

TI 2012-069/2

Tinbergen Institute Discussion Paper



# Effectiveness of Large Scale Water and Sanitation Interventions: The One Million Initiative in Mozambique

*Chris Elbers<sup>1</sup>*

*Samuel Godfrey<sup>2</sup>*

*Jan Willem Gunning<sup>1</sup>*

*Matteus van der Velden<sup>2</sup>*

*Melinda Vigh<sup>1</sup>*

<sup>1</sup> Faculty of Economics and Business Economics, VU University Amsterdam;

<sup>2</sup> UNICEF Mozambique.

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

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Effectiveness of Large Scale Water and Sanitation Interventions: the *One Million Initiative* in Mozambique<sup>+</sup>

Chris Elbers<sup>\*</sup>, Samuel Godfrey<sup>\*\*</sup>, Jan Willem Gunning<sup>\*</sup>, Matteus van der Velden<sup>\*\*</sup>  
and Melinda Vigh<sup>\*</sup>

Revised July 2012

## Abstract

The *One Million Initiative* aims to give one million people in rural Mozambique access to clean drinking water and adequate sanitation by constructing new water points and providing sanitation training. We use panel survey data for 1600 households to analyze the health impact of the Initiative. The paper moves beyond a black box evaluation by analyzing the contribution of various channels through which the interventions affected health. To our knowledge this is the first rigorous evaluation of such a large scale program in the water and sanitation sector. We find that the water point intervention had a sizeable impact on the use of improved water sources and on the health of young children (up to 3 years), while the sanitation training had a strong impact on latrine ownership and on the health of both adults and older children.

---

<sup>+</sup>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 on earlier drafts. 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).

<sup>\*</sup> VU University Amsterdam and Tinbergen Institute.

<sup>\*\*</sup> UNICEF Mozambique.

## 1. Introduction

In this paper we present an evaluation of a very large intervention, aimed at reaching a million people. It is not clear how such large programs are to be evaluated: although there has been an enormous advance in the use of rigorous statistical methods for evaluation, including randomised control trials (RCTs) these techniques are designed for small-scale interventions. Large programs are typically not randomly assigned and often randomisation is not even feasible. The literature suggests ways in which observational data can be used for the evaluation of large programs, e.g. World Bank (2006), Elbers *et al.*, (2009), Elbers and Gunning (2011). The current evaluation follows this approach to evaluate the *One Million Initiative*, a very large water, sanitation and hygiene program in Mozambique.

Fewtrell *et al.* (2005) and IEG (Independent Evaluation Group, 2008) review the literature on the impact of water and sanitation interventions. These reviews indicate that interventions at the source (supplying safe drinking water or chlorination) seem to have little if any effect on health outcomes. In particular, supplying safe water through standpipes or communal wells is not effective in improving health outcomes; only house connections are effective. Water treatment at the point of use (boiling, chlorination or filtration) *is* effective: in the studies reviewed by Fewtrell *et al.* the relative risk of diarrhoeal disease is 0.69 in treatment households.

Sanitation interventions typically promote latrine use, usually by providing latrines, sometimes by stimulating their construction through subsidies. Hygiene interventions rely on training to promote hygiene awareness, notably the importance of hand washing and latrine use. According to both reviews these interventions are generally effective, yielding relative risks of around 0.70 although there is some evidence of publication bias.<sup>1</sup>

In an older meta analysis of 144 studies Esrey *et al.* (1991) find a favourable impact of water and sanitation programs on a number of water borne diseases. They stress the importance of providing water close to the point of use and integrating hygiene education into water and health programs.

Jalan and Ravallion (2003) investigate the impact of access to piped water in rural India. They find a substantial reduction in the prevalence of diarrhoea, but there is an important interaction with income and education: in poor households the effect is very small, particularly if mothers are poorly educated. This suggests that investing in water supply is effective only when combined with sanitation and health education. The *One Million Initiative* is an example of such a combination.

---

<sup>1</sup> However, a number of recent impact studies found that households often do not even recall the hygiene and sanitation training they received let alone any lessons learnt (for example, IOB 2007, 2008, 2009).

A recent innovation is the Community Led Total Sanitation (CLTS) approach which aims at the promotion of latrine use by means of confrontational methods to convince households of the health risks of open defecation. While many developing countries have adopted this approach in recent years,<sup>2</sup> its effectiveness has rarely been assessed rigorously. An exception is Pattanayak *et al.* (2009) who finds that in India the CLTS shaming techniques in combination with a subsidy resulted in a 30 percent increase in improved latrine ownership. However, Pattanayak *et al.* (2007) do not find an impact on diarrhoea prevalence among under-5 children. Chakma *et al.* (2008) evaluate the impact of the Indian Total Sanitation Campaign by comparing health outcomes in Open Defecation Free (ODF) villages and non-ODF communities. They find that diarrhoeal morbidity and worm infestations were significantly less common in ODF-villages.

Mozambique is extreme in the extent of open defecation and, like most African countries, has high child mortality due to diarrhoea. Concerns about these outcomes have led to the *One Million Initiative*, a joint program initiated by the Government of Mozambique, UNICEF and the Government of the Netherlands in 2006. The program involves the creation or rehabilitation of improved communal water sources and a version of the CLTS approach aimed at eradicating open defecation and hygiene promotion.

Household panel survey data were collected for an evaluation of the Initiative (UNICEF and IOB, 2012). This paper presents evidence on the impact of the *One Million Initiative*, using the first two rounds of survey data.

Like Pattanayak *et al.* (2009) we find that the CLTS intervention succeeded in inducing people to build and use latrines. We also find that the creation of new improved water points induced many households to switch to the new water sources. The question is whether these intermediate outcomes resulted in improved health. We address this using a ‘black box’ regression of a health indicator on treatment indicators for the water and sanitation interventions. We find that the CLTS intervention does have a favourable effect on health, confirming the findings in the literature on the effectiveness of sanitation programs. The impact is driven by the outcomes for adults: for young children (under-5) we do not find a significant health impact of CLTS, confirming Pattanayak *et al.* (2009). Regarding the water intervention, we find a significant impact of the water intervention on the health of very young children. To investigate the mechanism underlying these black box results, we use instrumental variables (IV) regressions. We find that the direct effect of safe sanitary practices on health is positive and highly significant. The effect of using improved water sources is much weaker and largely driven by the effect of safer water on the health of very young children.

---

<sup>2</sup> For a complete list see <http://www.communityledtotalsanitation.org/where>.

The paper is organized as follows. Section 2 describes the *One Million Initiative* in detail. The survey and the data are discussed in section 3. Section 4 presents the identification strategy. The results are presented in section 5. Section 6 concludes.

## **2. The *One Million Initiative***

The *One Million Initiative*<sup>3</sup> (2006-2013) covers 18 districts in three provinces in central Mozambique: Manica, Sofala and Tete with a total population of 4.9 million people. The program therefore targets a significant part of the population. Among the goals of the Initiative are that at least 70 percent of the population use safe drinking water and at least 50 percent use improved sanitation facilities. The program provides sanitation and hygiene training and makes improved water sources available by creating or rehabilitating water points (mainly boreholes fitted with hand pumps).<sup>4</sup> These are community-level interventions and can therefore be considered as intentions to treat at the household level.

An important ultimate rationale for a water and sanitation (WASH) program is its impact on health. Using safe water is expected to reduce the prevalence of diarrhoea and other water borne diseases. Safe hygiene practices will enhance this effect, while unsafe practices can nullify it. The interventions can affect the health of the household members directly, by inducing them to use latrines or to make use of an improved water source, but also indirectly, through the effect of the interventions on the behaviour of other households in the same community. Since the interventions are at the level of the community our estimates of their impact are inclusive of such externalities (or spill-over effects).

From the outset the intention was to work through existing government institutions rather than setting up a separate donor run organization to implement the Initiative. However, UNICEF maintains very strict auditing controls over the program. In particular, UNICEF does not disburse funds unless the relevant government administrative bodies have worked out the plans for a particular component of the program in detail or provide evidence of completion of works. This combination of relying on government implementation and strict financial procedure makes the Initiative an interesting departure from sector budget support as commonly practiced, but this is not the topic of the present paper which focuses exclusively on the Initiative's health impact and its transmission channels.<sup>5</sup>

Many WASH programs promote awareness through a Participatory Hygiene and Sanitation Transformation (PHAST) program.<sup>6</sup> The effectiveness of PHAST is questionable: IOB (2007, 2008,

---

<sup>3</sup> See [http://www.unicef.org/mozambique/child\\_survival\\_3151.html](http://www.unicef.org/mozambique/child_survival_3151.html).

<sup>4</sup> The program also involved hygiene promotion but this intervention cannot be evaluated since it was implemented in all communities prior to the baseline. We may therefore underestimate the impact of the Initiative.

<sup>5</sup> Details about the placement of the interventions can be found in the supplemental material, section S4.

<sup>6</sup> See [http://www.who.int/water\\_sanitation\\_health/hygiene/envsan/phastep/en/](http://www.who.int/water_sanitation_health/hygiene/envsan/phastep/en/), accessed February 2011.

2009) describe several evaluations where households who did receive PHAST training did not even recall this fact let alone any lessons. In Mozambique the effectiveness of PHAST was also questioned by program officers of the *One Million Initiative*. Therefore, after initially using PHAST, the *One Million Initiative* switched to the ‘Community Led Total Sanitation’ (CLTS) approach in order to improve sanitary practices of the rural population. CLTS, a version of ‘Community Approaches to Total Sanitation’ (CATS) combines trainings with a system of rewards for communities that become open defecation free (ODF). CLTS itself is a highly confrontational method developed by Kamal Kar in Bangladesh (Kar with Chambers, 2008) and has been adapted to the Mozambique context. In essence it relies on shaming a community into using latrines by letting them discover the consequences of open defecation. CLTS involves the promotion of safe sanitation practices including the construction of latrines but implementation is left to the households who bear the full costs themselves.<sup>7</sup>

### 3. Survey design, data and attrition

The evaluation is based on survey data collected in two rounds, in 2008 and 2010 (UNICEF and IOB, 2011). A third round is envisaged for 2013. The sample consists of 80 communities randomly selected from 9 of the 18 districts in the three provinces. The Initiative is targeted at the poorer communities and these were oversampled to ensure sufficient statistical power. In the analysis we take this selective program placement into account, either by using first difference regressions or by including community dummies.<sup>8</sup>

Treatment and control communities were not assigned *ex ante*: UNICEF considered this unethical since it would involve telling some communities that they would not benefit from the program, or only later.

The sample data were collected in a household survey, focus group discussions with local leaders, and a water points survey. In each community 20 households were selected by systematic sampling from a randomly chosen contiguous group of approximately 100 households (corresponding to about 500 persons).<sup>9</sup> To ensure that survey households lived close to a potential new water point, communities were first asked which location they would prefer for a new water point. In the baseline households were then sampled in the neighbourhood of that location.<sup>10</sup> The household survey covers general household characteristics and health, water and sanitation practices. The participants in the focus group discussions

---

<sup>7</sup> For details about the reward system of CATS see Godfrey (2009).

<sup>8</sup> Since human development outcomes are typically worse in poorer communities any cross-sectional evidence of a favourable impact is likely to be an underestimate.

<sup>9</sup> For details on the sampling procedure, see UNICEF / WE Consult (2009).

<sup>10</sup> In some locations improved water sources were created shortly before the baseline survey. This does not seem to have had an effect on the baseline survey outcomes.

were local community leaders from the immediate vicinity of the sample households. The water point survey collected data on the water sources used by the sample households.

In each survey round 1600 households were interviewed and 1310 of them participated in both rounds. There are four groups of communities: those without any intervention, those with only a water intervention or only a sanitation intervention and those with both types of interventions. Since the interventions were targeted on poorer communities there are significant differences between the baseline characteristics of these four groups. There are also small but significant differences in a few characteristics not directly related to targeting: household size, age of the household members and education. The impact of these differences will be eliminated if household fixed effects are included. Since 290 of the baseline households dropped out selective attrition needs to be considered. Fortunately, a probit analysis of attrition suggests that there are no differences between those who left and those who remained except for household size and latrine ownership.<sup>11</sup>

The main health indicator in both the 2008 and 2010 rounds of the household survey is self-reported prevalence of ‘water borne diseases’ with a recall period of six months where the respondent reports for all household members. This is a modified MICS indicator.<sup>12</sup> The indicator is problematic since the recall period is too long for e.g. diarrhoea.<sup>13</sup> In addition, few households are likely to know the concept of ‘water borne diseases’. While it was not explained in the questionnaire enumerators were allowed to give the examples of diarrhoea, typhoid and cholera. To construct a more robust health measure we aggregated the indicator at the household level and converted it to a binary variable taking the value 1 if *any* household member was reported as having suffered from a water borne disease in the six months period and zero otherwise. Our analysis therefore rests on the assumption that if the degree of misreporting has changed between survey rounds the change is the same way for the four groups.<sup>14</sup>

#### **4. Identification strategy**

The analysis is based on Figure 1 which shows how the interventions of the *One Million Initiative* affect health: the provision of improved water sources affects their use and the CLTS intervention affects latrine

---

<sup>11</sup> See the supplemental material, section S6, for a description of the household and community characteristics and the regression results.

<sup>12</sup> MICS stands for Multiple Indicator Cluster Surveys. These surveys have been developed by UNICEF to produce internationally comparable indicators on a range of indicators in the areas of education and health. They have been used worldwide for the monitoring of the Millennium Development Goals.

<sup>13</sup> In 2010 households were also asked about water borne diseases during the two weeks preceding the interview. This is the common recall period used in health studies. However, the results implied far too low prevalence to be credible and could therefore not be used for evaluating program impact

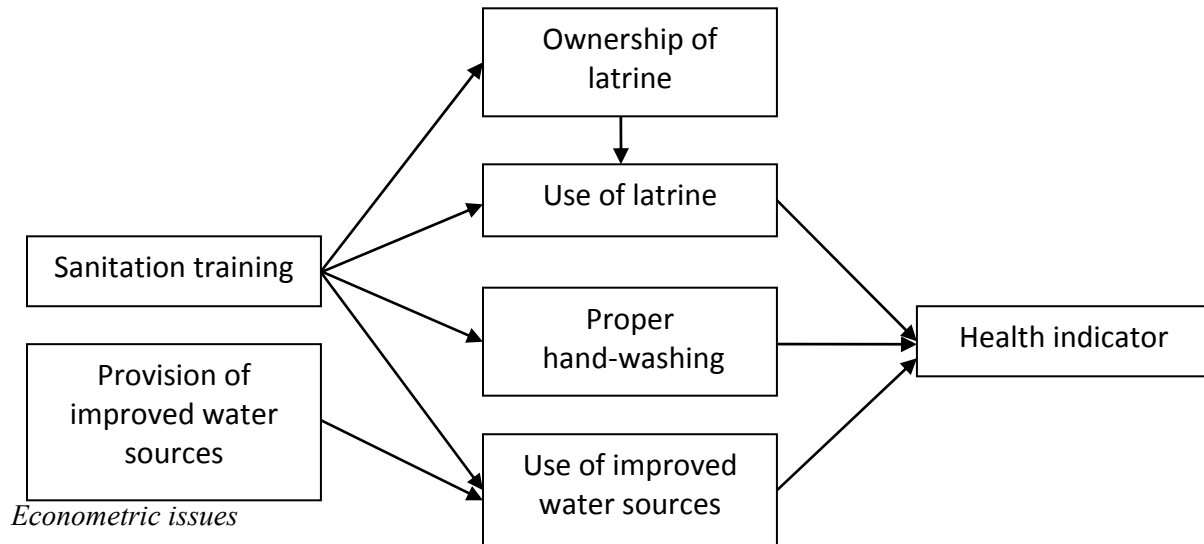
<sup>14</sup> A disadvantage of this procedure is that the impact of the Initiative cannot be measured in terms of prevalence as commonly defined.



use (both directly and indirectly, through increased latrine ownership), hand washing and the use of improved water sources. The Figure shows that health can be affected through three channels.<sup>15</sup>

This section discusses the econometric methodology used to estimate the effects shown in the Figure 1. (Readers not interested in methodological questions can skip to the discussion of the results in section 5.)

**Figure 1: Impact of water and sanitation interventions on health**



In the statistical analysis, we use binary outcome variables indicating whether a household uses a latrine or improved water sources. To evaluate the impact of the two interventions on these binary variables we specify the following probability model:

$$Y_{h,t} = F(\alpha_t + \beta_w W_{\ell(h)} d_t + \beta_s S_{\ell(h)} d_t + \gamma X_{h,t} + \lambda_{\ell(h)} + \eta_h + \varepsilon_{h,t}) \quad (1)$$

where  $Y_{h,t}$  is the binary outcome variable of interest for household  $h$  at time  $t$ ;  $W_{\ell(h)}$  and  $S_{\ell(h)}$  indicate whether there was a water point or sanitation intervention in the cluster (location)  $\ell$  of household  $h$ ;  $d_t$  is a time dummy (1 in 2010); and  $X_{h,t}$  denotes household-specific control variables like household size and wealth. Time, cluster and household fixed effects are denoted by  $\alpha_t$ ,  $\lambda_{\ell(h)}$  and  $\eta_h$  and  $\varepsilon_{h,t}$  is a random error.<sup>16</sup> For particular choices of the function  $F$  and the distribution of  $\varepsilon_{h,t}$  model (1) reduces to the normal, logistic or linear probability model. In this paper we report estimates for the linear probability model.<sup>17</sup>

<sup>15</sup> This Figure is based on the Figure in the supplemental material which itself is adapted from Waddington and Snilstveit (2009). The Figure does not allow for water treatment at the point of use since this is rare in Mozambique.

<sup>16</sup> The regressions in Tables 1-4 also include an interaction of distance to the water point and the water intervention dummy.

<sup>17</sup> Results for the logistic model are reported in the supplemental material.

We are interested in the coefficients  $\beta_w$  and  $\beta_s$ , which measure the treatment effect of the interventions.<sup>18</sup> To estimate these effects, we use the linear probability model. Logit models are estimated to verify the robustness of the results. We use both first difference and pooled estimators.

If  $F$  is the identity function then equation (1) reduces to the linear probability model. Taking first differences then gives

$$\Delta Y_{h,t} = \alpha + \beta_w W_{\ell(h)} + \beta_s S_{\ell(h)} + \gamma \Delta X_{h,t} + \Delta \varepsilon_{h,t} \quad (2)$$

where  $d_t$  drops out since  $d_2 - d_1 = 1$  and  $\alpha$  denotes  $\alpha_2 - \alpha_1$ . The advantage of this specification is that it removes all household-specific time-invariant unobservables (fixed effects) that can bias the estimates of the treatment effect due to non-random placement of the interventions. This comes at a cost: *observed* time-invariant variables can also contain information on the processes studied and this is also lost through differencing. In addition, differencing leads to larger standard errors.

If the household-specific effects are uncorrelated with the treatment status after controlling for the location of the household, i.e. if

$$W_{\ell(h)}, S_{\ell(h)} \perp \eta_h \mid \lambda_{\ell(h)} \quad (3)$$

then pooled estimators controlling for *location* fixed effects can be used to estimate the treatment effect consistently. This assumption seems plausible: since the interventions are at the cluster level location dummies are likely to deal with selectivity bias. Using pooled data, the distribution of the outcome variable can be estimated using

$$Y_{h,t} = F(\alpha_t + \beta_w W_{\ell(h)} d_t + \beta_s S_{\ell(h)} d_t + \gamma X_{h,t} + \lambda_{\ell(h)} + e_{h,t}) \quad (4)$$

where  $e_{h,t} = \eta_h + \varepsilon_{h,t}$ . We use this specification for the linear probability model to check the robustness of the earlier (differenced) estimates.

#### *Regressions for individual outcomes*

For the health outcome we can also use individual data. Unfortunately, the dataset does not allow for matching of the household members in the two survey rounds. Therefore, we use household and location fixed effects to deal with time-invariant unobservable factors affecting the health outcome of household members. In this case the distribution of individual health outcomes can be written as

$$Y_{m,h,t} = F(\alpha_t + \beta_w(A_{m,h,t}) W_{\ell(m,h)} d_t + \beta_s(A_{m,h,t}) S_{\ell(m,h)} d_t + \delta A_{m,h,t} + \gamma X_{m,h,t} + \lambda_{\ell(m,h)} + \eta_{m,h} + \varepsilon_{m,h,t}) \quad (5)$$

---

<sup>18</sup>Note that model (1) rules out synergy effects arising from the joint implementation of water point and sanitation interventions. We experimented with interaction terms for the interventions but found no significant effects.

where the binary variable  $Y_{m,h,t}$  denotes the health outcome of household member  $m$  in household  $h$  at time  $t$ . We allow for age dependence ( $A_{m,h,t}$ ) of the impact since we expect a higher disease risk for young children.

### *The impact of CLTS interventions*

Presumably the interventions were aimed at households who had not yet acquired a latrine or started using an improved water point. The question therefore arises how successful the interventions were in convincing *non-owners* and *non-users* to change their behaviour. To formalize the impact on adoption behaviour, consider the following linear probability model for latrine ownership in round 2 of the survey:

$$p_h = p_{\ell(h)}^N + (1 - p_{\ell(h)}^N) S_{\ell(h)} \pi + z_h \gamma + \varepsilon_h, \quad (6)$$

where  $p_h$  is the probability that household  $h$  in location  $\ell(h)$  owns a latrine,  $p_{\ell(h)}^N$  the probability that that household owns a latrine irrespective of whether it received CLTS training and  $\pi$  the probability that the CLTS training would convince a household to build one. Therefore,  $\pi$  is the success rate of the CLTS training for households who would not have owned a latrine in round 2 without the training. Note that  $p_h$  also depends on observed ( $z_h$ ) and unobserved ( $\varepsilon_h$ ) household characteristics which are taken to be independent of  $p_{\ell(h)}^N$ . Of course,  $p_{\ell(h)}^N$  can only be estimated for locations without CLTS training, but if we assume that the intervention indicator  $S_{\ell(h)}$  is independent of  $p_{\ell(h)}^N$  then the population average of  $p_{\ell(h)}^N$  can be estimated using the data for non-CLTS locations.

Denote the average latrine ownership in location  $\ell$  in round 1 by  $q_{1\ell(h)}$ . Assuming that  $q_{1\ell(h)}$  is independent of  $\varepsilon_h$  and that  $p_{\ell(h)}^N \perp S_{\ell(h)} \mid q_{1\ell(h)}, z_h$  we can use  $q_{1\ell(h)}$  as an instrumental variable.<sup>19</sup> The second assumption is reasonable if  $z_h$  includes a wealth indicator since placement of the CLTS intervention was targeted on the poorer locations.<sup>20</sup> Regarding the exogeneity of  $q_{1\ell(h)}$ , note that it is the latrine ownership rate at baseline very similar to  $p_{\ell(h)}^N$  and therefore unlikely to contribute to round 2 latrine ownership ( $p_h$ ), given  $p_{\ell(h)}^N$ . Moreover  $q_{1\ell(h)}$  is a location level variable. With a suitable set of controls  $z_h$  one would therefore expect it to be uncorrelated with  $\varepsilon_h$ . Taking conditional expectations in equation (6) gives

$$E(p_h \mid S_{\ell(h)}, q_{1\ell(h)}, z_h) \approx E(p_{\ell(h)}^N \mid q_{1\ell(h)}, \bar{z}_{\ell(h)}) + \pi S_{\ell(h)} (1 - E(p_{\ell(h)}^N \mid q_{1\ell(h)}, \bar{z}_{\ell(h)})) + z_h \gamma, \quad (7)$$

<sup>19</sup> The orthogonality symbol is used to indicate that two variables are not correlated.

<sup>20</sup> The correlation between average 2008 latrine ownership and the CLTS intervention is very low (0.06) and insignificant. Latrine prevalence has therefore not guided CLTS targeting.

where  $\bar{z}_{\ell(h)}$  denotes the mean value of  $z$  in location  $\ell$ . It is convenient to linearize the conditional expectations on the right hand side:

$$p_{\ell(h)}^N = E(p_{\ell(h)}^N | q_{1\ell(h)}, \bar{z}_{\ell(h)}) + \omega_{\ell(h)} = \alpha + \beta q_{1\ell(h)} + \delta \bar{z}_{\ell(h)} + \omega_{\ell(h)}. \quad (8)$$

Substitution in (6) gives the nonlinear regression equation

$$p_h = \alpha + \beta q_{1\ell(h)} + \delta \bar{z}_{\ell(h)} + (1 - \alpha - \beta q_{1\ell(h)} - \delta \bar{z}_{\ell(h)}) \pi S_{\ell(h)} + \gamma z_h + \varepsilon_h + \omega_{\ell(h)} (1 + \pi S_{\ell(h)}). \quad (9)$$

Note that the uncorrelatedness assumption implies that  $\omega_{\ell(h)} \perp S_{\ell(h)}$  and, since  $S_{\ell(h)}$  is a binary variable,  $\omega_{\ell(h)} (1 + \pi S_{\ell(h)}) \perp S_{\ell(h)}$ . We estimate (9) using nonlinear least squares both for latrine ownership and the use of improved water sources<sup>21</sup> in clusters with a water point intervention to find the impact  $\pi$  of the CLTS intervention on the adoption behaviour of non-user households.

#### *IV regression*

Equation (1) is a black box regression in the sense that it does not specify through which channels the interventions affect health. We like to go beyond this by quantifying the health impact of *using* latrine and improved water sources. Figure 1 shows the three likely channels from the interventions to the health outcome variable: using latrines, using improved sources and hand washing. To identify the impact of these three separately we need at least three instrumental variables. Consider the following linear probability model:

$$H_{h,t} = \alpha_t + \delta_1 B_{h,t} + \delta_2 L_{h,t} + \delta_3 P_{h,t} + \gamma X_{h,t} + \lambda_{\ell(h)} + \eta_h + \varepsilon_{h,t} \quad (10)$$

where  $H_{h,t}$  denotes the health outcome,  $B_{h,t}$  the use of an improved water source,  $L_{h,t}$  latrine use and  $P_{h,t}$  proper hand washing practices. In estimating (10) we instrument  $B_{h,t}$ ,  $L_{h,t}$  and  $P_{h,t}$  by the two interventions and by baseline cluster level indicators of latrine ownership and the presence of improved water sources. In the basic specification we assume  $\delta_3 = 0$  so that the impact of the interventions is channelled exclusively through the use of improved water sources and latrines. In subsequent regressions we test the restriction that  $\delta_3 = 0$  by adding other instrumental variables. We estimate (10) using the first difference model and also by pooling the data at household and individual level.

## **5. Results**

---

<sup>21</sup> In this case  $p_h$  is redefined as the probability of using improved water sources in 2010.

In this section we focus on four questions. First, whether the CLTS sanitation training has succeeded in convincing people to build and use latrines and employ proper hand washing practices. Second, whether the water point interventions and the sanitation training induced people to switch to improved water sources. Third, whether the *One Million Initiative* led to a reduction in reported water borne diseases. Finally, whether these results are consistent with the framework depicted in Figure 1.<sup>22</sup>

*Sanitary practices: latrine ownership, latrine use and hand washing*

The most important aim of the CLTS sanitation intervention is to convince a community to abandon the practice of open defecation and switch to latrine use. The success of this intervention can be investigated directly by looking at latrine ownership.<sup>23</sup> This is done in Table 1, regressions (1) and (2). Column (1) reports the results for the differenced indicator, which measures the change in latrine ownership status. The explanatory variables are the sanitation training and water point interventions and the interaction of the water point intervention and the distance to an improved water point.<sup>24</sup> Regression (2) pools data for all households in the two survey rounds, adding controls for household size and wealth as well as cluster and year dummies. The estimates in the two regressions are quite similar. They show a strong effect (12 to 14 percentage points) of the CLTS training on latrine ownership. The water point intervention has little if any impact consistent with Figure 1 where the water intervention has no effect on latrine ownership.

Since some households already owned a latrine prior to the sanitation intervention the impact of the CLTS training may be underestimated. To assess the effect of the CLTS intervention on non-owners, regression (1) in Table 2 reports the nonlinear least squares (NLS) estimates for equation (9). This regression uses all households in the second survey round and includes both household and location-level controls. The estimate of  $\pi$  implies that the CLTS training convinced around 21 percent of households who would otherwise not have built a latrine to do so.<sup>25</sup>

**Table 1: Impact on latrine ownership, hand washing practices and use of improved water sources**

	latrine ownership	proper hand washing	use of improved water source
--	-------------------	---------------------	------------------------------

<sup>22</sup> The regressions reported in Tables 1-5 are our preferred specifications. Extensive robustness checks are reported in the supplemental material.

<sup>23</sup> Shared use of latrines is uncommon.

<sup>24</sup> The distance is measured as the median distance from the location of the household (using GPS coordinates) to the nearest improved water source for households in a cluster. Using individual distances instead does not appreciably change the results but requires somewhat arbitrary decisions to deal with outliers. The disadvantage of using the median is that it makes distance the same for all households in a cluster. This reduces the precision of the estimates.

<sup>25</sup> Linear regressions without parameter restrictions gives quite similar estimates.

Regression	Household FD (1)	Household pooled (2)	Household FD (3)	Household pooled (4)	Household FD (5)	Household pooled (6)
Mean dependent variable	0.105	0.494	0.209	0.146	0.277	0.283
Water point intervention (wpi)	0.049 (0.061)	0.066 (0.059)	0.011 (0.083)	0.009 (0.069)	0.551*** (0.109)	0.567*** (0.101)
wpi*distance <sup>1</sup>	-0.046 (0.049)	-0.072* (0.040)	-0.019 (0.066)	-0.022 (0.055)	-0.414*** (0.130)	-0.412*** (0.112)
Sanitation training (CLTS)	0.118** (0.055)	0.141** (0.056)	0.100 (0.061)	0.070 (0.052)	0.130 (0.084)	0.112 (0.074)
Intercept	0.051* (0.026)	-0.166*** (0.024)	0.173*** (0.036)	0.124*** (0.026)	0.077* (0.042)	-0.030 (0.025)
Observations	1310	3200	1309	3199	1310	3200
Adjusted R-squared	0.018	0.42	0.005	0.130	0.229	0.592
First difference	yes	no	yes	no	yes	no
Year and location dummies	no	yes	no	yes	no	yes
Additional regressors <sup>2</sup>	no	yes	no	yes	no	yes
Clustered standard errors	yes	yes	yes	yes	yes	yes

Dependent variable: (1,3,5) change of binary indicator; (2,4,6) binary indicator. Proper hand washing is defined as washing hands with ash or soap, after defecation and before eating. The binary hand washing variable equals 1 if the respondent reports that *all* adults in the households practice proper hand washing. The distance variable is the median distance from the GPS coordinates of the household to the nearest improved water point.

<sup>1</sup> The distance variable is the median distance between GPS coordinates of cluster households and the closest improved water source.

<sup>2</sup> Household size, wealth.

Significance: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table 2: Impact of CLTS on latrine ownership for non-owners and on the use of improved water sources for non-users**

	latrine ownership	use of improved water source
Regression	(1)	(2)
Mean dependent variable	0.549	0.668
$\pi$ : effect of CLTS on non-owners	0.213*** (0.067)	
$\pi$ : effect of CLTS on non-users		0.069 (0.130)
$\alpha$ : constant	0.051 (0.150)	0.705* (0.404)
$\beta$ : share of latrines at baseline	0.895*** (0.054)	
$\beta$ : improved water point in cluster at baseline		0.151 (0.124)
$\delta_1$ : water point intervention (wpi)	0.071 (0.052)	
$\delta_2$ : wpi * distance	-0.112** (0.048)	-0.512 (0.450)
$\delta_3$ : average household size in cluster	-0.018 (0.030)	0.047 (0.087)
$\delta_4$ : average wealth in cluster	-0.029 (0.108)	-0.086 (0.135)
$\gamma_3$ : household size	0.012** (0.005)	-0.003 (0.012)
$\gamma_4$ : wealth	0.116*** (0.020)	0.017 (0.039)
Observations	1600	740

Nonlinear least squares estimator: equation (9). Clustered standard errors between parentheses.

Regression (1) includes all households, regression (2) only households in clusters with water point interventions.

Significance: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01.

Source: Household survey and program data on interventions, 2008 and 2010.

Owning latrines is almost equivalent to using one: more than 95 percent of the adults from households with a latrine also use it, but for children the share is lower, at two thirds. This is not surprising since households bear the full costs of latrine construction.

Proper hand washing at critical times can reduce the incidence of diarrhoea.<sup>26</sup> While the main aim of the CLTS intervention is to encourage a switch from open defecation to latrine use it also emphasises the importance of proper hand washing and hygiene in general. In this it does not appear to be effective. In the survey all adults report that they wash their hands at critical times: before eating, after defecation and after disposing of baby's faeces but people rarely use soap or ash. Columns (3) and (4) in Table 1 report the impact of the interventions on proper hand washing, defined as hand washing with soap or ash before eating and after defecation by all adults in the household. The estimated effect of CLTS training is positive but not significant.

#### *Using improved water sources*

As part of the Initiative a large number of water points were created or rehabilitated. Whether a household will start to use a new water source will clearly depend on its location. This is why we expect the use of an improved water source in columns (5) and (6) to depend on the interaction between the water point intervention and the household's distance from it.<sup>27</sup> As shown in the Table the new water points induced many households to use them. However, the effect is very sensitive to the distance to the source. At the mean distance in the sample (about 500 metres) the net effect is quite large: approximately 34 percentage points, but at a distance of more than a kilometre the effect is almost entirely eliminated.

In some of the communities there already existed functioning improved water sources at the time of the baseline survey, some of them constructed by the *One Million Initiative* just before the baseline. In 20 of the 80 clusters there were functioning water points in the neighbourhood according to the water points survey.<sup>28</sup> Obviously, the average use of improved water sources at the baseline is substantially higher in these communities. Hence the impact of the water point interventions might well be different in communities with improved water points prior to the interventions. We have therefore repeated the regressions separately for the clusters with and without a working water point in 2008. The impact of the water point intervention in the no-borehole clusters increases to 48 percentage points (at the mean of the

---

<sup>26</sup> For example Fewtrell *et al.* (2005).

<sup>27</sup> Distance should not play a role in determining latrine ownership or hand washing practices and indeed the interaction term is never significant (at the 5% level) in the regressions in columns (1)-(4).

<sup>28</sup> In 10 of these 20 clusters there was also a water point intervention by the One Million Initiative. Creating an additional water point in such communities makes sense if the number of households per water point was too high.



distance to the source). On the other hand, if the community already had an improved water point then the impact of an additional improved water point is only 1.5 percentage points. However, for this group the CLTS training increases the probability of using improved water sources by 26 percentage points.<sup>29</sup>

To analyze the impact of the sanitation intervention on the use of improved water sources further, regression (2) in Table 2 reports the results of applying the nonlinear model of equation (9) to the use of improved water sources using households form clusters with a water point intervention. The parameter  $\pi$  which captures the impact of CLTS training on a switch to using improved water sources by non-users in the water point intervention clusters is insignificant.

### *Impact on health*

The results for health are shown in Tables 3 and 4. Recall that we measure health status by a dummy variable indicating whether any household member was affected by water-borne diseases in the previous 6 months.<sup>30</sup> We find a substantial and significant effect of CLTS training: it reduces the probability of being affected by 8 percentage points and accounts for 20% of the decline between the two survey rounds.<sup>31</sup> Note that the estimated impact of CLTS is almost identical for the three specifications suggesting that the selective placement of CLTS training is successfully controlled for.

While the effect of CLTS is strong, access to improved water sources has no significant effect on health. This is not surprising since the water is often not safe at the source (even for ‘improved’ water sources) and there is considerable contamination of water with faecal (thermo-tolerant) bacteria between the source and the point of use, a common finding in WASH studies (e.g. IOB, 2006). An alternative explanation is that older household members have developed resistance to the effects of drinking contaminated water.

The results in columns (2) and (3) show no significant effect of changes in wealth, possibly because of lack of variation in the sample.

Whether the interventions affect young children differently is investigated in the last two regressions in Table 3. Our health indicator is still a household level aggregate but the aggregation is now restricted to the children in the household. Column (5) shows that for children under 3 years of age CLTS training has

---

<sup>29</sup> These results are not shown.

<sup>30</sup> This indicator was discussed in section 3. By construction the indicator is sensitive to household size. This variable is therefore included as a control in the regressions shown in columns (2)-(5).

<sup>31</sup> The autonomous decline of 13 percentage points is difficult to explain. It may reflect different weather conditions or differences in methods of enumerators in the two rounds.

no effect but the water point intervention has a substantial health impact, reducing our health variable by 14 percentage points.<sup>32</sup>

**Table 3: Health impact of interventions**

	Household FD	Household FD	Household pooled	Children under 5 pooled	Children under 3 pooled
Regression	(1)	(2)	(3)	(4)	(5)
Mean dependent variable	-0.163	-0.163	0.223	0.251	0.146
Water point intervention (wpi)	0.005 (0.054)	0.004 (0.053)	-0.023 (0.051)	-0.002 (0.062)	-0.141** (0.066)
wpi*distance	-0.027 (0.067)	-0.020 (0.061)	-0.029 (0.059)	-0.039 (0.063)	0.068 (0.074)
Sanitation training (CLTS)	-0.085** (0.043)	-0.082* (0.042)	-0.081** (0.040)	-0.074 (0.047)	-0.001 (0.049)
Household size		0.027*** (0.008)	0.013*** (0.005)	0.019*** (0.007)	0.009 (0.006)
Number of children under 5		0.030 (0.020)	0.032*** (0.010)	0.003 (0.018)	
Number of children under 3					-0.002 (0.028)
Wealth		-0.023 (0.031)	-0.008 (0.010)	0.011 (0.019)	0.003 (0.020)
Intercept	-0.127*** (0.030)	-0.119*** (0.030)	0.454*** (0.030)	0.481*** (0.055)	0.417*** (0.048)
Observations	1280	1279	3161	1798	1366
Adjusted R-squared <sup>2</sup>	0.004	0.023	0.093	0.082	0.055
First difference	yes	yes	no	no	no
Year and location dummies	no	no	yes	yes	yes
Additional regressors <sup>1</sup>	no	yes	yes	yes	yes
Clustered standard errors	yes	yes	yes	yes	yes

Dependent variable based on 6 months recall of disease occurrence (binary) in household. Regressions (3): all household members; (4): children under 5; (5): children under 3. Regressions (1,2) use change in disease occurrence at household level, 2008-2010.

<sup>1</sup>Household size, wealth.

Significance: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

<sup>32</sup> In column (4) this effect disappears: it is relevant only for the very young children.

Of the households without a water intervention in 2010 13% had under-3s who experienced diseases against 6% in other households. In this sense the water intervention halves the disease burden.

We can also investigate the effect of the interventions on the health of children by using the individual data on water borne diseases instead of aggregates at the household level.<sup>33</sup> This is done in Table 4 where pooled regression results are reported. The dependent variable is the health indicator for an individual household member. Children are defined as under-5s in (1) and as under-3s in (2). We allow for heterogeneity both in terms of intercept and the slope variables for the two interventions. The results indicate that the water intervention is important for both groups of children, unlike in Table 3 where the effect appeared to be limited to the under-3s. The effect is very large relative to the mean of the dependent variable: disease occurrence is reduced by more than 60% (0.044/0.071) for the under-5s and by over 70% for the under-3s. The effect of the CLTs training is clearly age-specific: it has no effect on the health of children (perhaps because they are too young to use latrines) but has a strong effect on the health of older household members.

What Tables 3 and 4 show for very young children is plausible: they do not yet use latrines and are more vulnerable to consuming highly contaminated water as their immune system is less resistant.

The newly constructed latrines are almost always traditional pit latrines rather than improved latrines as defined by the MICS. Improved latrines might well have stronger effects on health but the fact that the switch from open defecation to using traditional latrines already has a substantial effect on health is important for the policy debate. If only the adoption of improved latrines is counted as success in terms of the Millennium Development Goals there is a danger that the perfect becomes the enemy of the good.

### *Opening the black box*

Tables 1-4 reported on the intention-to-treat effect of the interventions of the *One Million Initiative*. However, we are also interested in the mechanism underlying these black box results. The channels from the interventions to the health outcome variable in **Error! Reference source not found.** suggest an instrumental variable regression of the health indicator on water source type, latrine use and hand washing, instrumented by the interventions and possibly also by baseline cluster level indicators of latrine ownership and availability of improved water sources.

---

<sup>33</sup> This allows us to investigate whether the effect is heterogeneous across age groups but this comes at a price: the household aggregate health indicator is likely to be a more robust measure.

**Table 4: Impact on individual health outcomes  
(pooled data)**

Regression	Children defined as under 5 (1)	Children defined as under 3 (2)
Mean dependent variable	0.071	0.071
Water point intervention for children in age category	-0.044** (0.020)	-0.051*** (0.019)
Water point intervention for older individuals	0.014 (0.018)	0.011 (0.009)
Sanitation training (CLTS) for children in age category	-0.022 (0.021)	-0.009 (0.020)
Sanitation training (CLTS) for older individuals	-0.037* (0.020)	-0.036*** (0.009)
Child in age category	0.072*** (0.008)	0.087*** (0.007)
Household size	0.005 (0.006)	0.006*** (0.002)
Number of children in age category	-0.0005 (0.009)	-0.008 (0.005)
Wealth	0.00002 (0.015)	-0.001 (0.006)
Intercept	0.402*** (0.016)	0.400*** (0.102)
Observations	17166	17148
Adj. R-squared <sup>2</sup>	0.215	0.215
First difference	No	no
Year, location dummies	Yes	yes
Household dummies	Yes	yes
Additional regressors <sup>1</sup>	Yes	yes
Clustered standard errors	Yes	yes
Dependent variable: (2) reported occurrence of water borne disease for individuals in last 6 months (binary); (1) change in disease occurrence.		
<sup>1</sup> Household size, number of children under 5, wealth.		
<sup>2</sup> McFadden's adjusted R <sup>2</sup> for logit models.		
Significance: '*' 0.1, '**' 0.05, '***' 0.01		
Source: Household survey and program data on interventions, 2008 and 2010.		

In Table 5 we show the results of three regressions. As before we show both double difference and pooled regressions. The IV-regressions confirm the importance of latrine use: the probability that a household reports water borne diseases is 38 percentage points lower if it uses a latrine. In all three regressions the P-value of the coefficient is about 5%. The use of improved water sources decreases the probability by an additional 10 percentage points but this effect is not significant.

In column (1) we only allow the channels of use of improved water and use of latrines. In columns (2) and (3) we also investigate the importance of the hand washing. The results indicate that proper hand washing has no significant effect on health. However, the first-stage F-statistic of the hand-washing variable indicates that the instruments are only weakly related to hand washing. We conclude that it is not possible to evaluate the hand washing channel with the current dataset. Note that adding the hand washing channel does not change the results for the other channels.

In terms of Figure 1 the interventions result in improved health and they do so largely through the ‘use of latrine’ channel. The two other channels, through ‘proper hand washing’ and ‘use of improved water sources’ are much less important except that the use of improved water sources is important for very young children.<sup>34</sup>

**Table 5: Health effect of hand washing, use of improved water sources and latrines (GMM estimates)**

Regression	Household FD	Household FD	Household pooled
	(1)	(2)	(3)
Mean dependent variable	-0.162	-0.162	-0.223
Use of improved water source	-0.106 (0.074)	-0.103 (0.080)	-0.115 (0.084)
Use of latrine	-0.384* (0.199)	-0.379* (0.218)	-0.379** (0.188)
Proper hand-washing		-0.018 (0.299)	-0.136 (0.588)
Household size	0.030*** (0.008)	0.029*** (0.008)	0.015** (0.006)
Number of children under 5	0.033 (0.020)	0.032 (0.020)	0.029 (0.017)
Wealth	0.003	0.003	0.040

<sup>34</sup> IV regressions at the individual level are reported in Table S14 of the supplemental material.

	(0.029)	(0.037)	(0.027)
Intercept	-0.075**	-0.072	0.404***
	(0.032)	(0.054)	(0.116)
Observations	1276	1275	3157
F-stat, first stage (water)	46.7	46.0	51.5
F-stat, first stage (latrine)	7.97	7.95	36.3
F-stat, first stage (hand washing)		2.88	6.47
J-test, degrees of freedom, probability	2.254, 4, 0.69	1.870, 3, 0.60	0.012, 1, 0.91

List of additional instruments <sup>1,2</sup>	Water×FB, FB, SLB, CLTS×SLB	Water×FB, FB, SLB, CLTS×SLB	Water×FB, CLTS×SLB
First difference	Yes	yes	no
Year, location dummies	No	No	yes
Household dummies	no	no	no
HAC robust standard errors	yes	yes	yes

Dependent variable: (4) reported occurrence of water related disease in household in last 6 months (binary); (1,2,3) change in reported disease occurrence

<sup>1</sup> Instrument list always includes Water, CLTS, HH size, number of children under 5, wealth.

<sup>2</sup> Baseline controls: FB is functioning borehole in cluster; and SLB is the share of latrines in cluster (excluding the household in question)

Significance: '\*\*\*' 0.01. '\*\*' 0.05. '\*' 0.1

Source: Household survey and program data on interventions, 2008 and 2010.

## 6. Conclusion

The *One Million Initiative* is an ambitious large-scale program aimed at improving the lives and the health of a million people in one of the poorest countries in the world. It uses the promotion of safe sanitation and hygiene practices and installation or rehabilitation of improved water sources to help achieve the Millennium Development Goals of creating access to improved drinking water and adequate sanitation.

This paper focused on the health impact of the *One Million Initiative*. Using a two-round survey we can trace not just the total impact but also the relative importance of the channels through which the Initiative operates: the switch to using safe water sources and a change in sanitary practices. Using a MICS-based<sup>35</sup> indicator for the prevalence of water borne diseases we find that between the 2008 and 2010 survey rounds the disease indicator in the sample declined from 30 percent to 14 percent. The analysis suggests that one fifth of this decline can be attributed to interventions under the *One Million Initiative*, in

<sup>35</sup> MICS denotes Multiple Indicator Cluster Surveys.

particular to the sanitation intervention. Assuming that our disease indicator is proportional to the prevalence of water borne disease, this is a substantial effect.<sup>36</sup>

The success of the CLTS contrasts with findings of similar evaluations studies (including earlier evaluations of WASH programs by the authors in Tanzania, Yemen, and Egypt) that hygiene training has little effect, if any. The *One Million Initiative* is a favourable outlier in this respect. The difference may be due to the confrontational nature of the CLTS approach which is quite different from the traditional PHAST training program.<sup>37</sup> The health impact of the sanitation intervention is mainly due to its success in convincing people to build latrines, but we also find that the CLTS training motivated non-user households to start using improved water sources in communities where improved water points were or became available. Our results show that the benefits of the sanitation intervention are mainly enjoyed by household members above the age of 5.

The water point intervention has a sizable impact on the use of improved water sources. It also substantially reduces the disease burden of the youngest children (under-3s) who are the most vulnerable. This favourable impact is achieved in spite of the substantial bacterial contamination which occurs between the source and the point (and time) of use of water. Such contamination can severely diminish the health impact of increased access to improved water sources. This suggests that once improved water sources are available in the community it is important to focus the intervention on the hygiene component of water use.

Whether the success of the Initiative, notably its CLTS component, is sustainable remains to be seen. Conceivably the effect may wear off and people may revert to open defecation. The final round of data collection (in 2013) will be used to investigate this issue.

---

<sup>36</sup> Note that the sample is not representative: locations receiving improved water facilities were overrepresented.

<sup>37</sup> See [http://www.who.int/water\\_sanitation\\_health/hygiene/envsan/phastep/en](http://www.who.int/water_sanitation_health/hygiene/envsan/phastep/en), accessed February 2011. It is important to note that we cannot use the Mozambique data to assess the effectiveness of PHAST since all sample households were exposed to that training. Strictly speaking we therefore only assess the effect of adding a CLTS training to PHAST instruction.

## References

- Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Chakma, T., Godfrey, S., Bhatt, J., Rao, P., Mishram, P. and Singh, S. (2008). Cross sectional health indicator survey of Open Defecation Free villages in Madhya Pradesh, India, *Waterlines Journal*, 27(3), pp. 236-247.
- Esrey, S.A., Potash, J. B., Roberts, L., and Shiff, C. (1991). Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization*, 69(5), pp. 609-621.
- Esrey, S.A. (1996). Water, Waste, and Well-Being: A Multicountry Study. *American Journal of Epidemiology*, 143(6), pp. 608 -623.
- Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., and Colford Jr., J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, 5(1), pp. 42-52.
- Godfrey, S. (2009). Preliminary Documentation and Evaluation of the Sanitation Component of the “One Million Initiative” in Mozambique.  
<http://www.communityledtotalsanitation.org/resource/preliminary-documentation-and-evaluation-sanitation-component-one-million-initiative-docume>
- GON and UNICEF (2006). *The Government of the Netherlands – UNICEF partnership: water, sanitation and hygiene. Mozambique project proposal*. The Hague and New York: GON and UNICEF. Revised July 2006.
- Independent Evaluation Group (IEG) (2008). *What works in water supply and sanitation: lessons from impact evaluation?* Washington, DC: World Bank.
- IOB (2007). *Water Supply and Sanitation Programmes Shinyanga Region, Tanzania 1990-2006*. The Hague: Ministry of Foreign Affairs.
- IOB (2008). *Support to Rural Water Supply and Sanitation in Dhamar and Hodeidah Governorates, Republic of Yemen*. The Hague: Ministry of Foreign Affairs.
- IOB (2009). *Drinking Water Supply and Sanitation Programme Supported by The Netherlands in Fayoum Governorate, Arab Republic of Egypt, 1990-2009*. The Hague: Ministry of Foreign Affairs.
- Jalan, J., and Ravallion, M. (2003). Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics*, 112(1), pp. 153-173.



- Kar, K. with R. Chambers (2008). *Handbook of Community-led Total Sanitation*. London: Plan and Brighton, Sussex: IDS.
- Pattanayak, S., Dickinson, K.L., Yang, J.-C., Patil, S.R., Praharaj, P., Poulos, C. (2007). Promoting Latrine Use: Midline Findings from a Randomized Evaluation of a Community Mobilization Campaign in Bhadrak, Orissa. Research Triangle Institute Working Paper 07-02.
- Pattanayak, S., Yang, J., Dickinson, K., Poulos, C., Patil, S., Mallick, R., Biltstein, J., Praharaj, P. (2009). Shame or Subsidy revisited: social mobilization for sanitation in Orissa, India, *Bulletin of World Health Organisation*, 87, pp. 580-587.
- UNICEF / WE Consult (2009). *Baseline survey final report*. UNICEF: Maputo.  
[http://www.unicef.org/mozambique/WASH\\_baseline\\_survey\\_UNICEF\\_270209.pdf](http://www.unicef.org/mozambique/WASH_baseline_survey_UNICEF_270209.pdf)
- UNICEF and IOB (2011). *Mid-term impact evaluation: UNICEF – Government of The Netherlands Partnership for Water, Sanitation and Hygiene ‘One Million Initiative’, Mozambique*. The Hague: Ministry of Foreign Affairs. <http://www.washblog.org/wp-content/uploads/2011/11/Nr-360-Mozambique-Report-FINAL-3.pdf>
- Waddington, H., and B. Snilstveit (2009). Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *Journal of Development Effectiveness*, 1 (3), pp. 295-335.
- WHO (n.d.). Costs and benefits of water and sanitation improvements at the global level. Retrieved October 5, 2011, from [http://www.who.int/water\\_sanitation\\_health/wsh0404summary/en/](http://www.who.int/water_sanitation_health/wsh0404summary/en/)
- WHO-UNICEF (2006). *Core Questions on Drinking-Water and Sanitation for Household Surveys*. Geneva: World Health Organization and UNICEF.
- World Bank (2006). *Impact evaluation: The experience of the independent evaluation group of the World Bank*. Washington, DC: World Bank.
- Wright, J., S. Gundry and R. Conroy (2004). Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. *Tropical Medicine and International Health*, 9 (1), pp. 106-117.

## Supplemental Material

This document contains supplemental material for the paper “Effectiveness of Large Scale Water and Sanitation Interventions: the One Million Initiative in Mozambique.” Section S1 discusses the evaluation framework in more detail. Section S2 contains descriptive statistics about the distribution of the data across the intervention groups and reported latrine use and hand-washing by gender and age categories, and reports regression results on attrition. Section S3 details regression results that were omitted from the paper due to the lack of space. Finally, section S4 gives an overview about the placement of the interventions.

### ***S1. Evaluation framework***

The pathways and barriers of disease transmission are summarized in Figure S1, which is adapted from Waddington and Snilstveit (2009)<sup>38</sup>. As the figure shows, faecal pathogens are carried into the body via fingers, flies, fields, food and unclean water. Water, hygiene and sanitation interventions reduce the risk of contracting diarrheal diseases by blocking the pathways of faecal pathogens. Sanitation interventions promoting the use of proper sanitation can provide a barrier to disease transmission from faeces to the environment (ground water, fields and flies). Hygiene interventions promoting proper hand-washing (at all critical times<sup>39</sup> using soap or ash and running water) aim to block transmission from fingers to drinking water, food and body. Water interventions aim at breaking down the transmission from water to food and body either by improving water quality at the source or providing water treatment methods at the point of use. However, any water, sanitation or hygiene intervention can only minimise risk along certain pathways. Therefore, in order to break the transmission cycle completely, a combination of the interventions is necessary. The One Million Initiative aims to achieve this using a combination of water point and CLTS sanitation training interventions.

Figure S2 (Figure 1 in the paper, reproduced here for convenience) depicts how the interventions of the One Million Initiative affect the barriers of disease transmission. The framework outlined

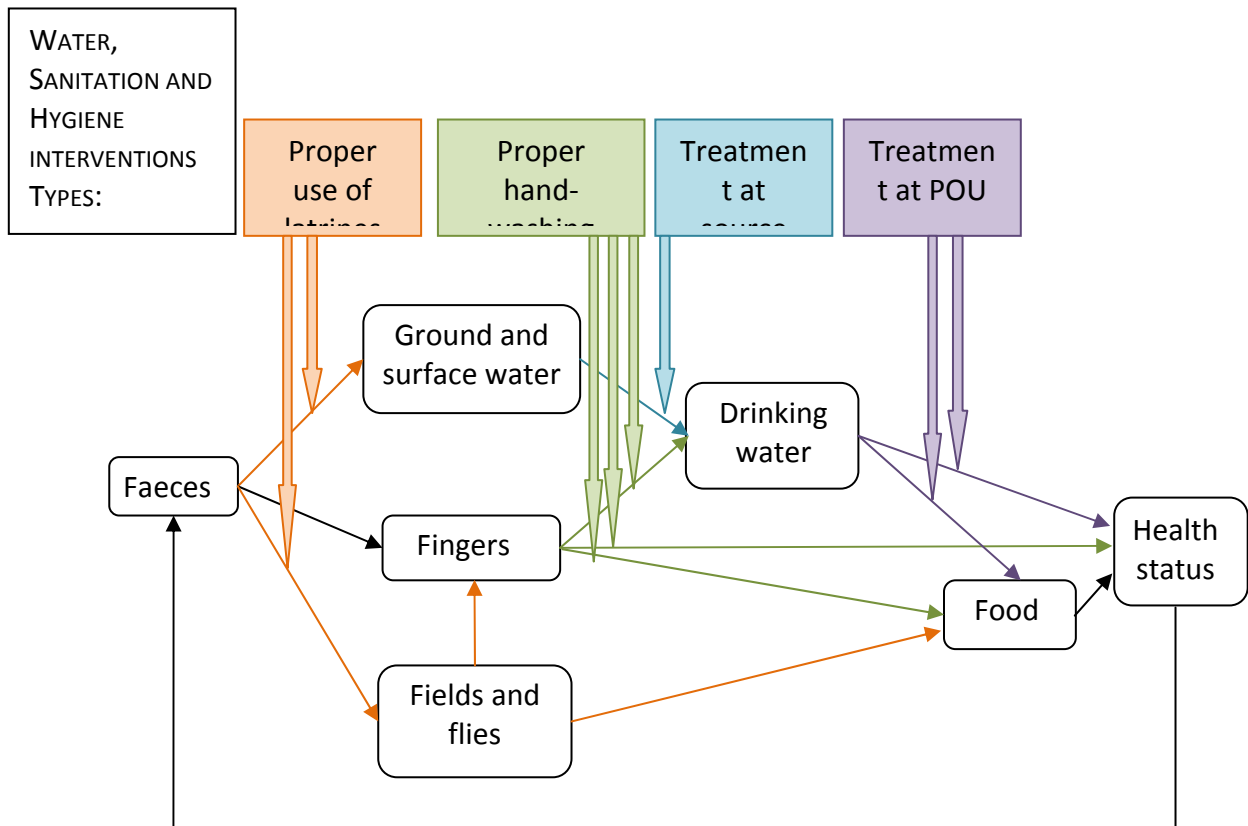
---

<sup>38</sup> Waddington, H., and B. Snilstveit (2009). Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *Journal of Development Effectiveness*, 1 (3), pp. 295-335.

<sup>39</sup> Critical times are before eating, after defecation and after disposing of baby’s feces.

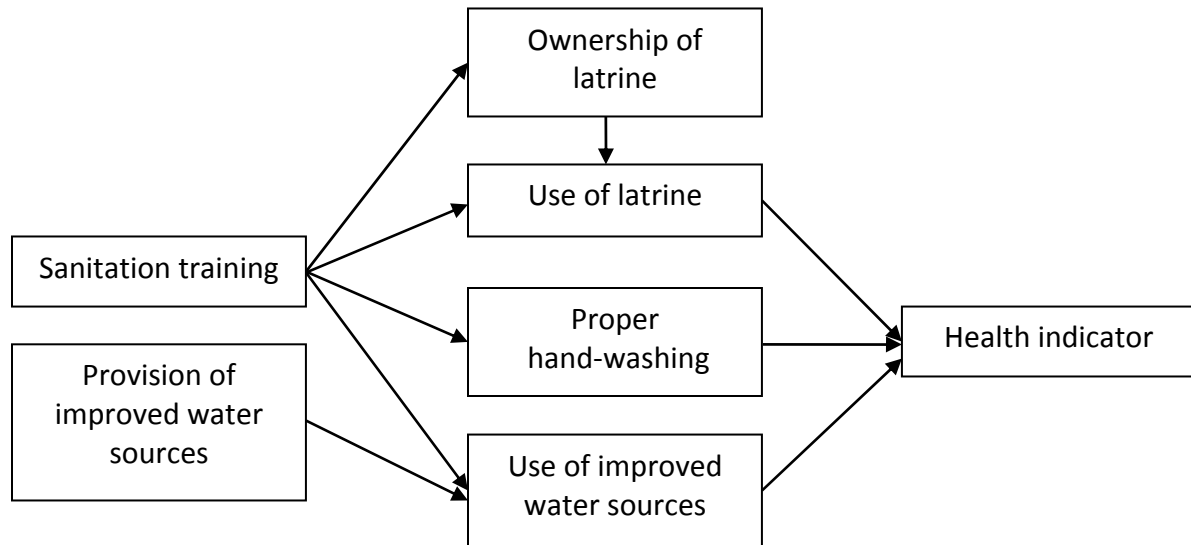
in the figure underlies the analysis in this paper. In the Figure the CLTS intervention is assumed to affect latrine usage (directly, and indirectly through increased latrine ownership) as well as proper hand washing practice. Water treatment at the source comes in the form of boreholes providing safe drinking water. Notably absent from Figure 1 is water treatment at the point of use which is not targeted by the One Million Initiative, therefore we cannot evaluate its impact. In addition, according to our survey results water treatment is not usual practice in Mozambique (see Table S2).

**Figure S1: Transmission pathways and barriers of faecal-oral disease**



Source: adapted from Waddington and Snilstveit (2009).

**Figure S2: Impact of water and sanitation interventions on health**



## ***S2. Descriptive statistics and attrition***

Table S1 reports outcomes for four different groups: those without interventions, those with only the water point intervention, those with only the CLTS interventions, and those with both interventions. The first column of the table reports the mean in the control group, and the subsequent columns report the differences relative to the control group and the significance of this difference. The table shows that in the 2008 round water related diseases were significantly more common and latrine use and ownership were significantly less common in the CLTS intervention group. This reflects the targeting of interventions at poorer villages. On the other hand, the water point intervention was more common in locations where households use improved water sources for drinking. This partly reflects the fact that some water point interventions had taken place just before the survey interview. Table S2 shows that in 2008 most other household characteristics were similar between the intervention groups. There are exceptions: wealth is lower and water treatment less common in the CLTS intervention group; as before this reflects intervention targeting. The other exceptions are education and hygiene knowledge, which in 2008 were somewhat better for the group receiving both the CLTS and the water point intervention. Table S3 summarizes community characteristics in 2008 including the median distance to the closest improved water point for the sampled households in clusters with a water point intervention. This distance variable is used in the regressions to control for treatment intensity. The remaining rows of the table show that

intervention communities differ from no-intervention communities in many respects so that it will be necessary to take these differences into account in the analysis below.<sup>40</sup>

Tables S4 and S5 summarize the reported latrine use and hand-washing practices of household members. These indicators were asked for the groups of adult males and females, school children and other children. Table S4 shows that almost all adults use a latrine if they own one, and conversely that very few households not owning a latrine report using one. This justifies the focus of the program on latrine ownership. The table also indicates that latrine use of children lags behind the use by adults. Hand-washing practices of the sampled households are shown Table S5: hand-washing before eating is universal, and almost all adults and most of the children wash their hands after defecation. However, at the baseline less than 20 percent of these households reported that they use soap or ash or running water for hand-washing.

Finally, Table S6 reports on attrition between the 2008 and 2010 survey rounds. A probit analysis of attrition probabilities shows that larger and ‘older’ households were somewhat more likely to be interviewed twice. These differences seem relatively unimportant for the present study. We therefore conclude that the survey is suitable for the impact analysis of WASH interventions provided proper care is taken of intervention targeting, either by including community dummies or by using double difference estimation.

Except for household size there is no substantial difference between the characteristics of the 1310 panel households and the other 290. All sample households and communities have now been visited twice, in 2008 and 2010, and will be visited again in 2013.

**Table S1: Comparison of outcome variables**

Variable	Year	Mean control group	Diff. water point	Diff. CLTS training	Diff. both interventions	N obs
Water related disease in household	2008	0.277 (0.448)	-0.001 (0.031)	0.170*** (0.046)	0.036 (0.028)	1588
	2010	0.183 (0.387)	-0.092*** (0.022)	-0.005 (0.036)	-0.076*** (0.021)	1574
Ownership of latrine	2008	0.479 (0.500)	-0.086*** (0.034)	0.014 (0.046)	-0.093*** (0.030)	1600
	2010	0.528 (0.500)	-0.071** (0.034)	0.129*** (0.044)	0.084*** (0.030)	1600
Use of latrine	2008	0.525 (0.500)	-0.108*** (0.034)	-0.011 (0.046)	-0.101*** (0.030)	1600
	2010	0.596 (0.491)	-0.104*** (0.034)	0.122*** (0.042)	0.040 (0.029)	1597

<sup>40</sup> Note that the differences are not statistically significant due to the small number of locations per treatment group.

Proper hand-washing	2008	0.183 (0.387)	-0.027 (0.025)	-0.026 (0.034)	-0.011 (0.023)	1600
	2010	0.377 (0.485)	-0.050 (0.033)	0.009 (0.045)	0.062** (0.030)	1599
Use of improved water source	2008	0.121 (0.326)	0.116*** (0.027)	-0.042 (0.026)	0.027 (0.021)	1600
	2010	0.206 (0.404)	0.394*** (0.032)	0.009 (0.038)	0.508*** (0.026)	1600
Share of households in both rounds		0.815	0.015	-0.008	0.005	1600
House is abandoned (share of HHs)		0.075	0.008	0.011	0.020	1600
New household in house (share of HHs)		0.033	-0.003	-0.012	0.006	1600
Number of households		720	300	140	440	1600

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S2: Comparison of selected household characteristics**

Variable	Year	Mean control group	Diff. water point	Diff. CLTS training	Diff. both interventions	N obs
Household size	2008	5.476	0.220	-0.019	0.267	1600
		(2.562)	(0.187)	(0.238)	(0.161)	
	2010	5.233	0.100	-0.262	0.078	1600
		(2.427)	(0.168)	(0.209)	(0.146)	
Number of children under 5 years	2008	0.873	0.017	-0.102	0.063	1599
		(0.887)	(0.063)	(0.082)	(0.056)	
	2010	0.864	0.103	-0.107	0.057	1598
		(0.940)	(0.070)	(0.080)	(0.057)	
Number of children under 3 years	2008	0.481	0.049	-0.066	0.081**	1600
		(0.615)	(0.047)	(0.056)	(0.040)	
	2010	0.503	0.067	-0.038	0.011	1600
		(0.648)	(0.048)	(0.056)	(0.039)	
Wealth index	2008	0.993	-0.048	-0.211***	0.011	1600
		(0.623)	(0.041)	(0.059)	(0.037)	
	2010	0.922	-0.013	-0.229***	0.034	1600
		(0.610)	(0.041)	(0.055)	(0.037)	
Mean household age	2008	23.513	-1.216	1.494	-1.351**	1599
		(11.920)	(0.750)	(1.208)	(0.680)	
	2010	24.520	-0.199	2.219	-0.171	1593
		(13.401)	(0.935)	(1.481)	(0.834)	
Female headed household	2008	0.113	-0.007	0.042	0.000	1445
		(0.316)	(0.023)	(0.035)	(0.020)	
	2010	0.152	-0.008	0.062	-0.007	1467
		(0.359)	(0.025)	(0.039)	(0.023)	
Household education	2008	0.832	0.057**	-0.044	0.051**	1586
		(0.374)	(0.023)	(0.038)	(0.021)	
	2010	0.868	-0.018	-0.082**	0.061**	1600
		(0.339)	(0.024)	(0.037)	(0.018)	
Water treatment	2008	0.089	0.008	-0.075***	0.041**	1600
		(0.285)	(0.020)	(0.015)	(0.019)	
	2010	0.124	-0.070***	0.019	0.051**	1600
		(0.329)	(0.018)	(0.032)	(0.022)	
Hygiene knowledge (maximum 5)	2008	3.396	-0.009	-0.139	0.225***	1600
		(1.185)	(0.077)	(0.111)	(0.068)	
	2010	3.776	0.110	-0.119	-0.200***	1599
		(1.030)	(0.072)	(0.098)	(0.062)	
Knowledge of practices to prevent diarrhea (max 8)	2010	3.527	-0.127	-0.320**	-0.218**	1592
		(1.546)	(0.110)	(0.123)	(0.094)	
Causes of diarrhea (max 6)	2010	2.987	-0.347***	-0.180	-0.426***	1592
		(1.338)	(0.099)	(0.115)	(0.079)	
Number of households		720	300	140	440	1600

Signif. Codes: \*' 0.1, \*\*\*' 0.05, \*\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S3: Comparison of community characteristics, 2008**

Variable	Mean control group	Diff. water point	Diff. CLTS training	Diff. both interventions	N obs
Median distance to improved water source in 2010	--	0.600 (0.462)	--	0.487 (0.402)	37
PHAST sanitation training in community	0.417 (0.500)	-0.017 (0.155)	-0.131 (0.202)	-0.144 (0.128)	80
Minutes drive from local center	72.222 (77.270)	2.111 (16.829)	7.063 (17.380)	-4.949 (24.970)	80
Mobile coverage in community	0.806 (0.401)	-0.234* (0.153)	-0.139 (0.221)	-0.139 (0.125)	77
Weekly market in community	0.194 (0.401)	-0.052 (0.118)	-0.028 (0.180)	-0.147** (0.082)	77
Primary school in community	0.806 (0.401)	0.128 (0.094)	-0.139 (0.221)	0.058 (0.100)	79
Health post in community	0.806 (0.401)	-0.139 (0.143)	0.028 (0.180)	-0.033 (0.113)	79
Size of community	2.250 (1.360)	-0.250 (0.383)	0.321 (0.575)	0.023 (0.411)	80
Number of clusters	36	15	7	22	80

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Community survey in 2008 and program data on interventions.



**Table S4: Latrine use by ownership**

<b>Latrine use by ownership (% of owners)</b>	<b>Men</b>	<b>Women</b>	<b>School children</b>	<b>Other children</b>
Household owns latrine in 2008	95.6	97.2	65.8	65.5
Household owns latrine in 2010	93.6	98.1	69.2	45.0
Household does not own latrine in 2008	6.0	6.2	3.9	4.2
Household does not own latrine in 2010	9.6	10.7	6.0	4.3

Source: Household survey, 2008 and 2010

**Table S5: Hand washing at critical times**

<b>Hand washing (% of respondents)</b>	<b>Year</b>	<b>Men</b>	<b>Women</b>	<b>School children</b>	<b>Other children</b>
Before eating	2008	100.0	100.0	99.9	96.4
	2010	99.9	100.0	99.8	98.7
After defecation	2008	96.1	96.0	84.7	66.7
	2010	96.9	95.6	83.8	73.0
After disposing baby's faeces	2008	--	89.5	74.0	55.7
	2010	--	90.0	71.5	56.3

Source: Household survey, 2008 and 2010

**Table S6: Probit estimation of attrition**

Regression	Probit (1)	Probit (2)	LPM (3)	LPM (4)
Mean dependent var	0.819	0.817	0.819	0.817
Water point intervention	0.066 (0.085)	0.106 (0.092)	0.015 (0.031)	0.024 (0.034)
Sanitation training (CLTS)	-0.049 (0.088)	-0.081 (0.095)	-0.011 (0.031)	-0.019 (0.034)
Water related disease prevalence at baseline	0.023 (0.084)	0.088 (0.091)	0.007 (0.021)	0.022 (0.021)
Use of improved water sources at baseline	-0.063 (0.106)	-0.121 (0.112)	-0.017 (0.032)	-0.033 (0.034)
Ownership of latrine at baseline	0.152** (0.077)	0.168** (0.081)	0.041** (0.023)	0.046** (0.025)
Proper hand washing practices at baseline	0.089 (0.102)	0.168 (0.110)	0.024 (0.023)	0.041 (0.025)
Household size at baseline	0.086*** (0.019)	0.106*** (0.024)	0.021*** (0.004)	0.024*** (0.005)
Number of children under 5 at baseline	-0.039 (0.047)	0.034 (0.060)	-0.009 (0.012)	0.007 (0.014)
Wealth at baseline	0.086 (0.062)	0.109 (0.069)	0.023 (0.022)	0.028 (0.023)
Female headed household at baseline		0.147 (0.134)		0.035 (0.034)
Mean age in household at baseline		0.010** (0.004)		0.002** (0.001)
Household education at baseline		0.041 (0.119)		0.016 (0.035)
Intercept	0.317*** (0.115)	-0.178 (0.230)	0.665*** (0.041)	0.547*** (0.074)
Observations	1587	1424	1587	1424
Adj. R-squared <sup>1</sup>	0.012	-0.027	0.018	0.024
Cluster robust s.e.	No	no	yes	Yes
HAC robust s.e.	Yes	yes	yes	Yes

Dependent variable: household is present in both rounds (binary)

<sup>1</sup> MacFadden's adjusted R<sup>2</sup> for probit models

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

### S3. Supplemental regression results

The tables of this section report regression results for alternative specifications: logistic regressions and results of the linear probability model omitting treatment heterogeneity due to distance. Table S7 reports these specifications for latrine ownership,

Table S8 for hand-washing practices with soap or ash and Table S9 for the use of improved water sources. Table S10 shows the importance of the distance of households to the nearest improved water source when deciding on using water from improved sources. The regression uses a distance measure calculated using the GPS coordinates of surveyed households and improved water points in the 2010 round of the survey.

The remaining tables report results on the health impact of the interventions. Table S11 shows the health impact on households aggregating over all household members, while Table S12 reports results for the households including only young children. Table S13 reports the logit regressions for the individual health outcome with differentiated treatment effects for young children and older individuals. Table S14 reports remaining specifications for the pooled IV (GMM) regressions. In the household level IV regressions two-step GMM estimation is used to allow for heteroskedasticity during the estimation procedure. Finally, Table S15 shows the IV regressions for individual household members. In this case, the large number of household dummies did not allow for the two-step feasible GMM estimation of the coefficients, therefore only the standard errors were corrected for heteroskedasticity.

**Table S7: Impact on latrine ownership**

Household FD	Household FD	Household pooled	Household FD logit	Household pooled logit
Regression	(1)	(2)	(3)	(4)
Mean dependent var	0.105	0.494	0.730	0.494
Water point intervention (wpi)	0.023 (0.046)	0.025 (0.049)	0.317 (0.727)	0.363 (0.479)
Wpi*Distance			0.367 (1.047)	0.062 (0.786)
Sanitation training (CLTS)	0.126** (0.056)	0.146** (0.057)	0.655 (0.607)	1.028** (0.454)
Intercept	0.059** (0.026)	-0.166*** (0.023)	0.781*** (0.294)	0.132 (4.041)
Observations	1310	3200	300	3200
Adj. R-squared <sup>2</sup>	0.037	0.419	0.087	0.349
First difference	Yes	No	yes	No
Year, location dummies	No	Yes	No	Yes
Household dummies	No	No	No	No
Additional regressors <sup>1</sup>	No	Yes	Yes	Yes
Cluster robust s.e.	Yes	Yes	yes	Yes
HAC robust s.e.	yes	Yes	Yes	Yes

Dependent variable: (2)-(4) latrine ownership (binary); (1) change in ownership; (3) obtained latrine=1, lost latrine=0 (binary)

<sup>1</sup> Additional regressors: household size, wealth

<sup>2</sup> MacFadden's adjusted R<sup>2</sup> for logit models

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S8: Impact on adults' hand-washing with soap or ash at crucial times**

Regression	Household FD	Household pooled	Household FD logit		Household pooled logit	
	(1)	(2)	(3)	mean effect	(4)	mean effect
Mean dependent var	0.209	0.146	0.780		0.146	
Water point intervention (wpi)	0.0002 (0.058)	-0.004 (0.053)	-0.340 (0.623)	-0.057	-0.110 (0.496)	-0.019
Wpi*Distance			0.694 (0.819)	0.116	0.377 (0.598)	0.064
Sanitation training (CLTS)	0.099* (0.059)	0.071 (0.052)	0.510 (0.422)	0.085	0.309 (0.359)	0.052
Intercept	0.165*** (0.036)	0.124*** (0.023)	1.045*** (0.237)	0.175	-1.878 (2.127)	-0.318
Observations	1309	3199	490		2890	
Adj. R-squared <sup>2</sup>	0.011	0.130	-0.001		0.097	
First difference	Yes	no	Yes		No	
Year, location dummies	No	yes	No		Yes	
Household dummies	No	no	No		No	
Additional regressors <sup>1</sup>	No	yes	Yes		Yes	
Cluster robust s.e.	Yes	yes	Yes		yes	
HAC robust s.e.	Yes	yes	yes		Yes	

Dependent variable: (2)-(4) hand-washing with soap/ash after defecation and before eating (binary); (1): change in variable, (3) 1=adopted, 0=abandoned (binary)

<sup>1</sup> Additional regressors: household size, wealth

<sup>2</sup> MacFadden's adjusted R<sup>2</sup> for logit models

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S9: Impact on use of improved water sources**

Regression	Household FD	Household pooled	Household FD logit		Household pooled logit	
	(1)	(2)	(3)	mean effect	(4)	mean effect
Mean dependent var	0.277	0.283	0.931		0.283	
Water point intervention (wpi)	0.320*** (0.088)	0.330*** (0.087)	77.741*** (20.164)	2.027	5.355*** (1.583)	0.382
Wpi*Distance			-84.614*** (25.351)	-2.207	-5.402*** (2.071)	-0.385
Sanitation training (CLTS)	0.158* (0.092)	0.142 (0.090)	83.341*** (27.680)	2.174	1.088 (1.131)	0.078
Intercept	0.069 (0.043)	-0.028 (0.028)	1.679 (1.402)	0.044	-20.396** (9.433)	-1.455
Observations	1310	3200	421		3200	
Adj. R-squared <sup>2</sup>	0.173	0.562	0.602		0.571	
First difference	yes	no	yes		No	
Year, location dummies	no	yes	no		Yes	
Household dummies	no	no	no		No	
Additional regressors <sup>1</sup>	no	yes	yes		Yes	
Cluster robust s.e.	Yes	yes	yes		Yes	
HAC robust s.e.	yes	yes	yes		Yes	

Dependent variable: (2)-(4) use of improved water source (binary); (1): change in variable, (3) 1=adopted use, 0=discontinued use (binary)

<sup>1</sup> Additional regressors: household size, wealth

<sup>2</sup> MacFadden's adjusted R<sup>2</sup> for logit models

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S10: Distance to improved water points as a determinant of their use**

**Dependent variable:** 1 if household used improved source in 2010. 0 otherwise

Mean dependent variable: 0.42

Coefficient	Estimate	Clustered Std. error
Intercept	0.784 ***	0.045
Improved source 10-250 m further away	-0.351 ***	0.083
Improved source 250-500 m further away	-0.55 ***	0.074
Improved source 500-999 m further away	-0.727 ***	0.052
Improved source 1 – 2 km further away	-0.641 ***	0.072
Improved source 2 - 3 km further away	-0.784 ***	0.045
Improved source 3 - 4 km further away	-0.764 ***	0.047

Signif. codes: '\*\*\*' 0.01, '\*\*' 0.05 '\*' 0.1

Source: Household Survey, 2010 round. Number of observations 1222. Adjusted R-squared 0.33.

**Table S11: Impact on health outcome**

Regression	Household FD	Household pooled	Household FD logit		Household pooled logit	
	(1)	(2)	(3)	mean effect	(4)	mean effect
Mean dependent var	-0.163	0.223	0.264		0.223	
Water point intervention (wpi)	-0.008 (0.039)	-0.039 (0.039)	-0.132 (0.386)	-0.024	-0.449 (0.315)	-0.068
Wpi*Distance			-0.184 (0.649)	-0.034	-0.139 (0.341)	-0.021
Sanitation training (CLTS)	-0.081 (0.042)	-0.079** (0.040)	-0.525* (0.285)	-0.096	-0.447* (0.237)	-0.068
Household size	0.027*** (0.008)	0.013*** (0.005)	0.131* (0.069)	0.024	0.079*** (0.030)	0.012
Number of children under 5	0.030 (0.020)	0.032*** (0.010)	0.246 (0.160)	0.045	0.206*** (0.062)	0.031
Wealth	-0.023 (0.031)	-0.008 (0.015)	-0.192 (0.219)	-0.035	-0.044 (0.095)	-0.007
Intercept	-0.119*** (0.030)	0.454*** (0.030)	-0.730*** (0.181)	-0.134	-0.302 (0.751)	-0.046
Observations	1280	3162	440		3162	
Adj. R-squared <sup>2</sup>	0.024	0.093	0.022		0.065	
First difference	Yes	no	yes		No	
Year, location dummies	No	yes	no		Yes	
Household dummies	No	no	no		No	
Additional regressors <sup>1</sup>	No	yes	yes		Yes	
Cluster robust s.e.	Yes	yes	yes		Yes	
HAC robust s.e.	Yes	yes	yes		Yes	

Dependent variable: (2)-(4) reported prevalence of water related disease in household in last 6 months (binary); (1): change in disease prevalence; (3) 1=disease in 2010 and not in 2008, 0= disease in 2008 and not in 2010 (binary)

<sup>1</sup> Additional regressors: household size, number of children under 5. Wealth

<sup>2</sup> MacFadden's adjusted R<sup>2</sup> for logit models

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S12: Impact on children's health outcome**

Regression	Children under 5	Children under 5	Children under 3	Children under 3		
	pooled (1)	pooled logit (2)	mean effect	pooled (3)	pooled logit (4)	mean effect
Mean dependent var	0.251	0.251		0.146	0.146	
Water point intervention (wpi)	-0.024 (0.046)	-0.344 (0.385)	-0.056	-0.103** (0.046)	-1.625** (0.782)	-0.176
Wpi*Distance		-0.132 (0.404)	-0.021		0.712 (1.271)	0.077
Sanitation training (CLTS)	-0.071 (0.046)	-0.323 (0.287)	-0.053	-0.005 (0.050)	0.057 (0.461)	0.006
Household size	0.019*** (0.007)	0.109*** (0.041)	0.018	0.009 (0.006)	0.081 (0.050)	0.009
Number of children in category	0.003 (0.018)	0.024 (0.106)	0.004	-0.001 (0.028)	-0.045 (0.229)	-0.005
Wealth	0.011 (0.019)	0.066 (0.113)	0.011	0.003 (0.020)	0.015 (0.173)	0.002
Intercept	0.481*** (0.055)	-0.217 (0.793)	-0.035	0.416*** (0.048)	-0.475 (3.411)	-0.052
Observations	1798	1798		1366	1366	
Adj. R-squared <sup>2</sup>	0.083	0.031		0.055	-0.009	
First difference	no	No		no	No	
Year, location dummies	yes	Yes		yes	Yes	
Household dummies	no	No		no	No	
Additional regressors <sup>1</sup>	yes	Yes		yes	Yes	
Cluster robust s.e.	yes	Yes		yes	Yes	
HAC robust s.e.	yes	Yes		yes	Yes	

Dependent variable: (4)-(5) reported prevalence of water related disease for children in age category in the household in last 6 months (binary); (1)-(2) change in disease prevalence; (3) 1=disease in 2010 and not in 2008, 0= disease in 2008 and not in 2010 (binary)

<sup>1</sup> Additional regressors: household size, number of children under 5. Wealth

<sup>2</sup> MacFadden's adjusted R<sup>2</sup> for logit models

Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01

Source: Household survey and program data on interventions, 2008 and 2010.

**Table S13: Impact on individual health outcome (logit regressions)**

Regression	Children defined as under 5		Children defined as under 3	
	(1)	Mean effect	(2)	Mean effect
Mean dependent variable	0.071		0.071	
Water point intervention for children in age category	-0.422 (0.391)	-0.019	-0.048 (0.439)	-0.002
Water point intervention for older individuals	-0.661** (0.331)	-0.029	-0.706** (0.317)	-0.031
Sanitation training (CLTS) for children in age category	-0.686 (0.451)	-0.03	-0.806* (0.485)	-0.036
Sanitation training (CLTS) for older individuals	-0.850** (0.413)	-0.037	-0.813** (0.412)	-0.036
Child in age category	1.069*** (0.191)	0.047	1.257*** (0.243)	0.055
Household size	0.100 (0.111)	0.004	0.134 (0.101)	0.006
Number of children in age category	-0.039 (0.212)	-0.002	-0.279 (0.211)	-0.012
Wealth	-0.054 (0.331)	-0.002	-0.066 (0.328)	-0.003
Intercept	-0.891*** (0.308)	-0.039	-0.421 (0.269)	-0.019
Observations	17150		17132	
Adj. R-squared <sup>2</sup>	0.017		0.017	
First difference	No		No	
Year, location dummies	Yes		Yes	
Household dummies	Yes		Yes	
Additional regressors <sup>1</sup>	Yes		Yes	
Clustered standard errors	Yes		Yes	
Dependent variable: (2) reported prevalence of water borne disease for individuals in last 6 months (binary); (1) change in disease prevalence.				
<sup>1</sup> Household size, number of children under 5, wealth.				
<sup>2</sup> McFadden's adjusted R <sup>2</sup> for logit models.				
Significance: '*' 0.1, '**' 0.05, '***' 0.01				
Source: Household survey and program data on interventions, 2008 and 2010.				



**Table S14: Health effect of the use of improved water sources and latrines using pooled data (GMM estimation)**

Regression	Household pooled (1)	Household pooled (2)
Mean dependent var	0.224	0.224
Use of improved water source	-0.109 (0.109)	-0.128* (0.068)
Use of latrine	-0.465 (0.337)	-0.393** (0.179)
Household size		
Number of children under 5	0.017*** (0.005)	0.016*** (0.004)
Wealth	0.026** (0.011)	0.027*** (0.010)
Intercept	0.05 (0.047)	0.04 (0.027)
Observations	3158	3158
F-stat, first stage (water)	182.7	51.5
F-stat, first stage (latrine)	21.4	36.3
J-test, degrees of freedom, probability		0.066, 2, 0.97
List of additional instruments <sup>1,2</sup>		Water×FB, CLTS×SLB
First difference	no	No
Year, location dummies	yes	Yes
Household dummies	no	No
Cluster robust s.e.	no	No
HAC robust s.e.	yes	Yes
Dependent variable: reported prevalence of water related disease in household in last 6 months (binary)		
<sup>1</sup> Instrument list always includes Water, CLTS, HH size, number of children under 5, wealth.		
<sup>2</sup> Baseline controls: FB is functioning borehole in cluster; and SLB is the share of latrines in cluster (excluding hh)		
Signif. Codes: '***' 0.01. '**' 0.05. '*' 0.1		
Source: Household survey and program data on interventions, 2008 and 2010.		

**Table S15: Impact of use of improved water sources and latrines on individual health outcome  
(IV estimation)**

Regression	Children under 5	Children under 3
	pooled (1)	pooled (2)
Mean dependent var	0.071	0.071
Use of improved water source for children in age category	-0.096** (0.042)	-0.127** (0.061)
Use of improved water source for older individuals	0.025 (0.029)	0.016 (0.029)
Use of latrine for children in age category	-0.128 (0.140)	-0.011 (0.219)
Use of latrine for older individuals	-0.279*** (0.089)	-0.263*** (0.090)
Child in age category	0.017 (0.054)	-0.014 (0.089)
Household size	0.007*** (0.002)	0.009*** (0.002)
Wealth	-0.0002 (0.005)	-0.009 (0.006)
Intercept	0.018* (0.010)	0.016 (0.011)
Observations	17150	17132
F-stat, first stage (water, child)	18.6	21.6
F-stat, first stage (water, older)	11.1	10.3
F-stat, first stage (latrine, child)	23.8	28.3
F-stat, first stage (latrine, older)	13.9	13.0
First difference	No	No
Year, location dummies	yes	Yes
Household dummies	yes	Yes
Cluster robust s.e.	no	No
HAC robust s.e.	yes	Yes

Dependent variable: (2) reported prevalence of water related disease for individuals in last 6 months (binary); (1) change in disease prevalence  
Signif. Codes: '\*' 0.1, '\*\*' 0.05, '\*\*\*' 0.01  
Source: Household survey and program data on interventions, 2008 and 2010.

#### **S4. Intervention placement**

The interventions in the One Million Initiative are purposefully not randomized across communities in the program area. Emphasis is put on reaching poorer, more vulnerable communities that do not have or have only limited access to safe water sources and sanitary facilities. This appendix describes the selection process of the intervention communities in the program districts.

In general, the district government decides on the location of water point interventions (rehabilitation and construction of new sources) based on information collected by an NGO responsible for implementing the Community Participation and Education (PEC) in the district. PEC consists of three components: (1) community mobilization and hygiene training, (2) water committee training, and (3) sanitation training. The PEC NGOs have the crucial role of communicating the program to the communities, and the situation in target areas to the local authorities. Information over the needs of communities is collected by inviting communities to formally apply for a water point intervention under the first component of PEC. Local authorities analyze and prioritize the applications regularly. Priority areas are set yearly by the government. 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.

The PEC NGO is also responsible for carrying out the water, sanitation and hygiene trainings of the program. Hygiene training is done in every community the PEC NGO visits (therefore in all communities in the survey sample) with particular attention to proper hand washing practices. A water committee training is carried out at locations where there is an improved water source or after the water point intervention was implemented. The training focuses on the importance of managing the improved water source and its maintenance through a dedicated water committee.<sup>41</sup>

The PEC NGOs can decide on the locations for the sanitation trainings, perhaps with the assistance of the local government. The sanitation component of the One Million Initiative from mid 2008 is Community Approach to Total Sanitation (CATS). This approach combines the Community Led Total Sanitation (CLTS) trainings with a reward system for communities that become open defecation free (ODF). PEC NGOs are rewarded for the number of ODF communities in their district, therefore it is in their interest to introduce CLTS at locations

---

<sup>41</sup> Setting up a water committee is also a precondition for applying for a water point intervention.

where their likelihood of success is high. Based on Kar et al. (2008) the success of CLTS is related to factors including health problems, leadership, size of community and geographical factors.