the standard errors for serial correlation, in addition to correcting for heteroskedasticity and clustering.

Average Treatment Effect on the Treated

In order to estimate the ATT, we use the standard difference-in-difference (DD) regression:

$$Y_{it} = \alpha_t + D_{ct}\tilde{\beta} + X_{it}\theta + \eta_i + u_{it} \tag{3}$$

where, the estimate of $\tilde{\beta}$ contains the treatment effect on the actual beneficiaries possibly including the effects of selective intervention placement. Hence, the results are not directly comparable to treatment effect estimates from randomized controlled trials. Comparing the estimates of the ATT and the ATE can help assessing the importance of the selectivity. Elbers and Gunning (2014) argue that for policymakers the treatment effect of the program inclusive of the effects of selection may actually be more relevant, unless the program is aimed at providing universal coverage.

We can estimate equation (3) using the within or the first difference transformation to filter out the household fixed effects (η_i). Taking the second approach allows us to consider more closely the changes in the outcome variables over time. In the current setting, these changes arise from two sources: adoption behavior of previous non-users ($Y_{i(t-1)} = 0$ and $Y_{it} = 1$) and non-retention of previous users ($Y_{i(t-1)} = 1$ and $Y_{it} = 0$). The effectiveness of the interventions may differ for these two groups.[27] For example,

[26] If the intervention allocation is based on incorrectly formed expectations, this weakens the correlation between the intervention variable and the treatment effect, thereby weakening the effect of selection on gains.

[27] See Tables A6 and A7 in Appendix A for the unconditional transition dynamics of these two groups with and without the CLTS intervention.

using a comparable specification, Gertler et al. (2015) finds that the program effects on the use of sanitation facilities can be primarily attributed to non-owner households at the baseline, in most context.

In order to investigate differential treatment effects on previous users and non-users, we estimate the following regression for the changes in the outcome variables (ΔY_{it}):

$$\Delta Y_{it} = \begin{cases} \alpha_{0t} + \Delta D_{ct}\tilde{\beta}_0 + \Delta X_{it}\theta_0 + X_{i(t-1)}\gamma_0 + v_{it} & \text{if } Y_{i(t-1)} = 0 \\ \alpha_{1t} + \Delta D_{ct}\tilde{\beta}_1 + \Delta X_{it}\theta_1 + X_{i(t-1)}\gamma_1 + v_{it} & \text{if } Y_{i(t-1)} = 1. \end{cases} \tag{4}$$

This regression is a variation of the first difference transformation of (3). In addition to the changes in the household-specific variables (ΔX_{it}), we also control for their past values ($X_{i(t-1)}$) assuming that both the level of the household-specific variables and their changes may influence the adoption and retention behavior of the households. The separate coefficients for $\tilde{\beta}_0$ and $\tilde{\beta}_1$ provide a decomposition of the ATT for the beneficiaries who did or did not adopt the desired outcomes before the start of the intervention.

It is important to note that even though we are interested in the synergies between the CLTS and WPI intervention, we do not estimate a model with an interaction term between 'CLTS (total)' and 'WPI (total)'. There are two main reasons for this. First, in the case of the 9 communities that received the two interventions in different waves, the coefficient of the interaction term would only measure the additional effect of the second intervention, while for the 21 communities it would measure the combined effect of the interventions. Splitting the intervention arms as shown in Table 1b avoids this problem.

Second, we are primarily interested in the effectiveness of the CLTS intervention, and whether it is more or less effective when implemented together with the WP intervention. Therefore, we want to test the size of the coefficients of the 'CLTS (alone)' against 'CLTS&WPI' intervention arm. In addition, we expect that the WP intervention alone does not significantly improve sanitation outcomes.

3.2 Treatment effect estimates

Table 2: Treatment effect estimates for latrine ownership and handwashing

	Treatment effect estimates					
	Latrine	Latrine	Latrine	HW	HW	HW
	(1)	(2)	(3)	(4)	(5)	(6)
CLTS (total)	0.076**	0.094**	0.032	0.128***	0.114**	0.114*
	(0.038)	(0.045)	(0.041)	(0.038)	(0.044)	(0.064)
WPI (total)	0.035	0.020	0.018	0.012	−0.006	−0.034
	(0.031)	(0.036)	(0.036)	(0.039)	(0.044)	(0.058)
Mean dep.var.	0.541	0.077	0.541	0.379	0.202	0.379
Observations	4206	2667	4206	4170	2632	4170
R^2	0.067	0.051	0.071	0.180	0.104	0.181
Regression model	DD	DD	CRS	DD	DD	CRS
Estimator	FE	FD	FE	FE	FD	FE
P-value($H_0 : \xi_{CLTS} = 0$)			0.030			0.792

Notes: Dependent variables: Latrine = latrine ownership, HW = handwashing with soap or ash after defecation. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Regression model DD estimates equation (3) and CRS estimates equation (2). Standard errors corrected for clustering at community level. All regressions control for HH size, wealth index, education and year. Sample includes HHs participating in at least 2 survey rounds.

We first discuss the effects of the CLTS intervention irrespective of the joint occurance of the interventions. Table 2 reports the treatment effect estimates for latrine ownership (columns 1-3) and handwashing (columns 4-6). In addition to the CLTS and WP interventions, all regressions control for household characteristics (wealth index, education and size), and household and time fixed effects. In columns 2 and 5 the first difference transformation (FD) is used to remove the fixed effects, while the other regressions use the within transformation (FE). As expected, the WP intervention on its own had no significant effect on the outcomes, and we do not discuss this result further.

The table shows that the CLTS intervention is associated with a 7.6 percentage point (pp) increase in latrine ownership (p=0.043) in the selected intervention communities. This effect is a 17% increase compared to the baseline latrine ownership in the CLTS intervention group. The general increase in latrine ownership was 9.3pp on average in all communities between 2008 and 2013, which can be partially attributed to increasing wealth levels and other sanitation trainings in the program area.[28]

Using the FD transformation to estimate the ATT, only the changes immediately after the intervention are taken into account. Hence, changes between 2010 and 2013 for the Wave 1 intervention group are not included in the FD estimate but they do affect the FE estimate. In the case of latrine ownership, the ATT of the CLTS intervention using the FD transformation (9.4pp in column 2) is higher than using the FE transformation (7.6pp in column 1) because latrine ownership decreased between 2010-2013 for the Wave 1 intervention group (see Figure 1).

Column 3 of Table 2 reveals that the estimate of ATT on latrine ownership is in a large part driven by the selection of the intervention communities: the average treatment effect in the total population of the program area would have been lower (ATE=3.2pp) and not significantly different from zero (p=0.431).This occurs due to the reduction in the effectiveness of the CLTS intervention between the Wave 1 and 2 intervention groups (see Figure 1), which we argue is driven by the selection of intervention communities. This effect is captured by the cross-terms $D \otimes (\bar{D} - \mu_{\bar{D}})$ with coefficient ξ_{CLTS}, which have a significant explanatory power in the CRS regression (p=0.030). If the NGOs selected communities where they expected the largest successes in terms of latrine ownership from the available pool of communities, we would also expect to see a continuous reduction in the size of the treatment effect in expectation. This is a likely scenario given that the NGOs were (financially) motivated to implement the CLTS intervention in communities where all households would start using latrines. More importantly, the validity of the ATE estimate depends on this assumption. If the change in the effectiveness of the CLTS intervention is caused by alternative explanations, for example, the population of the communities became saturated by all the sanitation trainings, then the CRS equation (2) would fail

[28] We only have self-reported information about other sanitation trainings but it is likely that these have also contributed to the rising number of latrines.

Table 3: Effect of CLTS was higher when implemented together with WP intervention

	Treatment effect estimates					
	Latrine	Latrine	Latrine	HW	HW	HW
	(1)	(2)	(3)	(4)	(5)	(6)
CLTS (alone)	0.069	0.079	−0.059	0.097**	0.070	0.091
	(0.048)	(0.057)	(0.078)	(0.049)	(0.053)	(0.104)
CLTS&WPI	0.115**	0.122*	0.010	0.157***	0.133**	0.085
	(0.049)	(0.063)	(0.072)	(0.054)	(0.061)	(0.136)
WPI (alone)	0.029	0.009	0.038	−0.013	−0.042	−0.136
	(0.030)	(0.030)	(0.075)	(0.045)	(0.050)	(0.092)
Mean dep.var.	0.541	0.077	0.541	0.379	0.202	0.379
Observations	4206	2667	4206	4170	2632	4170
R^2	0.068	0.051	0.072	0.181	0.105	0.182
Regression model	DD	DD	CRS	DD	DD	CRS
Estimator	FE	FD	FE	FE	FD	FE
P-value($H_0 : \beta_{CLTS} = \beta_{C\&W}$)	0.428	0.566	0.516	0.287	0.336	0.973
P-value($H_0 : \xi_{CLTS} = 0$)			0.078			0.970
P-value($H_0 : \xi_{C\&W} = 0$)			0.169			0.511

Notes: Dependent variables: Latrine = latrine ownership, HW = handwashing with soap or ash after defecation. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Regression model DD estimates equation (3) and CRS estimates equation (2). Standard errors corrected for clustering at community level. All regressions control for HH size, wealth index, education and year. Sample includes HHs participating in at least 2 survey rounds.

to estimate the Average Treatment Effect in the population.

Turning to the results on handwashing, columns 4-6 in Table 2 show that the treatment effect of CLTS on handwashing has been sizable and persistent over time. The average treatment effect of the CLTS intervention on handwashing with soap or ash after defecation is estimated at 11.4pp in the program area (column 6). This is only about 1pp lower than the treatment effect on the treated (column 4), and it is a 65% increase in handwashing compared to its baseline value. This also comes on top of a 36.3pp general increase in handwashing between 2008 and 2013, which is a 206% increase compared to the baseline.

Next, in Table 3 we investigate whether the treatment effect of CLTS differs when implementing the CLTS intervention with or without the WP intervention. The table shows that the ATT is larger for both latrine ownership and handwashing in the CLTS&WPI intervention arm (ATT=11.5pp and ATT=15.7pp, respectively) compared to implementing CLTS alone (ATT=6.9pp and ATT=9.7pp). This is an important finding because for latrine ownership (and also for the FD estimate for handwashing) the ATT is only significantly different from zero when the CLTS is implemented together with the WP intervention. We investigate this result further in section 5.

The estimate of the ATE shows that even when implementing CLTS together with the WP intervention, the treatment effect of CLTS on latrine ownership would not be statistically significant when implemented at randomly selected or in all potential target communities. The ATE for handwashing is also not significantly different from zero. The latter occurs partially due to the increase in standard error when controlling for the cross-terms ($D \otimes (\bar{D} - \mu_{\bar{D}})$), and partially due to the reduction in the ATE for the CLTS&WPI compared to the ATT. However, the difference between the coefficient estimates in column 4 and 6 are small, and the coefficient estimates of the cross-terms ($D \otimes (\bar{D} - \mu_{\bar{D}})$ and $D \otimes (\bar{X} - \mu_{\bar{X}})$) are not significantly different from zero (p=0.970 and p=0.511). This suggests that the ATT estimates in column 4 may estimate the ATE for handwashing reasonably well just like when estimating the total effect of CLTS on handwashing in Table 2.

Even if the CLTS intervention would have little additional effect on latrine ownership when implemented universally, it is worthwhile to consider whether the effects of the intervention among the beneficiaries come from new adopters ($Y_{i(t-1)} = 0$) or households that keep maintaining their latrines ($Y_{i(t-1)} = 1$). To address this question, Table 4 reports the estimates of ATT conditional on the before-intervention-period outcome of the households as described in equation (4). In column 1, we observe that the ATT of CLTS on latrine ownership can be fully attributed to the effect of the intervention on new adopters. The difference in ATT between the two groups is significantly different from zero (p=0.008). This is interesting given that the CLTS interventions were more likely to be allocated to communities with relatively higher latrine ownership at the baseline.

Table 4: Differential treatment effects for new adopters and the retention of earlier adopters

	Dependent variable:	
	Latrine ownership	Handwashing
	(1)	(2)
New adopters: CLTS (total)	0.180***	0.111***
	(0.067)	(0.042)
Retention: CLTS (total)	0.021	0.102**
	(0.029)	(0.046)
Mean dep.var.	0.077	0.202
P-value($H_0 : \tilde{\beta}_0 = \tilde{\beta}_1$)	0.008	0.866
Observations	2,667	2,632
Adjusted R^2	0.287	0.452

Notes: Conditions: 'New adopters': $Y_{i(t-1)} = 0$ and 'Retention': $Y_{i(t-1)} = 1$. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). First difference regression of equation (4) controlling for the past value and changes in wealth index, education and household size, and year. Controls for the WP intervention are excluded. Sample includes HHs participating in at least 2 survey rounds.

In addition, we also observe that 71% of the households owning a latrine emptied it or changed its location between 2008 and 2010, and 64% did so between 2010 and 2013. Households who emptied or moved the latrine, did so on average 1.3 and 1.4 times between 2008-2010 and 2010-2013, respectively. This signals a real behavior adaptation to using latrines.

The results in column 2 for handwashing show that the CLTS intervention was just as effective for households who did or did not wash their hands with soap/ash before the intervention (p=0.866). The finding that CLTS has a positive effect on retaining handwashing practice (ATT=10pp) is important considering that about 60% of the households stop using soap/ash when washing hands after defection in the communities without a CLTS intervention (Table A6d in Appendix A).

These results are observed both for the CLTS&WPI and CLTS alone intervention arms. There is very little difference in the coefficient estimates between the two intervention arms (see Table F1 in Appendix F).

The regression estimates of equation (4) for latrine ownership also reveal that a one standard deviation increase in the wealth index increases the likelihood of latrine ownership by about 4pp (p=0.0000) irrespective of latrine ownership in the previous period. Education and household size do not significantly affect changes in latrine ownership.

The likelihood of both the adoption and retention of the desired handwashing practices is increased by 11pp (p=0.007) when at least half of all adult household members have any level of formal education (results available upon request). We also observe an 8pp (p=0.004) increase in the adoption of handwashing practices when the education level increases in the household (results available upon request).

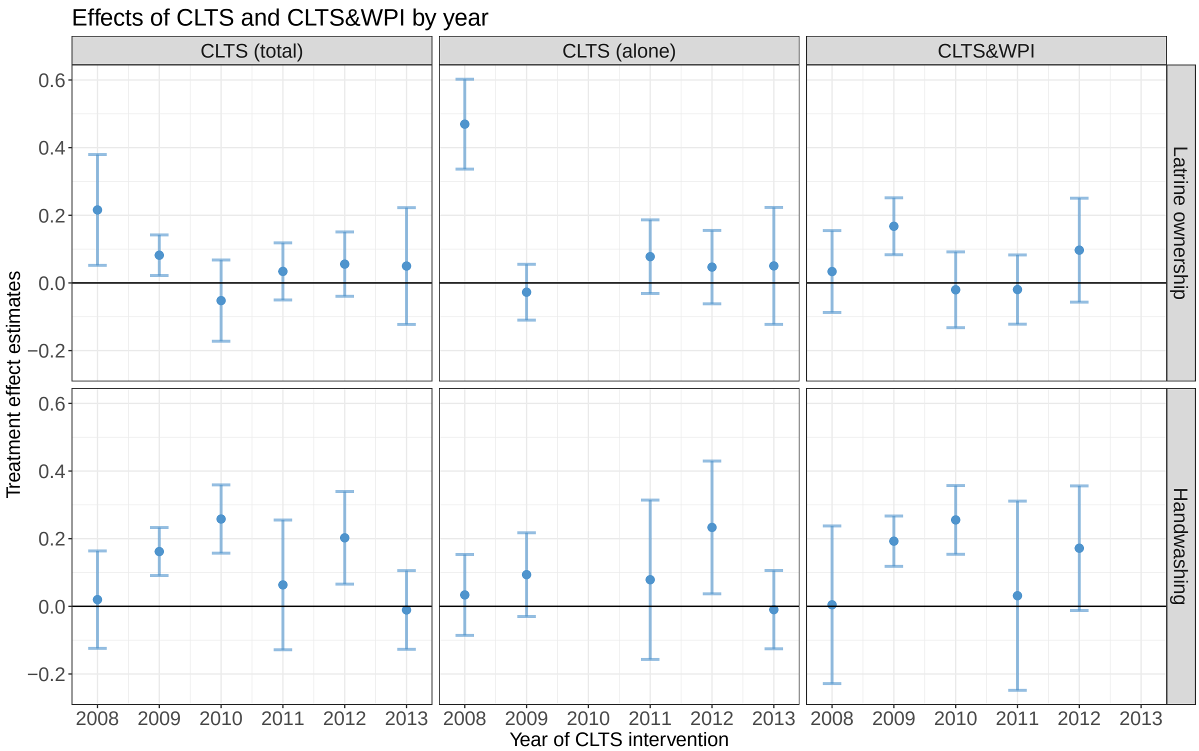
3.3 Robustness check

In order to calculate the ATE, we needed to make the assumptions that the CLTS intervention was delivered in the order of the size of the expected treatment and that these treatment effects do not change over time in the communities. Therefore, here we investigate whether the treatment effect of CLTS (on latrine ownership) indeed decreased over time. We can use the observation on the actual year when the intervention was implemented in the communities (s) to estimate the following regression specification:

$$Y_{it} = \alpha_t + \sum_{s=2008}^{2013} D_{ct}^{s}\beta_s + X_{it}\theta + \eta_i + u_{it} \tag{5}$$

where D_{ct}^{s} now indicates whether the intervention in year s was implemented in community c. We estimate a separate coefficient (β_s) for each intervention year. Note that the observations for the outcomes (Y_{it}) still come from the three survey rounds used in the main analysis.

Figure 2: Effects of CLTS and CLTS&WPI by year (95% confidence interval)



Note: 95% confidence interval around coefficient estimates.

Figure 2 shows the estimated size of the ATT by the year of CLTS intervention for all the CLTS interventions, and separately for the CLTS (alone) and CLTS&WPI treatment arms. Looking at the treatment effect of all CLTS interventions on latrine ownership (top left panel), we observe that the point estimate of the treatment effect was the highest in 2008, followed by 2009. However, these differences are not significantly different from the treatment effects in other years. Disaggregating the CLTS intervention into the CLTS alone (middle) and the CLTS&WPI (right) intervention arms, the treatment effect of the CLTS alone intervention delivered in 2008 stands clearly out with a treatment effect of over 40pp.

The CLTS interventions delivered in 2008 (after the baseline survey) could be problematic because these communities were promised a classroom or a borehole if their open defecation free (ODF) status was verified during the ODF evaluation campaign of 2008 (see Appendix B). Only 1 of the 4 communities with a CLTS intervention in 2008 became ODF in 2008 (in the CLTS&WPI intervention arm).

However, this incentive could have played a role in motivating communities to build latrines. Therefore, in Table 5 we also report the regression results omitting communities that received the CLTS intervention in 2008. The results do not change significantly compared to Table 3.

The treatment effects on handwashing with soap or ash after defecation (bottom left panel) show a different pattern. The figure suggests a fade-out effect as the size of the treatment effect increases towards the survey rounds (2010 and 2013). We indeed observe 30-40pp reversion rates for handwashing with soap (see Table A6c in Appendix A).

Additionally, we also investigate the robustness of the main results with respect to the functional specification of regressions (2) and (3). We use sampling weights when calculating the ATE to provide a representative sample; we use the probit instead of the linear probability model for both the ATE and the ATT; and we use propensity score matching as an alternative to the difference-in-difference method used to calculate the ATT. These results remain consistent with our main findings (see Appendix G).

3.4 Summary

We find that the CLTS intervention had an about 11pp additional effect on the adoption and retention on handwashing practices among the intervention beneficiaries. We expect to find the same effects had CLTS been implemented in randomly selected communities. The treatment effects are less sustainable on latrine ownership. The results indicate that the CLTS intervention increased latrine ownership among the intervention beneficiaries by 8pp on average, which stems from an 18pp increase in latrine ownership among new adopters. However, the treatment effect decreased over time, and the estimate of the ATE indicates that CLTS would not be effective on latrine ownership when implemented in the whole program area or in randomly selected communities.

Why do we observe such persistent effects for handwashing when the effect of CLTS on latrine ownership appears to be falling over time? First of all, buying a piece of soap is much more affordable than obtaining material and building a latrine.

Table 5: Robustness check: treatment effects omitting communities receiving CLTS in 2008

	Treatment effect estimates			
	Latrine	Latrine	HW	HW
	(1)	(2)	(3)	(4)
CLTS (alone)	0.020	0.008	0.099**	0.061
	(0.041)	(0.077)	(0.050)	(0.112)
CLTS&WPI	0.117**	−0.015	0.168***	0.057
	(0.046)	(0.062)	(0.048)	(0.126)
WPI (alone)	0.025	0.028	−0.020	−0.168*
	(0.024)	(0.074)	(0.044)	(0.092)
Mean dep.var.	0.538	0.538	0.383	0.383
Observations	3969	3969	3969	3969
R^2	0.064	0.068	0.186	0.188
Regression model	DD	CRS	DD	CRS
Estimator	FE	FE	FE	FE
P-value($H_0 : \beta_{CLTS} = \beta_{C\&W}$)	0.051	0.800	0.227	0.979
P-value($H_0 : \xi_{CLTS} = 0$)		0.168		0.835
P-value($H_0 : \xi_{C\&W} = 0$)		0.053		0.246

Notes: Dependent variables: Latrine = latrine ownership, HW = handwashing with soap or ash after defecation.

Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Regression model DD estimates equation (3) and CRS estimates equation (2). All regressions control for HH size, wealth index, education and year. Sample includes HHs participating in at least 2 survey rounds.

Interestingly, the lower investment costs may also result in less commitment. Table A5 in Appendix A shows that only about 60% of the households keep washing their hands with soap also in the next survey round (c) compared to above 80% of households who keep using a latrine (a). Second, in line with our identification strategy for the ATE, the NGOs were incentivized to select locations based on the likelihood of all households in the community adopting latrine use, and not handwashing. Given that handwashing and latrine ownership do not appear to be correlated with one another (see Table A4 in Appendix A), it is likely that even if the treatment effect on handwashing varies between communities and households it will not be correlated with the CLTS intervention placement (as we observe in Table 2).

Finally, we also find that CLTS appears to be more effective when implemented together with the WP intervention. We continue with examining the relationship between access and use of improved WPs and the program interventions.

4 Availability of improved water points and the effectiveness of CLTS

In this section, we investigate the program's impact on the use of improved WPs. Due to the WP intervention, we can investigate which factors affect the households' decision to switch to using an improved WP, and whether the CLTS intervention contributes to this decision.

4.1 Determinants of using improved water points

We start the analysis by documenting the increase in improved WPs as a result of the WP intervention. As discussed above, the interventions were specifically aimed at communities in need of an improved community WP. Table 6 shows the availability of improved WPs split by the timing of WP intervention. The information is based on the WP survey. The availability of improved WP equals 1 if an improved WP has been surveyed in the community, and 0 otherwise. The WP survey was implemented at all WPs used by the surveyed households. The table suggests a significant increase in

Table 6: WP intervention (total) increased the percentage of communities with a functioning improved WP

Intervention group	2008 (%)	2010 (%)	2013 (%)	Sample size
No WPI	42.4	45.5	46.9	33
Wave 1 WPI	20.0	82.9	91.4	35
Wave 2 WPI	8.3	33.3	75.0	12

Notes: Percentage of communities with a functioning improved WP according to the period of the WP intervention (WPI).

Table 7: Not all households used the improved WP when it was available

Variable	2008	2010	2013	Total
Improved WP use (%)	45.5	68.8	74.1	67.1
Sample size	440	960	1,119	2,519

Notes: Percentage of households using an improved WP conditional on its availability in the community.

availability as a result of the intervention. There is almost no change in the availability for the non-intervention communities. Regressing the availability of improved WPs on the WP intervention and community and year fixed effects, we observe that the WP intervention increases the probability of the availability of an improved WP by 55pp ($p<0.0001$).

Interestingly, we do not observe 100% availability of improved WP at the intervention communities. This is in a large part due to the fact that the new program WPs were not always installed in the vicinity of the sampled communities within the villages. We observe that three quarters of the surveyed households in 2010 fetch water from within 1 km from their dwelling. Therefore, if a new WP is installed on the other side of the village, they will not use it.[29]

Table 7 shows that, even when an improved water source becomes available, not all households will actually switch to using it. While improved WP use conditional on availability has been increasing over time, even in 2013, more than a quarter of the households choose to use traditional water sources. The most important factors determining the adoption decision to use an improved WP include the distance to the improved and traditional WPs and the perception of the water quality/safety at the improved and traditional WPs. These factors were the most often mentioned reasons by the households for using their chosen WP during the endline survey (see Table H1 in Appendix H), and regression analysis in Appendix H also confirms this finding.

4.2 Effects of CLTS on the use of improved water points

The CLTS intervention may raise the awareness of households regarding the benefits of using improved WPs and may also affect the perceived water quality of the traditional WPs. However, we expect that the WP intervention affects only the availability of the improved WPs.[30] We investigate these hypothesis below.

We estimate the effect of the CLTS and WPI interventions on the use of improved WPs conditional on access to an improved WP using the specifications for ATT and ATE. However, in the regressions we also control for the distance to the WPs and the perception of the water quality/safety. Distance is measured as the crow flies using GPS coordinates. We measure water quality using the mean valuation of the households within communities separately for the improved and traditional WPs (when available) because perceived water quality is only observed for the WP actually used by the households. Therefore, in the regression analysis we use community fixed effects instead of household fixed effects.

The results for the total effect of CLTS and WPI are reported in Table 8. Column 1

[29] We indeed observe that Intended Treatment communities with the WP intervention (where survey households were sampled in the neighborhood of the desired location of the new WP) are more likely to have a surveyed improved WP than Intended Control communities with the WP intervention.

[30] The WP intervention also usually affects the payment for water. However, this is not considered in the analysis.

Table 8: CLTS affected the use of improved WPs (conditional on the availability of improved WPs)

	Dependent variable:			
	Use of improved WP			
	(1)	(2)	(3)	(4)
CLTS (total)	0.184***	0.153**	0.149	0.104
	(0.062)	(0.071)	(0.152)	(0.135)
WPI (total)	0.106	0.069	0.219	0.083
	(0.070)	(0.089)	(0.162)	(0.133)
Distance to improved WP	−0.130***	−0.132***	−0.301***	−0.311***
	(0.027)	(0.046)	(0.040)	(0.043)
Distance to traditional WP			0.210***	0.211***
			(0.044)	(0.044)
Mean quality at improved WP	0.153***	0.169**	0.018	0.032
	(0.054)	(0.073)	(0.069)	(0.069)
Mean quality at traditional WP			−0.123***	−0.121**
			(0.047)	(0.048)
Mean dep.var.	0.709	0.709	0.573	0.575
R^2	0.117	0.128	0.195	0.206
Regression model	DD	CRS	DD	CRS
Estimator	FE	FE	FE	FE
P-value($H_0 : \xi_{CLTS} = 0$)		0.045		0.000
Observations	1,642	1,642	1,071	1,068

Notes: Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). All regressions control for HH size, wealth index, education, year and community fixed effects. Sample includes all HHs in communities where an improved WP (1&2) and both improved and traditional WP (3&4) were available in the given survey round.

Table 9: Effect of CLTS was higher when implemented together with the WP intervention (conditional on the availability of improved WPs)

	Dependent variable:			
	Use of improved WP			
	(1)	(2)	(3)	(4)
CLTS (alone)	0.145*	−0.101	0.149	−0.225*
	(0.086)	(0.104)	(0.152)	(0.125)
CLTS&WPI	0.363***	0.490***	0.368***	0.342***
	(0.118)	(0.106)	(0.070)	(0.093)
WPI (alone)	−0.002	−0.345		
	(0.091)	(0.245)		
Mean dep.var.	0.709	0.709	0.573	0.575
R^2	0.121	0.131	0.195	0.206
Controls for traditional WP	No	No	Yes	Yes
Regression model	DD	CRS	DD	CRS
Estimator	FE	FE	FE	FE
P-value($H_0 : \beta_{CLTS} = \beta_{C\&W}$)	0.095	0.000	0.176	0.000
P-value($H_0 : \xi_{CLTS} = 0$)		0.057		0.000
P-value($H_0 : \xi_{C\&W} = 0$)		0.365		0.917
Observations	1,642	1,642	1,071	1,068

Notes: Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). All regressions control for IWP characteristics, HH size, wealth index, education, year and community fixed effects. Sample includes all HHs in communities where an improved WP (1&2) and both improved and traditional WP (3&4) were available in the given survey round.

of the table shows the ATT estimates for CLTS and WP intervention when controlling for the characteristics of the improved WPs but not the traditional WPs. CLTS increased the probability of using an improved WP by 18pp (p=0.003) for the intervention beneficiaries who had access to an improved WP. This effect appears to be sustained, so that the population average of the treatment effect is estimated at 15pp (p=0.03). CLTS also had a similar effect in the sub-sample of communities where households use both improved and traditional WP in a given period (columns 3 and 4).

The coefficient estimates of distance and perceived water quality are not affected by the inclusion of the intervention variables compared to Table H2 in Appendix H. The estimate of CLTS is also not affected by excluding perceived water quality from the regression (results available upon request), suggesting that the effect of CLTS does not come from its effect on perceived water quality.

Conditional on the availability of improved WPs, the WP intervention had a positive (11pp) but insignificant (p=0.132 in column 1) effect on the use of improved WPs. This is in line with the expectation that the WP intervention mostly affected the access to improved WPs but not adoption conditional on access.

Looking further into the intervention arms, Table 9 shows that in both sub-samples, the effect of CLTS was driven by the CLTS&WPI intervention, which consistently increased the probability of using an improved WP by 36pp (p=0.002). CLTS alone had only a marginally significant effect (ATT=14pp, p=0.091), which was not sustained over time.

The results suggest that the CLTS intervention significantly increased the effectiveness of the WP intervention in terms of getting the households to start using the installed improved WPs even after controlling for the distance to the improved WPs.

5 Complementarity of the CLTS and WP interventions

Our findings in sections 3 and 4 indicate that CLTS was more effective when implemented together with the WP intervention. In this section, we discuss different mechanisms that could explain the synergies between the two interventions.

One explanation could be that access to improved WP induces households to invest in hygienic practices, like handwashing with soap and latrine use. Based on this explanation, we would expect that CLTS would have a more pronounced effect on households that live closer to the improved WP, as distance is one of the most important predictor for the use of improved WPs conditional on access. We test this hypothesis in section 5.1 below.

A related second explanation could be that the improved WPs of the program reduce the time and effort it takes to fetch water, which frees up resources to adopt more hygienic sanitation practices. This is not likely to happen in the survey area, because in most cases the closest traditional WPs (mean=0.67km and sd=0.72 in 2008) are not further away from the household than the closest improved WPs (mean=0.88km and sd=0.98 in 2013) when they are available.

A third explanation could be that water and sanitation interventions reinforce one another. In this case, we would expect that households adopt multiple of the desired outcomes. This hypothesis is discussed in section 5.2.

5.1 Distance to improved water points and the effects of CLTS

In column 1 and 2 of Table 10 we estimate the ATT of CLTS and CLTS&WPI on latrine ownership and handwashing conditional on the household's distance to improved WP. Because the distance to improved WP is only available for households where an improved WP is available in the community, an indicator of whether there is an improved WP within 1 km from the household is used in the regression.[31] The results show that the distance to improved WP does not affect the effects of the CLTS and

[31] For households in communities without an improved WP, we assume that there is no improved WP within 1 km of the households.

Table 10: Distance to improved WP and the ATT of CLTS and CLTS&WPI

	Dependent variable:			
	Latrine	HW	Use of IWP	
	(1)	(2)	(3)	(4)
CLTS (alone)	0.069*	0.142**	0.185**	0.157
	(0.041)	(0.064)	(0.077)	(0.122)
CLTS (alone) * Dist to IWP < 1 km	0.0002	−0.088		
	(0.051)	(0.063)		
CLTS (alone) * Dist to IWP			−0.097	−0.589
			(0.148)	(0.393)
CLTS&WPI	0.121***	0.119**	0.535***	0.734*
	(0.046)	(0.052)	(0.144)	(0.423)
CLTS&WPI * Dist to IWP < 1 km	−0.014	0.082		
	(0.051)	(0.064)		
CLTS&WPI * Dist to IWP			−0.245***	−0.398***
			(0.092)	(0.115)
WPI (alone)	0.029	−0.004	0.002	−0.302
	(0.024)	(0.033)	(0.089)	(0.212)
Mean dep.var.	0.541	0.379	0.709	0.709
R^2	0.068	0.182	0.133	0.164
Regression model	DD	DD	DD	CRS
P-value($H_0 : \xi_{CLTS} = \xi_{CLTS*dist} = 0$)				0.000
P-value($H_0 : \xi_{C\&W} = \xi_{C\&W*dist} = 0$)				0.000
Observations	4,206	4,170	1,642	1,642

Notes: Dependent variables: Latrine = latrine ownership, HW = handwashing with soap or ash after defecation, use of IWP = use of improved WP. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). All regressions control for HH size, wealth index, education, year and household (1&2) or community (3&4) fixed effects. (3&4) also controls for distance and quality of IWP. Sample for (1&2) includes HHs participating in at least 2 survey rounds, (3&4) includes all HHs in communities where an IWP was available in the given survey round.

CLTS&WPI interventions on latrine ownership and handwashing (p=0.996 and p=0.163, respectively for CLTS, and p=0.787 and p=0.2 for CLTS&WPI). The results are robust to using different cutoff values for the distance to the improved WP (results available upon request).

More generally, the results in the CLTS only treatment arm suggest that access to improved WPs in it self does not increase the effectiveness of the CLTS intervention on latrine ownership (and handwashing with soap after defecation), as the coefficient of the interaction term of distance and the CLTS (alone) intervention is insignificant in Table 10.[32]

For the use of improved WPs, we may also expect that CLTS was more effective when the improved WPs are closer to the households. Columns 3 and 4 in Table 10 find evidence for this conditional on access to improved WP in the community. For households living at a 1 km distance from an improved WP, the ATT was 9pp for the CLTS and 29pp for the CLTS&WPI intervention arms. However, the CLTS had no positive effect on the use of improved WP for households with a distance of 2 km or more away from the improved WP, irrespective of being implemented together with the WP intervention.

5.2 Adoption of multiple outcomes

If the CLTS and WP interventions reinforce one another, then we would expect that households are more likely to jointly adopt latrine ownership (LO), handwashing with soap after defecation (HW) and using an improved WP (IW) when the CLTS and WP interventions are implemented together.

Table 12 shows the treatment effect (ATT) estimates and general trend for the individual and combined outcomes conditional on access to improved WPs. Conditional on the access to improved WP, the treatment effect of CLTS on latrine ownership (column 1) and handwashing (column 2) is similar between the CLTS only and CLTS&WPI treatment arms.

[32] 40% and 54% of the households in the CLTS only intervention arm had an improved WP within 1 km from their dwelling in 2010 and 2013, respectively.

Table 11: Treatment effects (ATT) on combined water, sanitation and hygiene outcomes

	Treatment effect estimates						
	LO	HW	IW	LO&HW	LO&IW	HW&IW	LO&HW&IW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CLTS (alone)	0.069	0.097**	0.219***	0.107**	0.195***	0.201***	0.161***
	(0.048)	(0.049)	(0.080)	(0.046)	(0.061)	(0.045)	(0.037)
CLTS&WPI	0.115**	0.157***	0.416***	0.154***	0.297***	0.269***	0.192***
	(0.049)	(0.054)	(0.082)	(0.042)	(0.073)	(0.061)	(0.053)
WPI (alone)	0.029	−0.013	0.400***	0.002	0.246***	0.202***	0.119***
	(0.030)	(0.045)	(0.079)	(0.036)	(0.062)	(0.046)	(0.037)
Trend 2008-2013	0.088***	0.341***	0.095**	0.241***	0.066*	0.089***	0.069***
	(0.029)	(0.047)	(0.046)	(0.034)	(0.035)	(0.031)	(0.024)
Mean dep.var.	0.541	0.379	0.362	0.232	0.218	0.167	0.112
Observations	4206	4170	4206	4170	4206	4170	4170
R^2	0.068	0.181	0.348	0.159	0.212	0.198	0.150
Regression model	DD	DD	DD	DD	DD	DD	DD
Estimator	FE	FE	FE	FE	FE	FE	FE
P-value($\beta_C = \beta_{C\&W}$)	0.428	0.287	0.069	0.373	0.255	0.339	0.613

Notes: Dependent variables: LO = latrine ownership, HW = handwashing with soap or ash after defecation, IW = use of improved water point. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Regression model DD estimates equation (3). Standard errors corrected for clustering at community level. All regressions control for HH size, wealth index, education and year. Sample includes HHs participating in at least 2 survey rounds.

Table 12: Treatment effects (ATT) on combined water, sanitation and hygiene outcomes conditional on access to improved WP

	Treatment effect estimates						
	LO	HW	IW	LO&HW	LO&IW	HW&IW	LO&HW&IW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CLTS (alone)	0.135**	0.102	0.150*	0.126**	0.202***	0.178***	0.178***
	(0.065)	(0.068)	(0.085)	(0.063)	(0.071)	(0.054)	(0.051)
CLTS&WPI	0.131*	0.138	0.414***	0.169*	0.271***	0.242***	0.195**
	(0.074)	(0.096)	(0.134)	(0.103)	(0.096)	(0.086)	(0.091)
WPI (alone)	−0.006	−0.023	0.035	−0.0004	0.093**	0.136	0.100
	(0.043)	(0.121)	(0.078)	(0.080)	(0.044)	(0.115)	(0.075)
Trend 2008-2013	0.069	0.299***	0.060	0.237***	0.054	0.177***	0.139***
	(0.050)	(0.070)	(0.071)	(0.055)	(0.058)	(0.053)	(0.041)
Mean dep.var.	0.568	0.438	0.708	0.280	0.441	0.336	0.225
Observations	2190	2182	1645	2182	1645	1641	1641
R^2	0.065	0.122	0.132	0.141	0.087	0.119	0.119
Regression model	DD	DD	DD	DD	DD	DD	DD
Estimator	FE	FE	FE	FE	FE	FE	FE
P-value($\beta_C = \beta_{C\&W}$)	0.966	0.723	0.065	0.693	0.514	0.480	0.860

Notes: Dependent variables: LO = latrine ownership, HW = handwashing with soap or ash after defecation, IW = use of improved water point. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Regression model DD estimates equation (3). Standard errors corrected for clustering at community level. All regressions control for HH size, wealth index, education and year. Sample includes HHs participating in at least 2 survey rounds. Sample is conditional on access to improved water points.

The combined outcomes (columns 4-7) have mostly higher coefficient estimates for the CLTS only intervention arm compared to the individual outcomes. It suggests that households receiving the CLTS intervention adopt more than one of the promoted outcomes, in particular in combination with the use of improved WPs, compared to households in non-intervention communities.

Interestingly, the treatment effect estimates between the CLTS only and CLTS&WPI intervention arm only widen when the use of improved WP is included in the outcomes (columns 3, 5-7). The difference actually narrows when multiple outcomes are measured, and almost disappears when all three outcomes are included (column 7). These results are similar when we do not condition on access to improved WPs (results available upon request).

The results in the table suggest that combining the CLTS and WP interventions may induce some households who otherwise would not upgrade their hygienic practices to adopt handwashing or latrine ownership in addition to switching to using improved WP. For households that adopt all three outcomes, the CLTS intervention alone is sufficient (conditional that they have access to an improved WP).

6 Discussion and Conclusion

In this paper, we focus on a community-based intervention to promote safe sanitation, hand hygiene and water consumption. We find that the interventions in Mozambique were moderately successful at getting households to build traditional pit latrines and stop open defecation. Households that own a latrine also use and maintain (clean, empty and rebuild) them. The effect of the intervention is in the range reported by Garn et al. (2017) and Gertler et al. (2015) for similar sanitation interventions. However, only 5% of the households in the intervention group (and 4% in the comparison group) own latrines that satisfy all the CLTS requirements for safe sanitation.[33] This suggests

[33] The requirements for safe sanitation are in line with the UNICEF and WHO Joint Monitoring Programme. These include a durable slab (cement or another local material), a lid that properly closes the pit, a superstructure that provides privacy, and the presence of a hand-washing facility with soap or ash (IOB, 2011).

a need for a gradual approach to safe sanitation (Alzúa et al., 2020), and further visits by the implementing NGOs to Open Defecation Free communities with traditional pit latrine facilities.

Our data also show that open defecation has not completely been eliminated in communities that were certified to be Open Defecation Free: only 72.5% of the households owned a latrine in 2010 in communities that became ODF prior to 2010, and 61.8% in 2013 in communities that became ODF prior to 2013. It is possible that these communities were incorrectly certified ODF and not all households had a latrine. Alternatively, some households may have slipped back to practicing open defecation. Both explanations suggest that it is a long process to achieve total sanitation. Follow-up visits are necessary even after the ODF certification to help motivate communities to stay on track for achieving safe sanitation at 100% of the households. Alzúa et al. (2020) demonstrates that frequent follow-up visits before ODF certification can deliver higher adoption rate for latrine use (treatment effect of 33 percentage points). However, their study also found similar concerns about safe sanitation facilities and the practice of open defecation in certified ODF communities.

What distinguishes the program in Mozambique from the other CLTS studies is the possibility to study complementarities between community-based sanitation and water supply interventions. We find that households were more likely to use the installed improved water points after the CLTS intervention, and that the CLTS intervention was more effective in terms of latrine ownership and handwashing when it was implemented after a water point intervention. This suggests that households may be more willing to improve their sanitation situation when they have access to safe drinking water,[34] or adopt these safer practices when the same organization is involved in promoting them.

An important objective of combining safe water, sanitation and hygiene interventions is to reduce childhood diarrhea and stunting. This is also the case for the One Million Initiative, which was coordinated by UNICEF Mozambique. There are a

[34] Duflo et al. (2015) suggests that households in India are willing to invest into safe sanitation in order to get access to safe in-house tapped water. However, in our study, there was no conditionality involved between access to safe water and sanitation.

few recent studies that indicate that sanitation interventions can generate health gains for young children (Pickering et al., 2015, Duflo et al., 2015, Gertler et al., 2015), however, as of yet there is no consensus in the public health literature whether it is indeed the case (Schmidt, 2014, Cumming et al., 2019). Schmidt (2014) argues that it could be a design issue: "a good sanitation marketing campaign may require 5-10 years to achieve a marked increase in latrine coverage with the potential to impact on health" (p.3). This period is much longer than the length of recent randomized controlled trial studies.

This makes the case for turning to carefully designed evaluations of at-scale running programs. This study could be a pointer in this direction as it followed households for up to 5 years after the interventions. However, a number of details could have been designed more carefully. Specifically, the survey could have been better targeted at measuring childhood diarrhea and stunting by using the standard indicators for these[35] (see for example Pickering et al., 2015).

There are a few other sub-optimal circumstances regarding our study. The intervention in Mozambique took place in an environment where latrine ownership and the use of traditional latrines was on the rise in underdeveloped rural villages as the government aimed to achieve its ambitious Millennium Development Goals on access to safe water and sanitation facilities. The One Million Initiative was potentially only one of the programs in the surveyed villages advocating the use of latrines. Hence, in this respect our treatment effects are a lower bound of the program effects in the absence of other sanitation interventions.

A more serious concern is that the CLTS interventions were not randomly allocated among the surveyed villages. To tackle the issue of selection bias, we use a novel evaluation method following Vigh and Elbers (2017) to identify the average treatment

[35] In the surveys, we collected information on "the number of days affected by water-borne diseases" in the last 6 months and last 2 weeks for all household members, young and old (see WRD6m in Table A3 for the 6 months recall period). This formulation has been used in other WASH evaluations as well. However, the prevalence of water-borne diseases was greatly under-reported by the surveyed households compared to DHS figures on the prevalence of diarrhea in the last 2 weeks. This could have happened because the surveys were carried out in the dry season, when diarrhea is less common, or because the households were not familiar with the term "water-borne diseases."

effect of the CLTS interventions of the One Million Initiative. The method rests on the identifying assumption that the CLTS interventions were rolled out in the order of the expected treatment effect on latrine ownership based on the expectations of the NGOs, who were familiar with the program area. As we argued in the paper, this is a likely scenario in the case of the One Million Initiative.

Our results indicate that there is substantial heterogeneity in the treatment effect of CLTS, and that universally implementing the CLTS intervention in the program area would not result in a significant increase in latrine ownership above the 9 percentage points general increase in latrine ownership between 2008 and 2013. However, CLTS did increase latrine ownership by 8pp among the program beneficiaries (ATT), and the effects were higher in communities that also received the WP intervention (13pp).

Comparing the estimate of the average treatment effect in the population to that on the program beneficiaries suggests that allowing the NGOs to target CLTS at communities that were more likely to benefit from it increased the effectiveness of the program. Identifying the key characteristics of communities that make CLTS successful is beyond the scope of this paper (see Kar and Chambers, 2008, for recommendations).

At the same time, we find that CLTS had a persistent positive effect on handwashing with soap or ash after defecation (11pp) and on the use of improved WPs (15pp conditional on access). Also these effects were higher when CTLS and the WP intervention were implemented together (24pp and 49pp, respectively).

One implication of our findings is that combining community-based water supply and sanitation interventions can increase their joint effectiveness on using safe water sources and using a latrine facility. The One Million Initiative was implemented in poor rural communities often without access to any safe water sources prior to the interventions, and where a large share of population practiced open defecation. Whether water and sanitation interventions reinforce each other in communities with more developed water and sanitation infrastructure is a question for further research. However, it is important to note that precisely this synergy effect was the underlying motivation for integrating the water supply and sanitation interventions in the One Million Initiative and also in other similar programs.

Ultimately, the final objective of water and sanitation interventions is to reduce the transmission pathways of fecal contamination, and thereby improve the health of young children and other household members. Kremer et al. (2011) reports that after protecting spring water sources in Kenya, diarrhea among young children fell by nearly 25 percent as a result of improving the water quality at the source. Although not reported in this paper, we also find similar effects for the installed boreholes. This suggests that sanitation interventions without access to clean water may be less effective in improving child health.

References

Alzúa, M. L., Djebbari, H., and Pickering, A. J. (2020). A community-based program promotes sanitation. *Economic Development and Cultural Change*, 68(2):357–390.

Cameron, L., Olivia, S., and Shah, M. (2019). Scaling up sanitation: evidence from an rct in indonesia. *Journal of development economics*, 138:1–16.

Crocker, J., Saywell, D., and Bartram, J. (2017). Sustainability of community-led total sanitation outcomes: Evidence from ethiopia and ghana. *International journal of hygiene and environmental health*, 220(3):551–557.

Cumming, O., Arnold, B. F., Ban, R., Clasen, T., Mills, J. E., Freeman, M. C., Gordon, B., Guiteras, R., Howard, G., and Hunter, P. R. (2019). The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. *BMC medicine*, 17(1):1–9.

Duflo, E., Greenstone, M., Guiteras, R., and Clasen, T. (2015). Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation. *NBER Working Paper*, (Working Paper 21521).

Elbers, C. and Gunning, J. W. (2014). Evaluation of development programs: randomized controlled trials or regressions? *the world bank economic review*, 28(3):432–445.

Garn, J. V., Sclar, G. D., Freeman, M. C., Penakalapati, G., Alexander, K. T., Brooks, P., Rehfuess, E. A., Boisson, S., Medlicott, K. O., and Clasen, T. F. (2017). The impact of sanitation interventions on latrine coverage and latrine use: A systemati.c review and meta-analysis. *International journal of hygiene and environmental health*, 220(2):329–340.

Gertler, P., Shah, M., Alzúa, M. L., Cameron, L., Martinez, S., and Patil, S. (2015). How does health promotion work? evidence from the dirty business of eliminating open defecation.

Godfrey, A. (2009). Preliminary Documentation and Evaluation of the Sanitation Component of the 'One Million Initiative' in Mozambique. Technical report.

Guiteras, R., Levinsohn, J., and Mobarak, A. M. (2015). Encouraging sanitation investment in the developing world: a cluster-randomized trial. *Science*, 348(6237):903–906.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Hirano, K. and Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology*, 2(3):259–278.

INE (1997). Mozambique Population and Housing Census. Technical report, National Statistics Institute, Mozambique.

INE (2007). Mozambique Population and Housing Census. Technical report, National Statistics Institute, Mozambique.

IOB (2011). More than Water. Impact Evaluation of drinking water supply and sanitation interventions in rural Mozambique. Technical report, Policy and Operations Evaluation Department of the Netherlands Ministry of Foreign Affairs.

Kar, K. and Chambers, R. (2008). *Handbook on Community-Led Total Sanitation*.

Kremer, M., Leino, J., Miguel, E., and Zwane, A. P. (2011). Spring cleaning: Rural water impacts, valuation, and property rights institutions. *The Quarterly Journal of Economics*, 126(1):145–205.

NSO, M. and International, M. (2010). *Malawi Demographic and Health Survey*. National Statistical Office, Malawi.

Patil, S. R., Arnold, B. F., Salvatore, A. L., Briceno, B., Ganguly, S., Colford Jr., J. M., and 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).

Pattanayak, S. K., Yang, J.-C., Dickinson, K. L., Poulos, C., Patil, S. R., Mallick, R. K., Blitstein, J. L., and Praharaj, P. (2009). Shame or subsidy revisited: social mobilization for sanitation in Orissa, India. *Bulletin of the World Health Organization*, 87(8):580–587.

Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., and Alzúa, M. L. (2015). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. *The Lancet Global Health*, 3(11):701–711.

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.

UNICEF (2013). Evaluation of the WASH Sector Strategy 'Community Approaches to Total Sanitation' (CATS). Mozambique - Country visit report. Technical report.

UNICEF (2014a). Evaluation of the WASH Sector Strategy 'Community Approaches to Total Sanitation' (CATS) Final Evaluation Report. Technical report.

UNICEF (2014b). One Million Inititiative. Final report. Technical report. SC060665.

Vigh, M. and Elbers, C. (2017). Picking winners: Measuring the effectiveness of selectively placed policy interventions. *TI Discussion Paper Series*, 17(110/V).

Waddington, H. and Snilstveit, B. (2009). Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *Journal of development effectiveness*, 1(3):295–335.

WE Consult (2009). Baseline survey final report: a WASH baseline survey under the NL-UNICEF partnership for WASH. Technical report.

Whaley, L. and Webster, J. (2011). The effectiveness and sustainability of two demand-driven sanitation and hygiene approaches in zimbabwe. *Journal of Water Sanitation and Hygiene for Development*, 1(1):20–36.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1):137–150.

Appendix

A Descriptive statistics

Table A1: Variable definitions

- Latrine ownership: the HH owns any type of functioning latrine/toilet facility (verified by interviewers).
- Latrine use: adult males and females are both reported to use a latrine/toilet facility (self-reported).
- Open defecation: any adult males and females are reported to practice open defecation (self-reported).
- Handwashing: adult males and females are both reported to wash their hands with soap or ash after defecation (self-reported).
- Improved WP: the HH takes their drinking water from an improved water point (borehole, protected spring or well).
- Household size: number of people living in the household (HH).
- Children < 5 years: number of children younger than 5 years living in the HH.
- Wealth index: calculated using PCA of items owned by the HH, normalized to mean=0 and sd=1 in the full sample in 2008.
- Education (age>14): share of HH members (age>14) with any formal education.
- WRD6m: share of HHs that had any member affected by water related diseases in the past 6 months.
- WRD6m (under 5): share of HHs that had any member under age 5 affected by water related diseases in the past 6 months.

Table A2: Household characteristics in the regression sample

Variable	Year	N obs	Control	CLTS	WPI	CLTS& WPI	*p* CLTS	*p* WPI	*p* CLTS& WPI
Latrine ownership	2008	1387	0.454	0.549	0.405	0.440	0.024	0.171	0.703
	2010	1507	0.509	0.702	0.463	0.615	0.000	0.178	0.002
	2013	1312	0.601	0.675	0.587	0.653	0.076	0.711	0.152
Latrine use	2008	1387	0.501	0.580	0.423	0.488	0.062	0.027	0.715
	2010	1503	0.579	0.750	0.493	0.662	0.000	0.012	0.016
	2013	1306	0.649	0.726	0.640	0.691	0.054	0.798	0.237
Open defecation	2008	1387	0.422	0.363	0.516	0.472	0.150	0.008	0.169
	2010	1507	0.377	0.222	0.467	0.321	0.000	0.007	0.098
	2013	1306	0.300	0.250	0.279	0.258	0.201	0.533	0.223
Handwashing	2008	1351	0.198	0.193	0.174	0.155	0.870	0.385	0.124
	2010	1507	0.333	0.423	0.368	0.434	0.020	0.281	0.003
	2013	1312	0.489	0.651	0.562	0.650	0.000	0.044	0.000
Improved WP	2008	1387	0.201	0.177	0.098	0.096	0.478	0.000	0.000
	2010	1507	0.243	0.395	0.539	0.515	0.000	0.000	0.000
	2013	1312	0.277	0.571	0.658	0.596	0.000	0.000	0.000

Notes: Intervention arm is defined by the first intervention implemented in community if the timing of interventions differ. *p* CLTS reports the p-value for the difference of the variable mean in the intervention arms vs. the control group.

Table A3: Household characteristics in the regression sample

Variable	Year	N obs	Control	CLTS	WPI	CLTS& WPI	*p* CLTS	*p* WPI	*p* CLTS& WPI
Household size	2008	1387	5.549	5.447	5.700	6.088	0.640	0.433	0.007
	2010	1507	5.268	5.028	5.393	5.380	0.196	0.461	0.501
	2013	1312	5.399	5.151	5.929	6.122	0.230	0.010	0.000
Children <5 years	2008	1387	0.894	0.881	0.875	0.949	0.857	0.763	0.414
	2010	1507	0.895	0.746	0.945	0.890	0.046	0.462	0.939
	2013	1312	0.750	0.646	0.922	0.807	0.151	0.013	0.396
Wealth index	2008	1387	-0.072	-0.093	0.122	0.045	0.807	0.005	0.091
	2010	1507	-0.113	0.042	0.213	0.075	0.075	0.000	0.013
	2013	1312	0.066	-0.018	0.366	0.161	0.419	0.001	0.291
Formal education (age>14)	2008	1387	0.451	0.436	0.484	0.541	0.643	0.229	0.001
	2010	1507	0.464	0.538	0.551	0.558	0.018	0.001	0.000
	2013	1312	0.587	0.605	0.632	0.661	0.586	0.104	0.009
WRD6m	2008	1387	0.319	0.350	0.310	0.243	0.444	0.771	0.019
	2010	1507	0.204	0.137	0.103	0.125	0.029	0.000	0.002
	2013	1312	0.245	0.184	0.170	0.125	0.091	0.010	0.000
WRD6m (under 5)	2008	830	0.219	0.250	0.226	0.153	0.504	0.844	0.071
	2010	851	0.169	0.079	0.057	0.070	0.018	0.000	0.001
	2013	693	0.197	0.146	0.134	0.078	0.289	0.083	0.001

Notes: Intervention arm is defined by the first intervention implemented in community if the timing of interventions differ. *p* CLTS reports the p-value for the difference of the variable mean in the intervention arms vs. the control group.

Table A4: Correlation between outcome variables over time

	HW'08	L'08	IWP'08	HW'10	L'10	IWP'10	HW'13	L'13	IWP'13
HW 2008	1	0.013	0.035	0.129	0.0004	0.018	0.034	0.031	0.010
L 2008	0.013	1	0.043	0.007	0.549	0.019	0.083	0.502	- 0.007
IWP 2008	0.035	0.043	1	0.028	0.024	0.341	0.015	0.025	0.310
HW 2010	0.129	0.007	0.028	1	0.040	0.031	0.109	- 0.016	0.035
L 2010	0.0004	0.549	0.024	0.040	1	0.043	0.112	0.506	- 0.026
IWP 2010	0.018	0.019	0.341	0.031	0.043	1	0.016	0.077	0.532
HW 2013	0.034	0.083	0.015	0.109	0.112	0.016	1	0.149	0.007
L 2013	0.031	0.502	0.025	- 0.016	0.506	0.077	0.149	1	0.058
IWP 2013	0.010	- 0.007	0.310	0.035	- 0.026	0.532	0.007	0.058	1

Notes: HW = Handwashing with soap after defecation, L = Latrine ownership, IWP = Use of improved WP

Table A5: Transition dynamics: probability of being in the same state in the following period (column) conditional on the state now (row)

(a) Ownership of latrine

	2010	2013
2008	0.860	0.885
2010		0.834

(b) Non-ownership of latrine

	2010	2013
2008	0.688	0.601
2010		0.662

(c) Handwashing

	2010	2013
2008	0.521	0.626
2010		0.647

(d) No handwashing

	2010	2013
2008	0.642	0.417
2010		0.464

(e) Use of improved WP

	2010	2013
2008	0.853	0.918
2010		0.829

(f) Non-use of improved WP

	2010	2013
2008	0.636	0.535
2010		0.708

Table A6: Transition dynamics: probability of being in the same state in the following period (column) conditional on the state now (row)

(a) CLTS: Ownership of latrine

	2010	2013
2008	0.878	0.890
2010		0.815

(b) Control: Ownership of latrine

	2010	2013
2008	0.840	0.874
2010		0.843

(c) CLTS: Handwashing

	2010	2013
2008	0.630	0.728
2010		0.680

(d) Control: Handwashing

	2010	2013
2008	0.361	0.381
2010		0.562

Notes: CLTS group contains all communities that received the CLTS intervention during the evaluation period. Control group contains communities that received no program interventions.

Table A7: Transition dynamics: probability of being in the same state in the following period (column) conditional on the state now (row)

(a) CLTS: Non-ownership of latrine

	2010	2013
2008	0.558	0.544
2010		0.630

(b) Control: Non-ownership of latrine

	2010	2013
2008	0.762	0.663
2010		0.687

(c) CLTS: No handwashing

	2010	2013
2008	0.600	0.346
2010		0.369

(d) Control: No handwashing

	2010	2013
2008	0.676	0.472
2010		0.552

Notes: CLTS group contains all communities that received the CLTS intervention during the evaluation period. Control group contains communities that received no program interventions.

B Detailed program description

B.1 Overview

The One Million Initiative (OMI) was implemented in 18 districts of Manica, Sofala and Tete provinces in Mozambique between 2006 and 2013. Figure B1 shows the intervention areas of OMI. The program was a partnership between the Government of the Netherlands, the Government of Mozambique and UNICEF that was designed to contribute to the achievement of the Millennium Development Goals with respect to 70 percent of people having access to safe drinking water and 60 percent to adequate sanitation. Specifically, The objectives of OMI included that one million people in poor rural areas use safe and sustainable drinking water and use hygienic sanitation facilities (UNICEF, 2014b), which amounts to 40 percent of the population in the OMI intervention area (INE, 2007).[36]

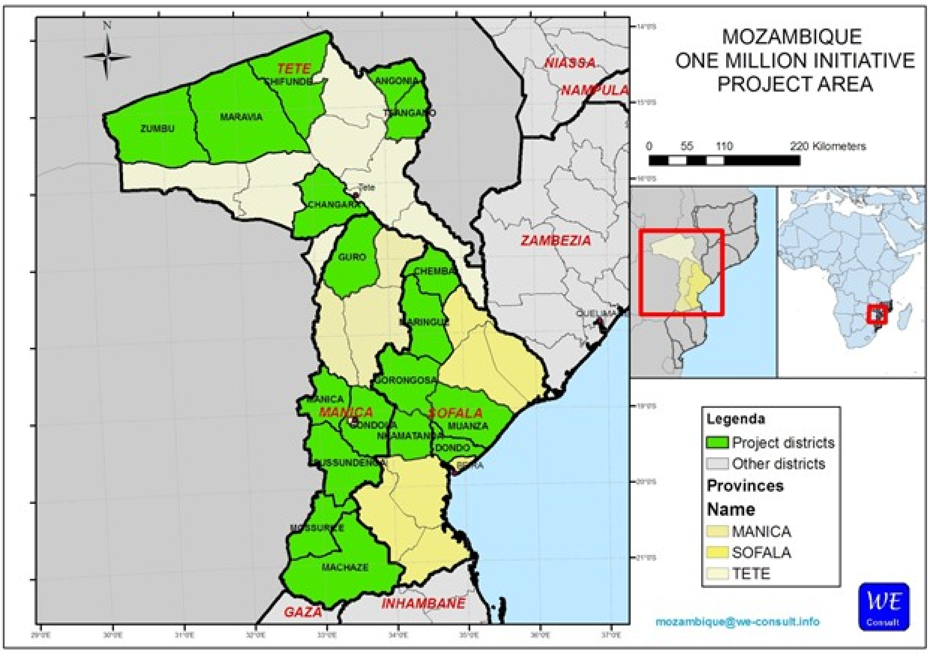
In order to achieve these goals, OMI carried out community water supply interventions, either by rehabilitating non-functional boreholes (2007-2009) or by creating new ones (2008-2013) coupled with water committee training;[37] and community-based sanitation and hygiene education through the Community Approaches to Total Sanitation (CATS), which combined the CLTS approach with an award scheme for ODF communities (2008-2013).

The training activities and planning of the interventions were carried out in the framework of the Community Participation and Education (PEC Zonal) program,

[36] The specific objectives of OMI were: (i) 1,000,000 people in poor rural areas use safe and sustainable drinking water; (ii) 200,000 people in rural areas have access to rehabilitated drinking water sources; (iii) 1,000,000 people in rural areas use hygienic sanitation facilities; (iv) 140,000 learners in 400 primary schools in rural areas use safe school water and sanitation facilities and adopt appropriate hygiene practices; (v) 1,200,000 people in rural areas adopt appropriate hygiene practices; (vi) 114 government staff members in 18 districts and three provinces have technical and managerial capacities for WASH planning, co-ordination and implementation; (vii) 4,572 people have relevant WASH skills and knowledge (planning, procurement, contract management, hygiene promotion, etc.); (viii) 21 institutions in 18 districts and three provinces have extra equipment (UNICEF, 2014b).

[37] OMI's focus on water supply intervention is motivated by the widely available groundwater resources in the intervention area. Investigations suggest that arsenic levels in groundwater are not a serious concern, but groundwater salinity is a problem in some areas (IOB (2011), p.34-35).

Figure B1: Districts covered by the One Million Initiative



Source: WE Consult (2009)

which has three pillars: community mobilization and hygiene training, water committee training and sanitation training. These activities were implemented by PEC activists (PEC NGOs) as described below.

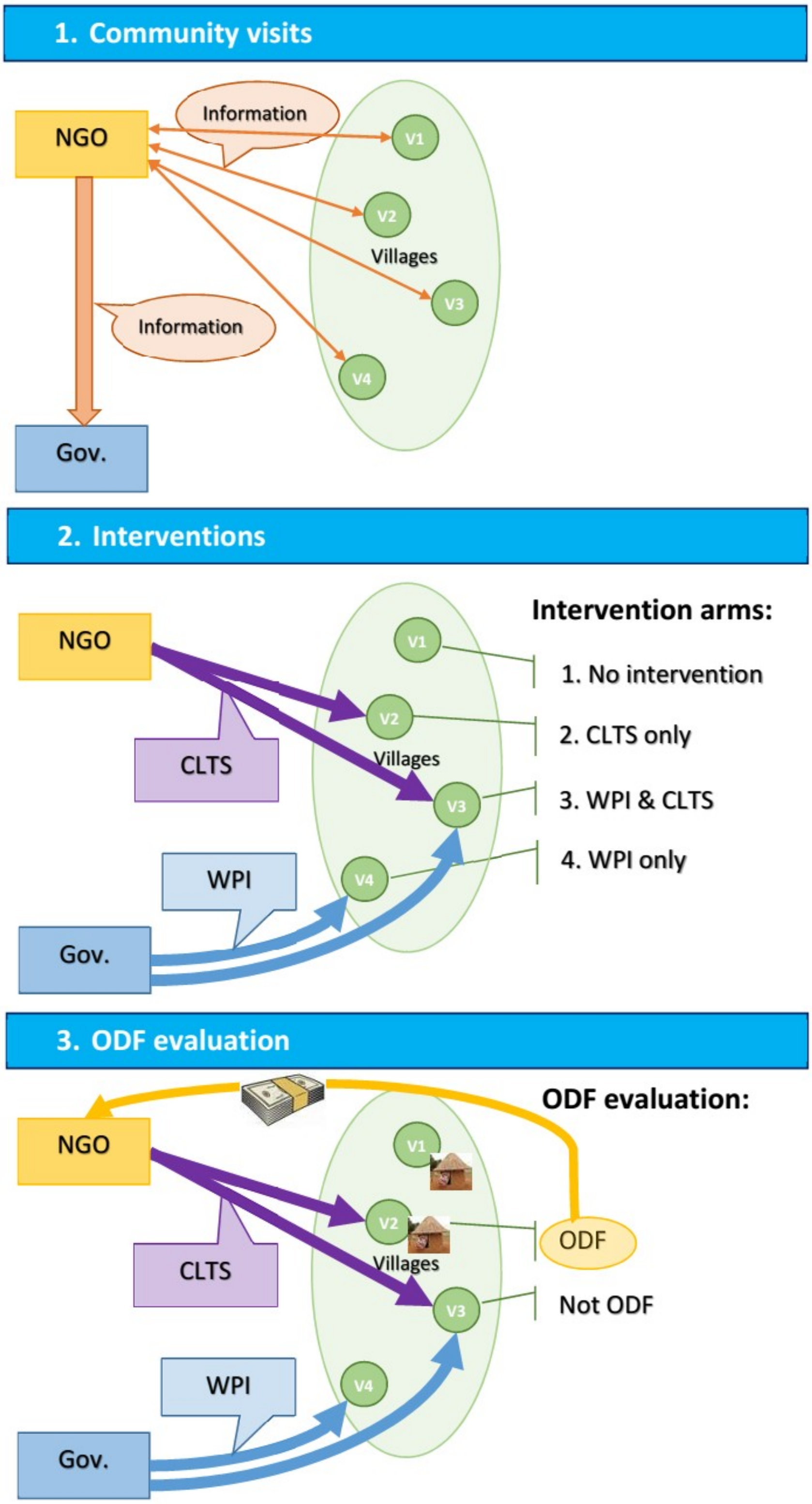
Overall, between 2008 and 2013, the One Million Imitative built 1,689 new water points reaching an estimated 1.16 million users, and it triggered 1,938 communities of which 890 were declared ODF during one of the ODF evaluation campaigns with a population of 0.7 million people. A schematic description of the program is displayed in Figure B2.

B.2 Community mobilization

The first task of the PEC NGO in a community was to mobilize the community towards a more hygienic attitude in their daily activities and to improve the organization of their water management. A typical scenario for the PEC NGO's visits to a community is the following:

Visit 1: PEC activists introduce themselves to the community leader and ask for a

Figure B2: Simplified overview of the One Million Initiative interventions in communities



meeting with the key community members (within the next couple of days usually).

Visit 2: PEC activists explain to the main stakeholders in the community the objective of their visit, and ask them to call a meeting with all the members of the community (within the next week).

Visit 3: PEC activists explain at the community meeting the importance of hygienic behavior, in particular proper hand washing at critical times, and of having a functioning water committee to manage the maintenance of their improved water sources. Emphasis is put on the commitment of the water committee to keep the water source and surroundings clean and organize the replacement of broken parts. It is stressed that the users should pay for the water, so that money is available to fix the water point when something breaks down. If there is no improved water source in the community or it is not sufficient to cater to the needs of all households, they encourage them to apply for a new water point at the district government. They discuss the procedure for application and the placement of such a new water point as well. If there is an improved water source in the community which is not functioning, the PEC activists encourage the community to apply for the rehabilitation of the water point. Finally, they ask the community to discuss the issues among themselves and decide whether they want to set up a water committee and apply for a new water point. Before parting, they make an appointment for the next visit.

Visit 4: The community gives their response whether they set up a water committee, and give the names of the members, which should include an equal share of women. At the meeting only the key members, in particular the water committee members, are present. They also indicate whether they agreed on a payment for water usage. At this point they can fill out the application form for the new water point (or rehabilitation of a water point). In the form they have to indicate the names of the water committee members and where they would like the new water point.

In further visits PEC activists meet only with the community members involved in the issues of hygiene and water management if applicable. If there is an improved water point already in the community, then PEC activists do a refreshment training on the management of the water point for the members of the new water committee (this is already pillar 2). A new community meeting is held only when the sanitation component (pillar 3) is introduced in the village.

The hygiene awareness pillar of PEC was already introduced before the baseline in most communities. As a result, a number of communities had a water committee already at the time of the baseline. The concept of water committees was introduced by the PEC, therefore it is not likely that a community had a water committee before the arrival of the PEC activists to the community.[38]

B.3 Water point interventions

OMI implemented two types of water point interventions: the rehabilitation of a broken improved water point, and the construction of new boreholes fitted with hand pump.[39] In both of these cases the community visits of the PEC NGOs were crucial. Through the application procedure for water points they help to identify the communities where new water points are needed and where the community is also organized to take care of them once they are constructed. From the list of applicants, the district authorities selected the locations for the new boreholes given the resources and focus of the year.[40] The boreholes were constructed or rehabilitated by engineering and drilling companies contracted by the provincial Departments of Public Works and Housing.

As a part of pillar 2 of PEC Zonal, once the new water point was constructed the PEC activists train the water committee members on how to manage the payments for

[38] Note that not all water points have a separate water committee. If there are more improved water points in the community not far away from each other, then a single water committee is responsible for all these water points. On the other hand, if the water points are scattered around the community at a distance, then each of the distant improved water points have a separate water committee.

[39] Most of the rehabilitation of water points were completed by mid 2008. Therefore, we did not include rehabilitated water points into the data analysis.

[40] Priority areas were set yearly by the government. For example, in 2008 the focus was on densely populated areas, while in 2009 it shifted to more remote areas and in 2010 schools and health centers were also targeted.

water and the maintenance of the water point.[41] They help them in finding an artisan who can supply them with spare parts if the pump needs fixing but also train them on easily reparable problems of the pump.

B.4 Sanitation interventions

From 2008 onwards, OMI implemented the Community-Led Total Sanitation (CLTS) approach in the sanitation component of PEC Zonal.[42] CLTS uses a direct and confrontational approach. It uses food and feces to make the people realize how easy it is for flies to contaminate food if human feces are lying around the community. It encourages people to draw up the map of the community with the locations where people go to defecate to make people aware of the problems of sanitation (community map), which are then visited during the 'walk of shame' (transect walk). The key triggering moment of CLTS is the demonstration of the fecal contamination of food (fecal-oral disease transmission). After the identification of the problem, the solution is discussed. The planning of latrines draws on local solutions using local materials. The PEC activists only give advice when asked, mostly about handwashing facilities and the use of slabs. The aim is to construct low-cost local solutions in a short time. The following visits by the PEC activists depend on the success of the triggering. According to UNICEF (2013), most communities reach ODF status within 3 weeks to 3 months after the CLTS triggering. Communities that do not achieve ODF status during this period often do not reach it at all. In such a case, a different approach should be tried in the community.

It is important to note that the communities for the sanitation interventions were

[41] Setting up a water committee was also a precondition for applying for a water point intervention. In addition, communities were usually required to contribute 2,500 MZN (56 EUR) before the construction of a water point began (IOB (2011), p. 49).

[42] Until mid 2008 the PEC NGOs used the PHAST (step-by-step) method instead of CLTS. Upon the introduction of CLTS, a total of 74 people were trained in CLTS including 2 supervisors from each of the 17 NGOs contracted by the districts to implement sanitation and hygiene promotion, representatives from Provincial and District governments from Tete, Manica and Sofala, and some UNICEF staff. The training was a combination of group discussion and hands-on field practice with the triggering of CLTS in 12 communities. (Godfrey, 2009)

mostly picked by the PEC NGOs, perhaps with the assistance of the local authorities. The PEC NGOs decided for themselves in which communities to introduce CLTS by considering the likelihood of success, i.e. that the community makes an organized effort to build latrines (and handwashing facilities) for every household and becomes open defecation free (ODF). The success of a CLTS training depends on factors like health problems, leadership, size of the community (smaller is easier) and geographical factors.[43] It is in their interest of the PEC NGO to act this way because their payment depends on the number of communities that become Open Defecation Free.

Finally, the connection between water point and sanitation interventions should be addressed. As both the water point and sanitation interventions originate in PEC Zonal, there is a large overlap between the water point and sanitation interventions. Until 2010, it was possible for a community to have a sanitation intervention without having an improved water point. However, it was planned to give ODF communities priority for receiving a water point.

B.5 CATS and open defecation free communities

The sanitation and hygiene component of OMI is implemented using UNICEF's Community Approaches to Total Sanitation (CATS), which combine a community-based sanitation program with the goal of eliminating open defecation, a focus on behavior and social change, community demand and leadership, and local innovation. The idea is that external agencies provide guidance rather than subsidies and regulation. Thus, households build toilets based on locally available materials using the skills of local technicians and artisans. CATS focus on building local capacities to enable sustainability. This includes the training of community facilitators and local artisans, and the encouragement of local champions for community-led programs. At the same time, CATS also involves government participation from the outset (at the local and national levels) to ensure the potential for scaling up (UNICEF (2014a), p. 9-10.)

[43] Godfrey (2009) found that 90 percent of the 34 communities that achieved the ODF status in 2008 had less than 300 households.

The subsidies in CATS were replaced by an award system for Open Defecation Free (ODF) communities, which consisted of four stages: application, evaluation, award ceremony and prizes. First, the triggered communities apply to certify their ODF status.[44] These applications are evaluated by multi-agency evaluation teams led by UNICEF.[45] The ODF evaluation involves household interviews and observations. The following ODF criteria are observed: whether all households own and use latrines; whether the latrines satisfy basic requirements (offer privacy, have a lid on the defecation hole, and a roof to protect the slab); whether there are any feces around the compounds; and whether there is a handwashing facility available near the latrine with water, soap or ash and evidence of use. Communities that satisfied all these criteria, received the ODF award. In 2008, prizes were given to the households (hygiene kit), community leader (bicycle), community (water point or classroom), chief of sub-district (mobile phone or radio) and district administrator (photocopier or computer). However, Godfrey (2009) pointed out that these awards were not sustainable and that the "reward system" had effectively become an "incentive scheme". Therefore, in 2009 prizes were only given to the households and community leader, and in 2010 only to the community leader (IOB (2011), p.52). In 2011, the award system was completely abandoned (UNICEF (2013), p.13).

UNICEF (2013) notes the importance of post-ODF follow-up visits to the communities in order to prevent reverting back to open defecation. The traditional community leader also plays a crucial role in keeping community members committed to ODF. However, households may get tired of the frequent need to rebuild their low-quality latrines and revert back to open defecation. For this reason, OMI introduced the concept of "CLTS+" and "ODF+" in 2011, which promote the use of more durable materials (such as cement slabs or improved roof materials). However, CLTS+ was only targeted at communities where the financial capacity is considered sufficient for households to respond.

[44] Note that only communities that were triggered using CLTS were allowed to participate in the ODF evaluation and award system. Hence, communities with PHAST were excluded.

[45] In 2008, the evaluation teams included staff from the Ministry of Health, Ministry of Education, DNA, provincial government and personnel from WSP, UNICEF, SNV and OXFAM (Godfrey (2009), p.10).

C Sample distribution of the interventions

Table C1 shows the cumulative percentage of IT and IC communities that received the CLTS, WP and any of the two interventions. As expected, the Intended Target communities were more likely to receive the interventions, particularly between the baseline (2008) and midline (2010) surveys (Wave 1). Most of the WP interventions took place during this period. The CLTS interventions were more evenly spread out over the evaluation period (Wave 1 and 2), but also here we observe that the IT communities received the interventions earlier. Overall, 95% of the IT and 50% of the IC communities received at least one of the program interventions.

Table C2 summarizes the realized intervention history of the CLTS and water point (WPI) interventions in the surveyed communities. The first column of the table shows the number of communities that did not receive the CLTS intervention during the evaluation period, while the second and third columns show the number of communities that received the CLTS intervention between the baseline and midline (Wave 1) and between the midline and endline surveys (Wave 2), respectively. Similarly, the first row of the table shows the number of communities where no WP intervention was carried out, and the second and third rows show the number of communities receiving the WP intervention between the baseline and midline (Wave 1) and between the midline and endline surveys (Wave 2), respectively. Overall, 22 of 80 communities did not receive any program interventions and these serve as our comparison group.

For the analysis, we mainly use CLTS and WP interventions as overlapping intervention groups, which we will refer to as 'CLTS (total)' and 'WPI (total)' or the total effects of the interventions. These contain the 41 and 47 communities that received a CLTS and WP intervention, respectively. However, we also report results for the three independent intervention arms: CLTS (alone), WPI (alone) and CLTS&WPI. The CLTS&WPI intervention arm contains those 21 communities that received both the CLTS and WP interventions in the same period (15 between the baseline and midline, and 6 between the midline and endline). The CLTS only intervention arm contains those 20 communities that received the CLTS intervention without the WP

Table C1: Program exposure in the Intended Target (IT) and Intended Control (IC) communities

	IC	IT	p-value
WP intervention			
2008	0.0%	0.0%	1.000
2010	17.5%	70.0%	0.000
2013	32.5%	85.0%	0.000
CLTS intervention			
2008	0.0%	0.0%	1.000
2010	15.0%	42.5%	0.001
2013	40.0%	62.5%	0.009
Any intervention			
2008	0.0%	0.0%	1.000
2010	27.5%	80.0%	0.000
2013	50.0%	95.0%	0.000

Table C2: Period and overlap of the CLTS and WP interventions in the surveyed communities

	No CLTS	Wave 1 CLTS	Wave 2 CLTS	Total
No WPI	22	6	5	33
Wave 1 WPI	13	15	7	35
Wave 2 WPI	4	2	6	12
Total	39	23	18	80

Notes: WPI: water point intervention.
Wave 1 interventions were implemented between 2008-2010 (before midline survey). Wave 2 interventions were implemented between 2010-2013 (after midline survey).

intervention (6+5 communities) or the WP intervention was implemented in a different period (7+2 communities). Similarly, the WPI only intervention arm contains those 26 communities that received the WP intervention without the CLTS intervention (13+4 communities) or the CLTS intervention was implemented in a different period (7+2 communities).

Hence, 9 communities appear both in the CLTS and WPI intervention arms. Because in these communities one of the interventions happened before the midline and the other after the midline survey, we are able to measure the effect of both interventions separately (unlike for the communities in the CLTS&WPI intervention arm).

D Selection of the intervention communities

Given that the program interventions were not randomly assigned, it is important to have a better understanding about the selection of the intervention communities. We first look at the distribution of the main outcome variables in selected program areas (districts and communities) compared to the general population using data from the Demographic and Health Surveys (DHS) in Mozambique in 2009 and 2011. Then we investigate whether the intervention placement can be predicted using observed community characteristics collected at the baseline survey.

In order to investigate the selection of the program districts, Tables D1 and D2 show the distribution of the outcome variables within the program provinces (Manica, Sofala and Tete) using data from the DHS in 2009 and 2011, respectively. These data were collected when the program was already under way. Nonetheless, we observe that the use of improved WPs was on average lower in the surveyed program districts (Table D1b) compared to the province averages Table D1a) in 2009 (with the exception of Manica province), and the difference between the surveyed program districts compared to other districts is statistically significant (Table D1c). This finding suggests that the program indeed targeted locations in need of improved WPs.

Looking at the distribution of the use of toilet facilities (mostly traditional latrines) in Tables D1a and D1b, we observe a large variation between and within the provinces. There is no indication that the program selected districts based on latrine use. However, it is important to note that the more prevalent use of latrines in the survey districts in Tete province are due to districts located at the border to Malawi, where basic sanitation facilities are more widespread.[46] Data on handwashing is only available in DHS 2011, and this variable looks at the presence of a handwashing facility in the dwelling not at handwashing occasions. The figures in Tables D2a and D2b indicate that there is a large variation in the presence of a handwashing facility between the provinces.

Comparing the figures in Tables D1b and D3 for the 2009 DHS sub-sample and the 2008 program survey data, we can compare the distribution of the main outcome

[46] Open defecation was estimated at 11.3% in the rural population in 2010 by NSO and International (2010).

Table D1: Differences in water and sanitation outcomes between the sampled program districts and non-program districts in the program provinces (using DHS 2009 data)

(a) Water and sanitation outcomes in the program provinces

	Manica	Sofala	Tete	Total
Use of toilet facility (2009)	50.7	31.7	37.0	39.4
Use of improved WPs (2009)	33.4	59.5	42.5	44.5

(b) Water and sanitation outcomes in the sampled program districts

	Manica	Sofala	Tete	Total
Use of toilet facility (2009)	39.2	24.0	61.4	50.4
Use of improved WPs (2009)	34.1	12.6	31.1	30.3

(c) Difference in outcomes in sampled program districts compared to other districts (p-value)

	Manica	Sofala	Tete	Total
Use of toilet facility (2009)	0.0002	0.241	0	0
Use of improved WPs (2009)	0.813	0	0.000003	0

Notes: Based on own calculation using data from DHS Mozambique 2009. Sample includes rural clusters only. 1025 observations used. Sampling weights are provided by DHS 2009.

Table D2: Differences in water and sanitation outcomes between the sampled program districts and non-program districts in the program provinces (using DHS 2011 data)

(a) Water and sanitation outcomes in the program provinces

	Manica	Sofala	Tete	Total
Use of toilet facility (2011)	62.7	18.4	53.5	47.7
Use of improved WPs (2011)	77.7	43.9	37.1	47.8
HW facility at dwelling (2011)	23.9	28.8	52.5	40.7

(b) Water and sanitation outcomes in the sampled program districts

	Manica	Sofala	Tete	Total
Use of toilet facility (2011)	43.6	23.4	71.3	58.9
Use of improved WPs (2011)	72.4	30.1	41.6	45.8
HW facility at dwelling (2011)	6.2	30.9	56.6	43.1

(c) Difference in outcomes in sampled program districts compared to other districts (p-value)

	Manica	Sofala	Tete	Total
Use of toilet facility (2011)	0	0.043	0	0
Use of improved WPs (2011)	0.011	0.00001	0.010	0.114
HW facility at dwelling (2011)	0	0.469	0.020	0.055

Notes: Based on own calculation using data from DHS Mozambique 2011. Sample includes rural clusters only. 2621 observations used. Sampling weights are provided by DHS 2011.

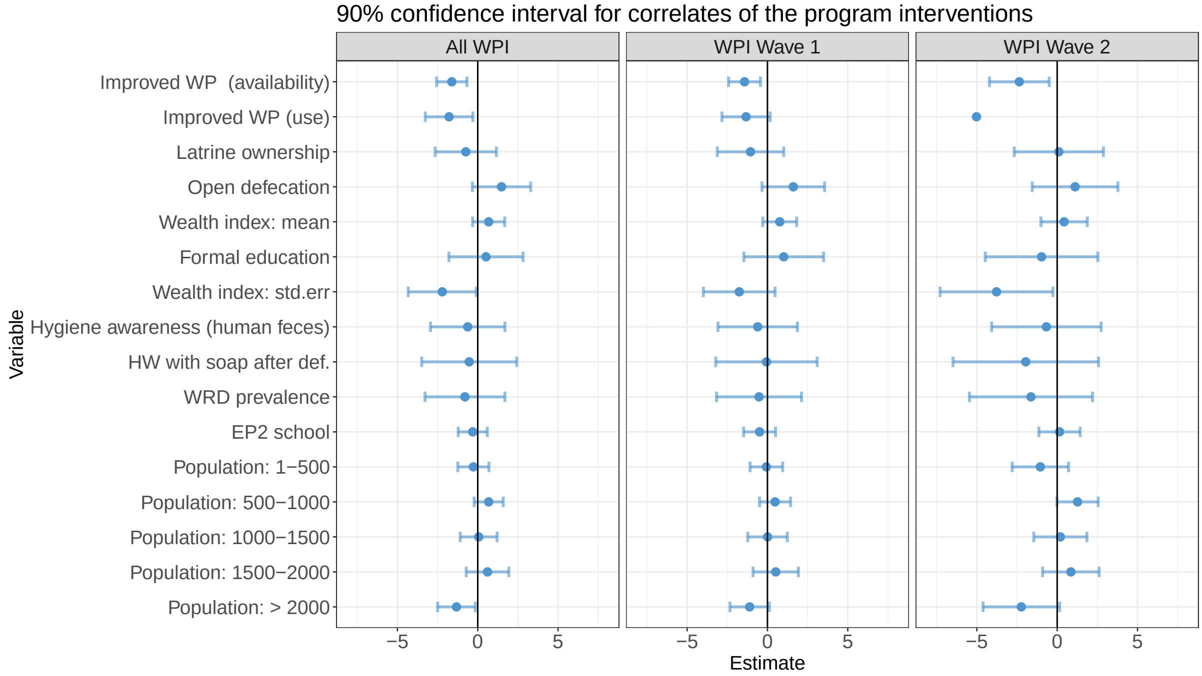
variables in sampled program communities compare to the estimated district means. Looking at the figures on the use of improved WPs in the two tables, we observe that the surveyed program sample over-represents communities not using an improved WP even within the surveyed program districts. Latrine ownership shows similar variation between the provinces in the program survey in 2008 (Table D3) as in the DHS survey (Table D1b). Hence, we can conclude that the surveyed program sample targeted communities in need of an improved WP and not sanitary practices.

Regarding the actual allocation of the interventions in the surveyed communities, we investigate whether observed community characteristics predict the interventions. As discussed before, the CLTS and WP interventions follow different allocation rules. We can safely assume that the WPI interventions were allocated independently from the CLTS interventions, because it was decided by local government officials who had different priorities, and were likely not informed about the (future) location of the sanitation interventions. However, the allocation of the CLTS interventions takes into account the location of the WP intervention.

For the allocation of the WP interventions, we estimate the probability that the communities receive the WP intervention based on baseline community characteristics using a logistic regression. Bivariate multinomial logit regression is used to estimate the factors affecting the timing of the WP intervention. Figure D1 reports the results of the logit (left, for all WP interventions) and multinomial logit regressions (middle and right, for Wave 1 and Wave 2 WP interventions) estimating the probability of the WP intervention arms using baseline community level variables one by one as predictors. The variables are measured relative to their district mean in order to remove variations arising from different characteristics of the districts. This is important because the placement of the interventions are separately decided in each district, and because we observe significant differences in the district means for example for latrine ownership. The figure shows the 90% confidence interval around the point estimates.

The figure shows that the WP interventions were indeed allocated based on need. Communities without an improved WP at baseline were significantly more likely to receive the intervention. Other than availability, villages with less income inequality (in terms of the standard deviation of the wealth index) and medium-small size

Figure D1: Variables predicting the WP interventions (90% confidence interval)



Notes:
1. Coefficient and errorbar estimates from bivariate estimation results from logit for All WP interventions and multinomial logit for Wav
2. Community level variables at the baseline are relative to the district mean.

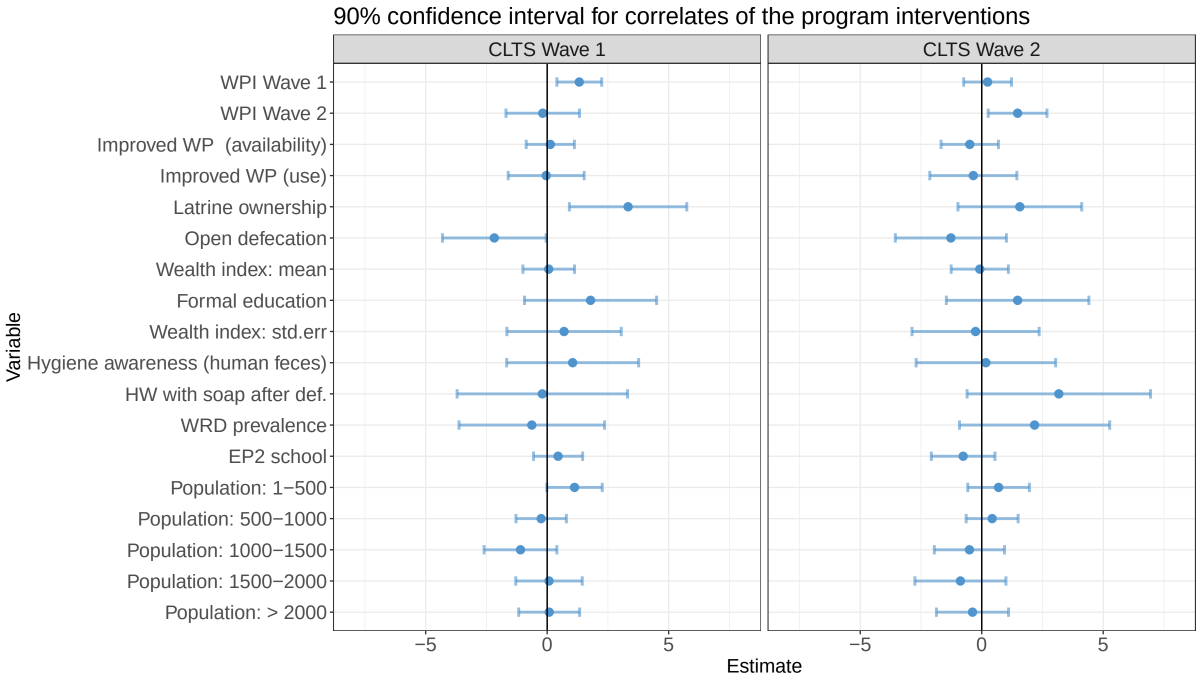
(population of 500-1000) are found to be correlated to receiving the WP intervention. Notably, we find no correlation between the placement of WP intervention and the community mean of other outcome variables (latrine ownership and handwashing practices) and a health measure (water related disease prevalence). Based on the community characteristics in Figure D1, the allocation of the WP interventions followed similar patterns in Wave 1 and Wave 2.

In Figure D2, we carry out a similar analysis for the allocation of CLTS intervention. We estimate the probability of the communities receiving the CLTS intervention based on baseline community characteristics and the timing of the WP intervention. We use a multinomial logit model to estimate the 3 possible intervention arms for CLTS: Wave 1, Wave 2 and no CLTS.[47]

Figure D2 shows that the placement of WP intervention had a significant effect

[47] We omit the logit model on any CLTS because the timing of WP intervention affects the timing of CLTS.

Figure D2: Variables predicting the CLTS interventions (90% confidence interval)



Notes:
1. Coefficient and errorbar estimates from bivariate estimation results from multinomial logit regression.
2. Community level variables at the baseline are relative to the district mean.

on the placement of the CLTS intervention. Moreover, the CLTS intervention closely followed the timing of the WP intervention. As discussed above, only 5 out of 30 communities with both interventions received the CLTS intervention prior to the water point intervention.

We observe some differences in the observable baseline community characteristics between the Wave 1 and Wave 2 CLTS interventions. The Wave 1 interventions were more often targeted at communities that already had higher rates of latrine ownership at the baseline relative to other communities in the their district. These were also more likely to be small villages (population of less than 500 people). Given that building a latrine requires non-negligible investment, it is important to point out that the CLTS interventions are not correlated with the relative wealth level of the communities nor with wealth inequality within the communities.[48]

[48] The findings remain the same in a multivariate regression.

Table D3: Water and sanitation outcomes in the provinces before the program (using program data)

Variable	Manica	Sofala	Tete	Total
Latrine ownership (2008)	29.3	11.9	71.6	43.9
Use of toilet facility (2008)	32.2	13.6	76.9	47.6
Open defecation (2008)	61.1	69.7	22.9	46.3
Handwashing with soap/ash after defectation (2008)	18.8	15.8	17.6	17.6
Use of improved WPs (2008)	22.2	11.1	9.6	14.2
Handwashing facility at dwelling (2010)	17.4	17.2	36.9	25.9

Notes: Percent of households with outcome are reported. Data from the program baseline survey (2008) and midline survey (2010).

E Selectivity of attrition

The attrition between the survey rounds is reported in Table E1. Are household characteristics at the baseline and later realization of the program interventions are correlated with the probability that a household will take part in all 3 rounds of the household survey? In Table E2, we find that the program interventions are not correlated with attrition. Further, larger households are less likely to drop out of the sample, and that there is a lower mobility in the northern province Tete (at the border of Malawi and Zambia). Not controlling for provinces, households owning a latrine at the baseline are predicted to have a higher probability of remaining in the sample which is due to the fact that a much larger proportion of the sample had latrines in Tete (about 70 percent vs. 20-30 percent in the other 2 provinces).

In order to control for the education level of the household, we include the share of adult (older than 14 years) household members that are reported to have had at least primary education. Table E2 shows some evidence that more educated households are less likely to remain in the sample for all three survey rounds.

Unfortunately, we have very little information about the characteristics of the household head. However, if the household head or spouse was interviewed, we know the gender of the household head. Column 4 in Table E2 shows that female headed households did not have different probability of staying in the sample compared to male headed households.

Table E1: Attrition in the surveys

	Total HHs	IT HHs	IC HHs
Sample in ...			
Baseline (2008)	1,600	800	800
Midline (2010)	1,600	800	800
Endline (2013)	1,579	800	779
Interviewed in ...			
2008 and 2010	1,372	682	690
2008 and 2013	1,177	585	592
2010 and 2013	1,297	653	644
All 3 rounds	1,161	580	581
At least 2 rounds	1,524	760	764

Notes: IT=Intended Treatment, IC=Intended Control

Table E2: Regression results on attrition

	Dependent variable:			
	Household is interviewed in all 3 rounds			
	(1)	(2)	(3)	(4)
WPI (alone)	−0.001			
	(0.027)			
CLTS (alone)	−0.006			
	(0.028)			
CLTS&WPI	0.030			
	(0.029)			
Latrine ownership	0.043	0.044*	0.128***	0.053*
	(0.027)	(0.026)	(0.022)	(0.028)
Using soap after def.	0.028	0.027	0.027	0.033
	(0.028)	(0.028)	(0.028)	(0.029)
Using improved WP	−0.008	−0.012	−0.039	−0.021
	(0.034)	(0.034)	(0.033)	(0.036)
Health outcome (WRD 6m)	−0.013	−0.016	−0.016	−0.022
	(0.025)	(0.025)	(0.025)	(0.026)
HH size	0.026***	0.026***	0.023***	0.029***
	(0.004)	(0.004)	(0.004)	(0.005)
Wealth index	0.017	0.016	0.006	0.014
	(0.012)	(0.012)	(0.012)	(0.013)
Formal education	−0.064**	−0.060**	−0.072**	−0.080***
	(0.030)	(0.030)	(0.030)	(0.031)
Female HH head				0.022
				(0.041)
Sofala Province	−0.021	−0.014		−0.009
	(0.033)	(0.032)		(0.035)
Tete Province	0.150***	0.152***		0.148***
	(0.031)	(0.029)		(0.031)
Constant	0.527***	0.527***	0.582***	0.525***
	(0.044)	(0.038)	(0.034)	(0.042)
Adjusted R^2	0.054	0.055	0.035	0.06
Observations	1556	1556	1556	1403

Notes: Standard errors are reported in brackets (* 0.10 ** 0.05 *** 0.01).

F Additional regression results

Table F1: Differential treatment effects for new adopters and the retention of earlier adopters (CLTS alone and CLTS&WPI intervention arms)

	Dependent variable:	
	Latrine ownership	Handwashing
	(1)	(2)
New adopters: CLTS (alone)	0.202**	0.099*
	(0.091)	(0.051)
Retention: CLTS (alone)	0.004	0.106*
	(0.039)	(0.057)
New adopters: CLTS&WPI	0.164*	0.120**
	(0.091)	(0.057)
Retention: CLTS&WPI	0.040	0.096
	(0.034)	(0.072)
Mean dep.var.	0.077	0.202
P-value($H_0 : \tilde{\beta}_0 = \tilde{\beta}_1$)	0.020	0.935
Observations	2,667	2,632
Adjusted R^2	0.287	0.452

Notes: Conditions: 'New adopters': $Y_{i(t-1)} = 0$ and 'Retention': $Y_{i(t-1)} = 1$. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). First difference regression of equation (4) controlling for the past value and changes in wealth index, education and household size, and year. Controls for the WP intervention are excluded. Sample includes HHs participating in at least 2 survey rounds.

Table F2: Treatment effect estimates for the use of latrine facility and open defecation by intervention arms

	Treatment effect estimates					
	LU	LU	LU	OD	OD	OD
	(1)	(2)	(3)	(4)	(5)	(6)
CLTS (alone)	0.077	0.093*	−0.041	−0.082	−0.113*	0.072
	(0.047)	(0.055)	(0.075)	(0.051)	(0.059)	(0.095)
CLTS&WPI	0.082*	0.084	−0.016	−0.096*	−0.091*	0.023
	(0.047)	(0.058)	(0.065)	(0.050)	(0.054)	(0.067)
WPI (alone)	0.042	0.005	0.089**	−0.043	0.007	−0.096*
	(0.029)	(0.030)	(0.044)	(0.035)	(0.037)	(0.049)
Mean dep.var.	0.585	0.082	0.585	0.366	-0.082	0.366
Observations	4196	2655	4196	4200	2661	4200
R^2	0.081	0.057	0.087	0.066	0.047	0.077
Regression model	DD	DD	CRS	DD	DD	CRS
Estimator	FE	FD	FE	FE	FD	FE
P-value($H_0 : \beta_{CLTS} = \beta_{C\&W}$)	0.927	0.896	0.797	0.826	0.766	0.662
P-value($H_0 : \xi_{CLTS} = 0$)			0.019			0.003
P-value($H_0 : \xi_{C\&W} = 0$)			0.190			0.152

Notes: Dependent variables: LU = all adult HH members use a latrine facility, OD = adult HH members practice open defecation. Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Regression model DD estimates equation (3) and CRS estimates equation (2). All regressions control for HH size, wealth index, education and year. Sample includes HHs participating in at least 2 survey rounds.

G Specification tests

G.1 Controlling for observable determinants of the intervention allocation

In the main analysis, we mostly considered unobserved determinants of the selectivity in the intervention allocation. At least part of the selectivity may be driven by observable characteristics from the perspective of the researchers. In that case, weighting the sample by the inverse of the propensity score can be used to control for selection on observables following Hirano and Imbens (2001).

In order to determine which community-specific observable characteristics drive the intervention placement, we first estimate a multinomial regression on the endline intervention status of the communities:

$$P(D^j_{cT} = 1) = Q_c\delta_j + u_c \tag{6}$$

where j denotes the intervention arms (j^{th} element of the vector of intervention arms) and Q_c are community-specific variables at the baseline. Note that (6) is a cross sectional regression estimating the probability of receiving the intervention for each of the intervention arms at the endline period. Hence, we estimate different coefficients for each intervention arm (δ_j). Based on the regression results of (6) we select the variables $Z_c \subseteq Q_c$ such that the coefficient estimate for the variables in Z was significant at least for one of the intervention arms.

However, we calculate the propensity scores using a probit regression on the communities' probability of receiving any program intervention during the evaluation period using Z as control variables:

$$P\left(\sum_j D^j_{cT} \neq 0\right) = Z_c\theta + u_c. \tag{7}$$

When reporting the regression results for the ATT in equation (3) corrected for propensity score weighting, we weight the observations using the inverse of the probability of the community being in the realized value of the intervention[49]: $1/\Phi(Z\hat{\theta})$ for communities with any program interventions, and $1/(1-\Phi(Z\hat{\theta}))$ for communities without any program interventions. In addition, we only keep observations where the fitted values from equation 7 have support in both the intervention and comparison groups.

Column 2 of Tables G1-G2 show that adjusting for observable characteristics affecting the intervention allocation does not significantly change the (unweighted) ATT estimates (reproduced in column 1). If anything, the ATT becomes larger when using propensity score weighting. More importantly, the fact that controlling for selection on observables does not move the treatment effect estimates closer to the ATE for latrine ownership (Table G1) suggests that selection on unobservables was an important factor for this outcome.

G.2 Representative sample

In the main regression analysis, we used the the surveyed program sample without adjusting the distribution for the oversampling of the Wave 1 intervention communities (Intended Target group). Using a sample that is not representative of the population of the program area does not bias the estimates of the ATT. However, for the estimation of the ATE we need to assume that that the observations are representative of the relevant population. Therefore, in column 5 of Tables G1 and G2 we report the estimates of ATE using representative sampling weights obtained by correcting for the communities' probability of being sampled.[50] These corrected estimates of the ATE (column 5) show a similar pattern as the non-weighted estimates of the ATE (reproduced in column 4).

[49] Note that $\Phi(z)$ denotes the cumulative normal density function.

[50] The probability of being sampled was 2.2% for the Intended Target communities and 0.3% for the Intended Control communities.

Table G1: Robustness checks for the treatment effects on latrine ownership

	Treatment effect estimates for latrine ownership					
	(1)	(2)	(3)	(4)	(5)	(6)
CLTS (alone)	0.069	0.081	0.074	−0.059	−0.111	−0.074
	(0.048)	(0.052)	(0.050)	(0.078)	(0.069)	(0.078)
CLTS&WPI	0.115**	0.181***	0.116**	0.010	0.034	0.008
	(0.049)	(0.061)	(0.049)	(0.072)	(0.092)	(0.071)
WPI (alone)	0.029	0.045	0.031	0.038	−0.120	0.051
	(0.030)	(0.036)	(0.030)	(0.075)	(0.096)	(0.073)
Mean dep.var.	0.541		0.541	0.541		0.541
Treatment effect	ATT	ATT	ATT	ATE	ATE	ATE
Sampling weight	No	PS	No	No	Repr	No
Model	LPM	LPM	Probit	LPM	LPM	Probit
P-value($H_0 : \xi_{CLTS} = 0$)				0.078	0.123	0.027
P-value($H_0 : \xi_{CxW} = 0$)				0.169	0.645	0.139
Observations	4,206	2,754	4,206	4,206	4,206	4,206
R^2	0.068	0.705		0.072	0.725	

Notes: Columns 1-3 estimate equation (3), columns 4-6 estimate equation (2). Sampling weight: PS = inverse propensity score weighting, Repr = population representative sampling weight. Model: LPM = linear probability model with HH fixed effects, Probit = average partial effect of the probit correlated random effects regression is reported at baseline mean values. All regressions include HHs interviewed in at least 2 survey rounds, and control for HH size, education, wealth and year. LPM standard errors are robust to clustering, heteroskedasticity and serial correlation; Probit standard errors are robust to clustering at community level (* 0.10 ** 0.05 *** 0.01).

Table G2: Robustness checks for the treatment effects on handwashing with soap or ash after defecation

	Treatment effect estimates for handwashing					
	(1)	(2)	(3)	(4)	(5)	(6)
CLTS (alone)	0.097**	0.122*	0.064*	0.091	0.120	0.046
	(0.049)	(0.070)	(0.038)	(0.104)	(0.143)	(0.075)
CLTS&WPI	0.157***	0.181**	0.114**	0.085	0.139	0.045
	(0.054)	(0.088)	(0.045)	(0.136)	(0.150)	(0.097)
WPI (alone)	−0.013	0.015	−0.012	−0.136	−0.127	−0.092
	(0.045)	(0.068)	(0.034)	(0.092)	(0.087)	(0.064)
Mean dep.var.	0.379		0.379	0.379		0.379
Treatment effect	ATT	ATT	ATT	ATE	ATE	ATE
Sampling weight	No	PS	No	No	Repr	No
Model	LPM	LPM	Probit	LPM	LPM	Probit
P-value($H_0 : \xi_{CLTS} = 0$)				0.970	0.185	0.917
P-value($H_0 : \xi_{CxW} = 0$)				0.511	0.442	0.396
Observations	4,170	2,728	4,170	4,170	4,170	4,170
R^2	0.181	0.482		0.182	0.494	

Notes: Columns 1-3 estimate equation (3), columns 4-6 estimate equation (2). Sampling weight: PS = inverse propensity score weighting, Repr = population representative sampling weight. Model: LPM = linear probability model with HH fixed effects, Probit = average partial effect of the probit correlated random effects regression is reported at baseline mean values. All regressions include HHs interviewed in at least 2 survey rounds, and control for HH size, education, wealth and year. LPM standard errors are robust to clustering, heteroskedasticity and serial correlation; Probit standard errors are robust to clustering at community level (* 0.10 ** 0.05 *** 0.01).

H Determinants of the use of improved water points

The observation that households do not always use an improved WP when it is available in their community raises the question which factors infuence the households' decision to use them. Table H1 lists the most often mentioned reasons by the households for using their chosen WP during the endline survey. The most important factors include the distance to the improved and traditional WPs and the perception of the water quality/safety at the improved and traditional WPs.

We model the households' decision problem on choosing to use of improved WPs (IWP) with the following latent variable model:

$$
\begin{aligned}
IWP_{it}^{**} &= \gamma_3 DI_{it} + \gamma_4 DT_{it} + \gamma_5 QI_{it} + \gamma_6 QT_{it} + \gamma_8 X_{it} + \eta_i + \varepsilon_{it} \\
IWP_{it}^{*} &= \begin{cases} 1 & \text{if } IWP_{it}^{**} > 0 \\ 0 & \text{if } IWP_{it}^{**} \leq 0 \end{cases} \\
IWP_{it} &= \begin{cases} IWP_{it}^{*} & \text{if } W_{it} = 1 \\ 0 & \text{if } W_{it} = 0 \end{cases}
\end{aligned} \tag{8}
$$

where the first equation (IWP^{**}) shows the latent demand function for the use of an improved WP based on the characteristics of alternative water sources and household characteristics; the second equation (IWP^{*}) shows whether the household would choose to use an improved WP based on the value of IWP^{**}; and the last equation (IWP) indicates the observed outcome on the use of an improved WP, which is conditional on actually having the option to use an improved WP. W_{it} stands for the availablity of an improved WP in the community.

In the equation for the latent demand, DI_{it} and DT_{it} stand for household i's distance to the closest improved and traditional WP in cluster at time t, respectively. QI_{it} and QT_{it} represent the perceived water quality at the improved and traditional sources. As before, X_{it} contains other observable household characteristics that may affect the households' decision to use an improved WP (education, wealth index and household size).

In the analysis, distance to the improved and traditional WPs is measured as the

Table H1: Reasons for choosing to use improved and traditional WPs

	Traditional WP	Improved WP
Closest WP to HH	72.6	70.9
Short waiting time	9.1	5.3
WP always has water	16.9	14.9
Water is free	13	4.4
Reasonable cost	0.8	9.3
Good water quality	5.9	59.5
Other	34.8	11.2

Notes: Percent of households indicating response during the household survey in 2013.

crow flies using GPS coordinates. The perceived water quality is only available for the WP actually used by the households. Therefore, we use the mean valuation of the households within their communities separately for the improved and traditional WPs (when available) in each survey round. Therefore, in the estimation of (8) we use community fixed effects instead of household fixed effects (η_i).

We estimate equation (8) as a linear probability model using observations on all households in communities where an improved WP is available in the given survey round. As a robustness check, we also estimate equation (8) using only the observations immediately before and after the WP was installed.

Table H2 shows the estimates of equation (8) for the linear probability model using community fixed effects instead of household fixed effects (η_i). We use observations on all households in communities where an improved WP is available in the given survey round. As a result, communities where an improved WP was available earlier are included in the sample more often. Distance to the improved and traditional WPs are measured as the crow flies using GPS coordinates. The perceived water quality is only available for the WP actually used by the households. Therefore, we use the mean valuation of the households within their communities separately for the improved and

Table H2: Factors influencing the use of improved WPs

	Dependent variable:		
	Use of improved WP		
	(1)	(2)	(3)
Distance to improved WP	−0.134***	−0.255***	−0.313***
	(0.030)	(0.056)	(0.042)
Distance to traditional WP			0.201***
			(0.044)
Mean quality at improved WP	0.167***	0.071	0.069
	(0.056)	(0.068)	(0.069)
Mean quality at traditional WP		−0.099**	−0.090*
		(0.046)	(0.047)
Wealth index	−0.008	−0.010	−0.008
	(0.008)	(0.013)	(0.013)
Formal education	0.060**	0.080**	0.107***
	(0.027)	(0.038)	(0.039)
HH size	0.002	0.0003	0.001
	(0.003)	(0.005)	(0.005)
Year 2010	0.125***	0.100	0.083
	(0.035)	(0.087)	(0.090)
Year 2013	0.146***	0.148*	0.131
	(0.032)	(0.077)	(0.083)
Mean dep.var.	0.709	0.597	0.573
R^2	0.098	0.141	0.181
Observations	1,642	1,195	1,071

Notes: Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Sample includes all HHs in communities where an improved WP (1) and both improved and traditional WP (2&3) were available in the given survey round. Community fixed effects are included.

traditional WPs (when available) in each survey round.

The first column of Table H2 shows the regression results controlling only for the characteristics of improved WPs and of households. The table suggests that the distance to improved WPs is the most important determinant of the choice to use an improved WP. Households where more than half of the adult members have any level of formal education are more likely to use an improved WP. Importantly, we observe that the wealth index has no predictive power. In communities where the mean perceived water quality at the improved WP was higher, households were more likely to use it (column 1). However, when controlling for the characteristics of traditional WPs (columns 2 and 3), this association reduces in size and becomes insignificant. At the same time, the columns 2 and 3 indicate that households are more likely to use an improved WP in communities where the perceived water quality of the traditional WP is worse.

Including information on traditional WPs reduces the sample size to communities where households use both improved WP and traditional WP within the same survey year. Excluding communities where everyone switched to using an improved WP, increases the effect of distance to the improved WP and education on the choice to use an improved WP. An additional distance of 1 km from the improved WP reduces the probability of using it by up to 31pp in column 3. For the sub-sample where we can measure the households' distance to the closest traditional WP, we find that the distance to the traditional WP also affects the choice to switch to using an improved WP: an additional distance of 1 km from the traditional WP increases the probability of using the improved WP by 20pp.

As a result of using the community mean of the households' valuation of the improved and traditional WPs (when available) for each survey round, the identification of the coefficients on perceived water quality comes exclusively from the changes in mean perceived water quality over time in communities that were included in the sample multiple times. However, the coefficient estimates for perceived water quality change only marginally when we estimate the model assuming random effects, which estimates the coefficients using the whole sample (results available upon request).

The regressions in Table H2 overweight communities where an improved WP has been installed earlier. In addition, the precision of the variables used to proxy perceived water quality are correlated with the share of households using the type of WP. In order to establish a causal relationship between the decision to use an improved WP and the perceived water quality, in Table H3 we use observations on each community only once, at the first observation after the improved WP was installed. This way, we can use the perceived water quality at the traditional WP from the previous survey round:[51]

$$IWP^{**}_{it} = \gamma_3 DI_{it} + \gamma_4 DT_{it} + \gamma_6 QT_{i(t-1)} + \gamma_8 X_{it} + \eta_i + \epsilon_{it}. \qquad (9)$$

This regression allows us to identify the determinants of the choice to use improved WP when it first became available. Due to the water point intervention, we have observations on 567 households (35.5% of the sample in cross-section) using either a combination of 2008-2010 or 2010-2013 data.

Table H3 confirms that distance to the improved WP plays a crucial role in the adoption decision: an additional distance of 1 km reduces the probability of using an improved WP by 35pp. The findings on the effect of education and the insignificance of wealth remain unchanged.

The coefficient of perceived water quality at traditional WP is smaller than in Table H2 (coef=-0.036, p=0.056). However, this is a conservative estimate because we are using information on the valuation of water quality from two to three years before the observation on the use of improved WP and before the improved WP were installed. Another difference in the two coefficient estimates for the perceived quality at traditional WP is that while in Table H2 the identification of the coefficient comes from the variation in the mean perception within the community over time, in Table H3 the identification comes from the variation in perceived water quality within the community before the improved WP was installed.

Unfortunately, we cannot look at the effect of payment for the use of improved WP because it would be a community-specific variable. However, we observe that 81% of the communities with an improved WP have regular payment for these WPs in 2013 (80% with and 85% without the WP intervention). At the same time, 75% of the

[51] Notice that the coefficient on the perceived water quality at improved WP cannot be identified because it is only measured for households actually using it and the cluster mean is not identified due to community fixed effects.

Table H3: Factors affecting households' decision to switch to using improved WPs when it becomes available

	Dependent variable:	
	Use of improved WP	
	(1)	(2)
Distance to improved WP (t)	-0.349^{***}	-0.360^{***}
	(0.063)	(0.069)
Distance to traditional WP (t-1)		0.044
		(0.073)
Quality at traditional WP (t-1)	-0.032^{*}	-0.036^{*}
	(0.019)	(0.019)
Wealth index (t)	-0.012	-0.014
	(0.017)	(0.017)
Formal education (t)	0.085^{***}	0.096^{***}
	(0.029)	(0.026)
HH size (t)	0.012^{**}	0.012^{*}
	(0.006)	(0.006)
Mean dep.var.	0.711	0.712
R^2	0.175	0.182
Observations	567	555

Notes: Standard errors are robust to clustering at the community level, heteroskedasticity and serial correlation (* 0.10 ** 0.05 *** 0.01). Community fixed effects are included. t can be 2010 or 2013 depending on when an improved WP was installed in the community. The sample consists of time-location combinations where improved WP becomes available. Each location is only included once. Observations are used before $(t-1)$ and after (t) the improved WP is placed in the community.

households in communities with regular payment for (all) improved WP use these WP compared to 63% of HHs in communities without payment. This suggests that payment for the use of IWP is not a limiting factor for most households.

Overall, we can conclude that the distance to improved WP, the quality of the water at the traditional WP and the education of households are the main determinants of the use of improved WP, supporting the responses of households in Table H1.