# Access to safe drinking water: experimental evidence from new water sources in Bangladesh

**Serena Cocciolo**, Institute for International Economic Studies

**Selene Ghisolfi**, Institute for International Economic Studies

**Ahasan Habib**, NGO Forum for Public Health

**SMA Rashid**, NGO Forum for Public Health

**Anna Tompsett**, Stockholm University

**Grantee Final Report**

**Accepted by 3ie: October 2019**

International Initiative for Impact Evaluation

## Note to readers

This impact evaluation has been submitted in partial fulfilment of the requirements of grant DPW1.1006 issued under Development Priorities Window 1. This version is being published online as it was received. A copy-edited and formatted version will be available in the 3ie Impact Evaluation Report Series in the near future.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Anna Tompsett at anna.tompsett@ne.su.se.

The 3ie technical quality assurance team comprises Rosaine Yegbemey, Kanika Jha Kingra, Sayak Khatua, Marie Gaarder, an anonymous external impact evaluation design expert reviewer and an anonymous external sector expert reviewer, with overall technical supervision by Marie Gaarder.

Funding for this impact evaluation was provided by UK aid through the Department for International Development. A complete listing of all of 3ie's donors is available on the 3ie website.

Suggested citation: Cocciolo, S, Ghisolfi, S, Habib, A, Rashid, SMA and Tompsett, A, 2019. *Access to safe drinking water: experimental evidence from new water sources in Bangladesh, 3ie Grantee Final Report*. New Delhi: International Initiative for Impact Evaluation (3ie)

## Acknowledgements

We acknowledge financial support from the International Initiative for Impact Evaluation (3ie), as well as technical review and support throughout the study. We also acknowledge financial support from the Swedish Research Council and the International Growth Centre (the IGC). In Bangladesh, we thank Md. Rezwanul Karim, Jahirul Islam, Sabbir Ahmed, Rezaul Karim, Ranjan Kumar Mohonto, Md. Tariqul Islam, Mir Abu Raihan, Tahmid Sharif, and Md. Shahadat Hossain for exemplary research assistance and support. In Sweden, we thank Cristina Altomare, Tillmann von Carnap, Merve Demirel, and Lorenzo Schirato for excellent research assistance.

# Contents

## List of Figures



## List of Tables

## Abbreviations and acronyms

| | |
|---|---|
| BAMWSP | Bangladesh Arsenic Mitigation Water Supply Project |
| BBS | Bangladesh Bureau of Statistics |
| BD | Bangladesh |
| BDT | Bangladeshi Taka |
| BSDS | Bangladesh Social Development Services |
| DPHE | Department of Public Health Engineering |
| HH | Household |
| IGC | International Growth Centre |
| IV | Instrumental Variable |
| MICS | Multiple Indicator Cluster Survey |
| NGO | Non-governmental organization |
| ppb | Parts per billion |
| SRC | Swedish Research Council |
| Treatment Unit | A community of between 50 and 250 households |
| WS | Water source |
| WHO | World Health Organization |

# Executive summary

**We evaluate the impact of a program to provide safe sources of drinking water in rural Bangladesh.** The program consists of a package of subsidies and technical advice to install deep tubewells. Deep tubewells access water from deep aquifers which is free from fecal contamination and from arsenic contamination, a major naturally occurring problem in much of rural Bangladesh.

**The program installs wells that provide water which is almost free from arsenic contamination, but not from fecal contamination.** The program successfully installs new water sources, installing a total of 107 tubewells in 129 communities. In communities where we successfully install new water sources, the median household is 1.6 minutes walk from the new source. The wells installed provide arsenic safe drinking water, essentially eliminating arsenic contamination at source for those who adopt the wells. However, 34% of installed tubewells unexpectedly test positive for fecal contamination, compared to 46% of other tubewells in the same communities, suggesting that the program wells only reduce exposure to fecal contamination by around 26%. This result is unexpected because the source water is isolated from fecal contamination, meaning that fecal contamination must occur either because shallow groundwater enters the tubewell system or through the pump body. The test for fecal contamination is coarse, however, meaning that we cannot evaluate whether levels of fecal contamination are lower in the installed wells, only whether fecal contamination is present. We may therefore not fully capture the reduction in exposure to fecal contamination at source.

**The program reduces arsenic contamination in household drinking water, but not fecal contamination.** Each tubewell installed under the program leads to a reduction in contamination of household drinking water equivalent to elimination of arsenic contamination at the WHO level for about 5 households, but an increase in fecal contamination equivalent to introducing fecal contamination into drinking water for about 2 households (although we cannot reject a small reduction or no effect on fecal contamination in household drinking water).

**Modest improvements in source water quality with respect to fecal contamination are offset by increases in travel time and possibly by changes in storage behaviour.** The program somewhat improves fecal contamination at the source level but also slightly increases travel times and induces small changes in storage behaviour, both of which increase the risk of fecal contamination in drinking water. Our best estimates suggest that walking an extra minute to collect drinking water increases the risk of fecal contamination by around 1.7% while storing drinking water in the house increases the risk of fecal contamination by around 7%. The consequences of these negative effects are modest, however, because few households walk more than a minute to collect drinking water, and the majority of households do not change their storage behaviour as a result of the intervention.

**Key takeaways** Our results suggest that while deep tubewells can feasibly provide arsenic-safe water in rural Bangladesh, deep tubewell construction programs may have a limited effect on fecal contamination. These results allay fears that deep tubewell construction programs may substantially increase exposure to fecal contamination. However, they also suggest that construction of deep tubewells is insufficient to resolve the fecal contamination problem in villages in rural Bangladesh, in the absence of improvements to tubewell design or maintenance practices. Future research is needed to quantify the extent of fecal contamination in deep tubewells and to understand the channels for contamination of deep tubewells.

# Chapter 1

# Introduction

Sustainable Development Goal 6 sets out the challenge of ensuring availability and sustainable management of water and sanitation for all. However, access to safe drinking water remains limited, particularly in rural areas where safe sources may be few and far between. In 2015, 663 million people worldwide still lacked access to improved sources of drinking water; 1.8 billion people drank fecally-contaminated water; and 1000 children a day died from diarrheal disease, associated with poor water quality and sanitation (United Nations, 2016).

In Bangladesh, the focus of this evaluation, the problem of access to safe drinking water is particularly acute. In the 1970s and 1980s, infant mortality in Bangladesh was extremely high, largely as a result of high levels of diarrheal disease resulting from fecal contamination of surface water, used for drinking (Caldwell et al., 2003). Education campaigns encouraged people to shift to obtaining drinking water from groundwater sources instead, followed by a decline in child mortality (Caldwell et al., 2003). However, in the 1990s high but naturally-occurring levels of arsenic were discovered in the groundwater. Arsenic is undetectable without water quality tests. By the time the arsenic contamination problem was discovered, an epidemic of diseases associated with arsenic exposure was already established, called "the largest poisoning of a population in history" (Smith et al., 2000). Despite years of effort by the Bangladeshi government, non-governmental organizations and international aid agencies, progress on safe drinking water in Bangladesh remains elusive (Human Rights Watch, 2016). Today, almost 100 million people still drink fecally-contaminated water, and 39 million people drink water that is contaminated with arsenic at international standards (BBS and UNICEF, 2015).

The magnitude of the problem of providing access to safe drinking water is clear. With respect to arsenic contamination, the remedy is technically straightforward, albeit costly: switching to an arsenic-safe source of drinking water. However, with respect to the reduction of exposure to fecal contamination, there is far less consensus regarding the potential solutions. Drinking water may be contaminated with pathogens at source, during transport from the source, or during storage (Wright et al., 2004). Disentangling these channels empirically is difficult because households that live nearer safe water sources likely differ in other respects that also affect their drinking water quality e.g. income or education. As a result, prior evidence is mixed as to which of these channels is most important in determining bacterial contamination of household drinking water (e.g. Fewtrell et al., 2005; Clasen et al., 2006). Further, in Bangladesh, recent studies raise the concern that efforts to reduce exposure to arsenic have had the unintended consequence of increasing bacterial contamination of drinking water, via increased transport and storage times associated with the use of more distant, arsenic-safe water sources (Field et al., 2011; Wu et al., 2011).

These uncertainties make it more difficult to design effective interventions to improve access to safe drinking water. In particular, they raise the risk that providing safer but more distant sources may increase exposure to pathogens via contamination in transport. These questions are particularly salient in Bangladesh, where policy-makers must design policy to reduce exposure to arsenic contamination without increasing exposure to fecal/bacterial contamination. Our evaluation measures the impact of a program of new safe drinking water source construction in rural Bangladesh and goes on to measure the impacts of source water quality and transport time on household water quality in the same context.

**Overview** Chapter 2 provides a brief overview of the intervention, the theory of change and the research hypotheses. Chapter 3 describes the context and Chapter 4 outlines the timeline. Chapter 5 describes the evaluation design, methods and implementation. Chapter 6 provides a more detailed description of the programme evaluated. Chapter 7 describes the analysis and the results of the impact evaluation and Chapter 8 discusses these results. Chapter 9 draws conclusions for policy and practice.

# Chapter 2

# Intervention, theory of change and research hypotheses

## 2.1 The intervention

We evaluate the effects of a program designed to improve access to safe drinking water in rural Bangladesh. The program consists of a package of subsidies and technical advice to build new sources of water, which are intended to provide drinking water that is free of both arsenic and bacterial contamination. The program is fully implemented by the Bangladeshi NGO "NGO Forum for Public Health", who is a partner on this impact evaluation.

The new safe sources of water are deep tubewells, which draw water from aquifers that are sufficiently deep to be safe from both bacterial contamination and arsenic contamination. In rural Bangladesh, deep tubewells are the most commonly proposed and implemented solution to the arsenic contamination problem. After installation, we test all sources to confirm that the water is indeed arsenic free.[1] Table 2.1 describes the program in a logical framework.

The subsidies range in value from 90 to 100% of the cost of water source installation. Communities decide the location of new water sources by unanimous consensus in community meetings. We carry out the intervention in "treatment units", consisting of groups of between 50 and 250 households, dividing larger villages into several treatment units along natural boundaries. We refer to "treatment units" or "communities" interchangeably throughout the document. We offer to install one new water source in smaller treatment units, and two new water sources in larger treatment units.[2]

[1]We also confirm that the sources are manganese free; manganese is another drinking water pollutant that affects some areas of Bangladesh.

[2]We designed the rules to allocate tubewells to achieve the goals of a parallel study regarding the effect of group size on collective action. Specifically, we implemented one of two rules: i) we assigned tubewells to villages as a function of village size, then divided these among the designated treatment units within each village; ii) we assigned tubewells to treatment units to keep the ratio of households to tubewells as close as possible to 125:1.

Table 2.1: Logical framework

| | Summary | Indicators | Means of verification | Assumptions |
|---|---|---|---|---|
| Impact | Improvement in drinking water quality | Arsenic contamination in household drinking water<br>Fecal contamination in household drinking water | Water testing program | Improved water quality leads to improved health |
| Outcomes | Adoption of new drinking water sources | Number of households who report using new sources | Follow-up survey | Adoption of new sources leads to improved household water quality (see Section 2.4) |
| Outputs | Construction of new drinking water sources | Number of sources constructed | Project records | |
| Activities | Safe drinking water program (see Section 2.1 for details) | | | |
| Inputs | Subsidies, technical advice and support, community engagement/participation | | | |

Installation costs on average 60,000 Bangladeshi Taka per deep tubewell.[3] We assigned communities to one of three contribution requirements. Communities assigned to the **cash contribution** approach are required to raise 6,000 Bangladeshi Taka per installed water source, and the decision on how to divide this amount among households is entirely delegated to the community. Communities assigned to the **labour contribution** approach are responsible for providing a total of 18 person-days of labour to assist the mason group in the installation work. Each person-day corresponds to a 6-hour shift, consistent with local norms for unskilled labour, and is valued at 300 Bangladeshi Taka,[4] or a total of 5,400 Bangladeshi Taka, similar to the cash contribution requirement.[5] The implementation of cash and labour contributions rule are designed in order to maximize comparability between the two treatments.[6] Communities assigned to the **contribution waiver** treatment received the program without a required contribution.

A key feature of program delivery is the active involvement of targeted communities in the decision-making process regarding: (i) how many water sources to install in the community; (ii) where to construct them; (iii) how to divide the required contributions between households; (iv) which households should take responsibility for the management and maintenance of each new water source. Communities take all decisions at meeting(s) organized by project staff at which project staff play a strong facilitatory role. We impose minimum participation requirements to hold a community meeting, and require that all decisions are taken by unanimous consensus, during the meeting in the presence of project staff. We did not implement the project in communities where an agreement is not found after a maximum of three community meetings.[7] The rules and procedures imposed on the decision-making process are designed to encourage participation, reduce the likelihood that influential groups or individuals could co-opt the decision-making process, and ensure that everyone is guaranteed the right to express his/her voice, at least *de jure*.

Implementation of the program was carried out between March 2016 and August 2017, with some piloting beginning in October 2015. More details are given in Chapter 4.

## 2.2 Theory of change

Many programs aim to improve access to safe drinking water by providing new, safe sources of drinking water. A simple theory of change underlies these programs: new sources are built; households adopt the new sources; source water quality improves; and thereby household water

---

[3]Exchange rate approximately 80 Bangladeshi Taka to 1 US$.

[4]The average daily unskilled wage in rural Bangladesh.

[5]The contractor is paid 3,000 Bangladeshi Taka less per deep tubewell under the labour contribution approach. The unskilled labour provided by the community does not fully substitute for the relatively skilled labour required by the contractor.

[6]In case of installation failures due to hydrogeological constraints, we returned cash contributions to households and we compensate households contributing labour with 300 Bangladeshi Taka per man per shift.

[7]In practice, only one community failed to reach an agreement. In this community, they declined to hold further meetings after a second meeting was unsuccessful in reaching agreement.

quality also improves. Figure 2.1a illustrates this simple theory of change.

However, a more nuanced theory of change recognises that source water quality is only one of the determinants of household water quality,[8] and that household water quality is also affected by transport distances and storage practices. Specifically, longer transport and storage times provide more opportunities for recontamination between the point of collection and the point of use, decreasing household water quality.[9] Figure 2.1b illustrates this more complete theory of change, which acknowledges that not all households will adopt a new source; that among those who do adopt a new source, many will also alter their transport and storage practices; and that these changes in storage and practice may in turn have separate effects on household water quality.

Households that adopt new sources may either increase or decrease their transport distances and storage times. Households who adopt the new source because it is closer than their previous source will decrease their transport distance and possibly reduce storage time. Households may also adopt the new source when it is further away than their previous source, if the new source is better quality than their previous source. These households will increase their transport distance and likely store water for longer.[10]

Providing new sources of safe drinking water may therefore have unintended conse- quences for some households, depending on how they change transport times and storage practice in response. In particular, the gains from improvements in source quality may be partially or even completely offset by the increases in contamination via transport and storage, for those who increase transport and storage time as a result of adopting the new source. Consistent with this more complete model of change, the closest previous randomized evaluation, Kremer et al. (2011), found that reducing source contamination by 66% only reduced household water contamination by 24%, interpreting these findings as evidence for recontamination via transport and storage. In the Bangladeshi context, Field et al. (2011) raised the concern that actions taken to reduce exposure to arsenic may have increased exposure to bacterial contamination, as households switched from using nearby arsenic-contaminated wells to more distant arsenic-safe wells.

An additional key assumption that underlies the standard model is that a substantial number of individuals will adopt new sources of safe drinking water. However, there is considerable evidence from our own previous work (Madajewicz et al., 2018), from anecdotal evidence, and from other studies (e.g. Wu et al., 2011; Human Rights Watch, 2016) that new sources do not always translate into widespread changes in use. This may be because of inherent preferences for

---

[8]For example, Wright et al. (2004) point out a systematic and considerable gap between source water quality and household water quality.

[9]Related, Waddington and Snilstveit (2009) conclude that the most effective interventions to reduce diarrhea are those that reduce bacterial contamination at point of use.

[10]We observed these effects in our previous study (Madajewicz et al., 2018). Households that switched from unsafe to safe water increased the distance they walked to collect safe water by approximately 50% on average. Households that used safe water at both baseline and follow-up on average decreased the distance they walked to safe water, because some of these houses switched from more distant safe sources to new, nearer safe sources.

local sources,[11] or awareness of the risks of transporting water over greater distances and storing water for longer time periods. However, it may also be because new sources are built in places that favor use by elites, rather than the community as a whole, or because elites or landowners explicitly restrict use of the source.

Figure 2.1a: Theory of Change: Naïve model

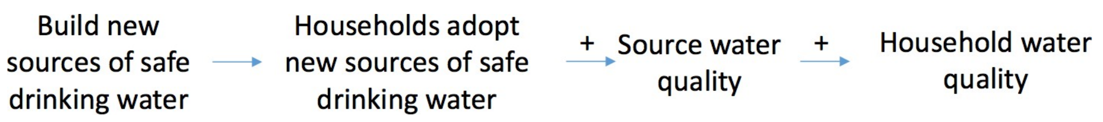
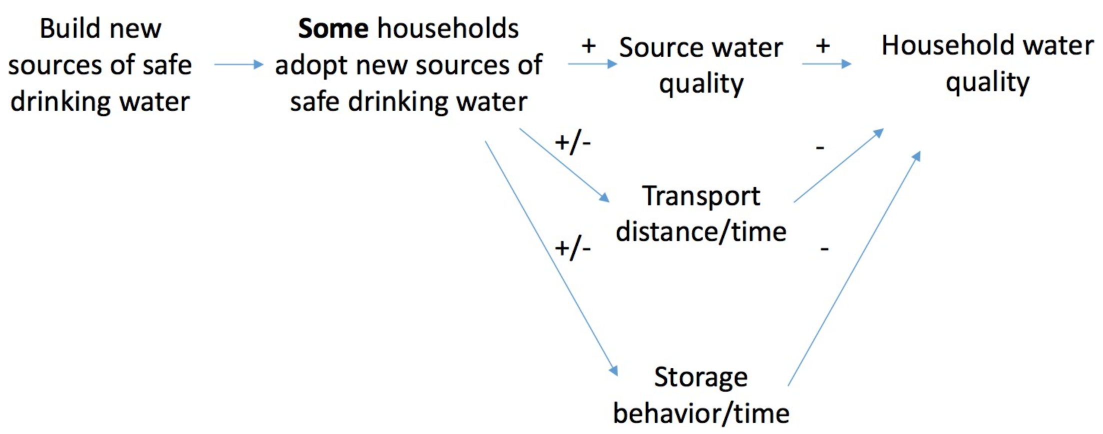



Figure 2.1b: Theory of Change: More complete model




Finally, we should note that although the theory of change we outline here relates source construction to household water quality, the real policy objective is usually improved health. We focus in this study on household contamination for two reasons. First, our measures of drinking water quality have the benefit of being largely objective measures of project impact, and less susceptible to reporting bias. Second, health changes in response to reduced arsenic exposure, in particular, will be difficult to detect because arsenic is a cumulative pollutant in the body, meaning that health consequences are the result of life-time exposure.

[11]In our current study, of those using some unsafe water at baseline, 94% reported that they would switch to a new source if it was within 1 minute of their compound, but only 3% would switch if it was 7 minutes walk from their compound. Observed adoption rates for new sources were considerably lower than this.

## 2.3 Key research questions

The evaluation questions of our study are the following. These questions are the same as those outlined in our pre-analysis plan.

1. What is the average effect of the program on household water quality, measured by:
   (a) arsenic contamination in drinking water?
   (b) fecal contamination in drinking water?
2. How does the program change behaviour with respect to obtaining water for drinking and cooking?
   (a) What is the average effect of the program on water quality of the source used by the household, measured by source arsenic contamination?
   (b) What is the average effect of the program on water quality of the source used by the household, measured by source fecal contamination?
   (c) What is the average effect of the program on distance walked to collect water?
   (d) What is the average effect of the program on household water storage practices?
3. What is the causal effect of the behavioral channels on household water quality?
   (a) What is the causal effect of water source quality on household water quality?
   (b) What is the causal effect of transport distance on household water quality?
   (c) What is the causal effect of storage practice on household water quality?

We note that the average effects may conceal considerable heterogeneity. For example, households who adopt new sources may in principle either reduce or increase their transport times. We explore this heterogeneity in Section 7.5

# Chapter 3

# Context

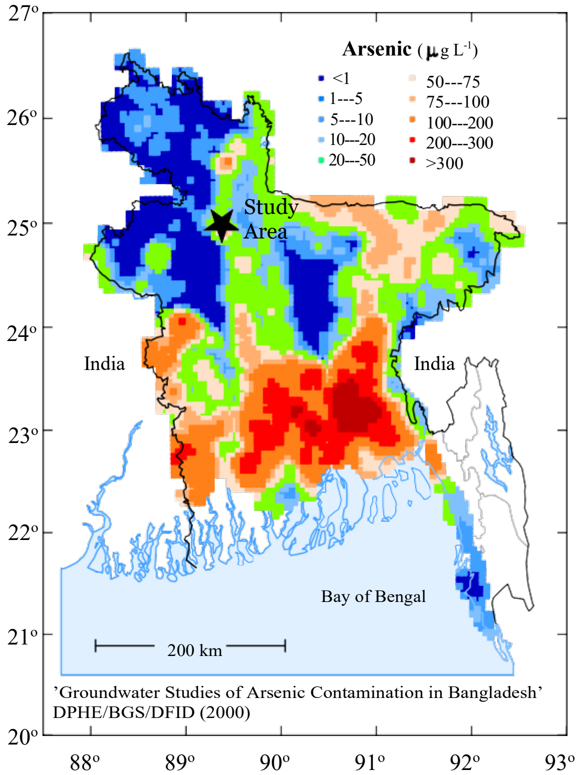
The context for this study is rural Bangladesh, where access to safe drinking water remains elusive (Human Rights Watch, 2016), despite having large volumes of renewable freshwater, even relative to its high population density (FAO, 2016). The problem is primarily a lack of access to high quality drinking water sources in rural areas. The vast majority of Bangladesh' rural population, consisting of more than 100 million individuals, according to World Bank statistics, now relies on drinking water obtained from approximately 10 million shallow, hand-pumped tubewells (Human Rights Watch, 2016).

Use of drinking water from these shallow tubewells was originally extensively promoted via education campaigns, as a safe alternative to the use of surface water for drinking. The switch from surface water to tubewell water was indeed followed by a sharp decline in child mortality (Caldwell et al., 2003), from under 5 mortality of above 200 per 1000 live births in the 1960s and 1970s to around 30 today, according to World Bank statistics.

However, shallow groundwater in Bangladesh is contaminated with naturally-occurring arsenic, a fact which was unknown when shallow tubewells were promoted because arsenic is undetectable without water quality tests. Long-term exposure to arsenic leads eventually to a number of serious health conditions, including internal and skin cancers. Daily use of arsenic-contaminated water at the Bangladeshi safe water standard of 50 parts per billion (ppb) — which is itself five times higher than the WHO standard of 10ppb — is associated with an additional 1 in 100 lifetime risk of cancer, rising to more than 1 in 10 for water that is highly contaminated (Smith et al., 2000). By the time the arsenic contamination problem was discovered, an epidemic of diseases associated with arsenic exposure was already established, called "the largest poisoning of a population in history" (Smith et al., 2000).

The primary solution proposed and implemented is the installation of deep tubewells. However, progress has remained elusive. Despite years of effort by the Bangladeshi government, non-governmental organizations and international aid agencies, around 39 million people drink water that is contaminated with arsenic at international standards (BBS and UNICEF, 2015). Source locations may be chosen for political purposes, rather than targeted to those areas in most need, and few sources are monitored after installation, meaning that some become unknowingly recontaminated (Human Rights Watch, 2016).

Figure 3.1: Arsenic contamination in Bangladesh and study site location




Additionally, recent studies have raised the concern than some efforts to reduce exposure to arsenic have increased exposure to fecal contamination, either because arsenic and fecal contamination are negatively correlated for hydrogeological reasons (Wu et al., 2011) or because households that adopt safer but more distant sources increase their exposure to fecal or bacterial contamination of drinking water through increased transport or storage times. Today, almost 100 million people still drink fecally-contaminated water (BBS and UNICEF, 2015).

Our intervention is located in north-western Bangladesh, in Shibganj and Sonatala Upazilas in Bogra District and in Gobindaganj Upazila in Gaibandha District, as shown in Figure 3.1. The study area is not in the epicentre of the arsenic contamination problem, but government officials and national media reported high levels of arsenic contamination in the specific study region (Daily Observer, 2014), and the distance from the epicentre of the epidemic meant that the area had received relatively low levels of prior intervention. Our implementing partner, NGO Forum for Public Health, viewed this as a major advantage, because they expected the marginal impact of providing deep tubewells to be larger in areas where few deep tubewells had been installed previously.

Within the study area, we targeted communities with high levels of arsenic contamination, using the limited data on arsenic contamination available before our study to pre-select candidate communities and then screening these candidate communities using water source testing. The final criteria for selection into the project was that either more than 25% of community water sources were contaminated with arsenic, or more than 15% of community water sources were contaminated and these sources were spatially clustered. We provide further details on recruitment to the study in Chapter 6.

The study population consists of primarily agricultural communities. Among the study sample, 40% of households are employed in agriculture, 12% are day labourers and 12% are small business owners. Communities are mostly poor or low income, but not extremely poor: 3.6% of households self-report as very poor, 22% report as poor, 38% report as low income, 34% report as middle income and 2.3% report as upper income. Table 3.1 [1] shows baseline socio-economic characteristics of the household sample for our study, including household size, religion, education levels, assets including land and livestock and other proxies for wealth including measures of housing quality.

Table 3.1 and Table 3.2, which presents baseline characteristics of the household sample with respect to their access to safe drinking water, also show comparable statistics from the national rural population, obtained from the Multiple Indicator Cluster Survey for 2012- 2013 (MICS) generated by the Bangladesh Bureau of Statistics - Ministry of Planning in collaboration with the United Nations Children's Fund (BBS and UNICEF, 2015). These comparisons allow us to evaluate how representative the study communities are of the rural population in Bangladesh and thus to what extent the results are likely to generalize. Table 3.1 shows that study population is largely representative of the national rural population, although households are somewhat smaller, the households are more likely to be Muslim, and they may be slightly poorer than the rural average: they are less likely to own a mobile phone, although they are slightly more likely to own a motorized vehicle.

Table 3.2 describes access to safe drinking water at baseline. We include measures of household water quality (variables labelled as "HH test") and water quality of the primary source of drinking water used by the household (variables labelled as "WS test"). As we do throughout the paper, we report arsenic contamination using both the Bangladeshi national threshold of 50 parts per billion (ppb) and the more conservative WHO threshold of 10 ppb. There is increasing evidence that the risks of exposure to between 10 and 50 parts per billion are still considerable (Human Rights Watch, 2016). Before our intervention, the local population was primarily dependent on shallow, privately-owned tubewells, the vast majority of which were owned by the household or another close relative. Correspondingly, the mean total time required to collect drinking water is around 2 minutes, lower on average in our sample than in the rural population as a whole, which also includes parts of Bangladesh where there is greater water scarcity in quantitative terms.[2]

[1]Appendices E.1 and E.2 provide a detailed description of how each variable reported in Tables 3.1 and 3.2 is constructed.

[2]However, there were also differences in how we measured this variable, and how the variable is measured in the MICS study which may also account for the differences. We asked the number of minutes

The rate of fecal contamination in water sources is higher than the national average, although the rate of fecal contamination in household drinking water is very similar. The amount of water in litres collected per day is larger in the national population sample than in our study sample, although this may partially reflect the difference in average household sizes. The rates of arsenic contamination are higher, which is unsurprising since we specifically recruit communities who face arsenic contamination problems.

it takes to walk to the primary water source from the respondent's house. In MICS, on the contrary, the question the interviewer asked was "How long does it take to go to the water source, get water, and come back?". To improve comparability, we calculate $Total\,time = Walking\,time * 2 + Queueing\,time + 0.5$, which is the value we report in the table. The distribution in MICS remains right skewed compared to our data, partially accounting for the large difference in means.

Table 3.1 : Socio-economic characteristics - Descriptive statistics

| | Study sample | National Population (rural) |
|---|---|---|
| Household size | 3.9<br>(0.022) | 4.6<br>(0.015) |
| The household head is muslim | .94<br>(0.012) | .87<br>(0.006) |
| The household head has no education | .42<br>(0.009) | .46<br>(0.004) |
| The household owns livestock | .76<br>(0.009) | .74<br>(0.004) |
| The household owns land for cultivation | .53<br>(0.011) | .48<br>(0.004) |
| Land owned by the household (acres) | 1<br>(0.049) | 1.2<br>(0.053) |
| HH has some toilet facility | .84<br>(0.008) | .94<br>(0.003) |
| Number of rooms to sleep | 1.9<br>(0.016) | 2<br>(0.009) |
| The floor is made of earth or sand | .84<br>(0.008) | .85<br>(0.004) |
| The roof is made of metal | .96<br>(0.005) | .92<br>(0.003) |
| Mobile phone ownership | .6<br>(0.017) | .83<br>(0.003) |
| Ownership of a motorized vehicle | .065<br>(0.004) | .051<br>(0.001) |

Notes: The table reports means and standard errors (in parentheses), obtained from a regression with no constant of each control on indicators for the study sample and the nationally representative sample. Standard errors are clustered at Primary Sampling Units level ("Treatment unit" for the study sample and "Cluster" for the Nationally representative sample).

Table 3.2: Water-related characteristics - Descriptive statistics

| | Study sample | National population (rural) |
|---|---|---|
| Arsenic contamination (WHO) (HH test) | .63<br>(0.017) | .61<br>(0.009) |
| Arsenic contamination (BD) (HH test) | .24<br>(0.016) | .17<br>(0.006) |
| Bacteria contamination (HH test) | .65<br>(0.016) | .63<br>(0.010) |
| Arsenic contamination (primary WS) | 37<br>(2.248) | 34<br>(1.586) |
| Arsenic contamination (WHO) (primary WS) | .69<br>(0.018) | .59<br>(0.011) |
| Arsenic contamination (BD) (primary WS) | .31<br>(0.017) | .19<br>(0.008) |
| Bacteria contamination (primary WS) | .54<br>(0.010) | .39.<br>(0.011) |
| Storage dummy (observed) | .73<br>(0.011) | .19.<br>(0.002) |
| The water is treated before drinking (primary WS) | .087<br>(0.008) | .035<br>(0.002) |
| Time needed to collect water (mins) | 2.2<br>(0.038) | 15<br>(0.268) |
| Water collected per day (litres) | 59<br>(1.199) | 79<br>(1.160) |

Notes: The table reports means and standard errors (in parentheses), obtained from a regression with no constant of each control on indicators for the study sample and the nationally representative sample. Standard errors are clustered at Primary Sampling Units level ("Treatment unit" for the study sample and "Cluster" for the Nationally representative sample).

# Chapter 4

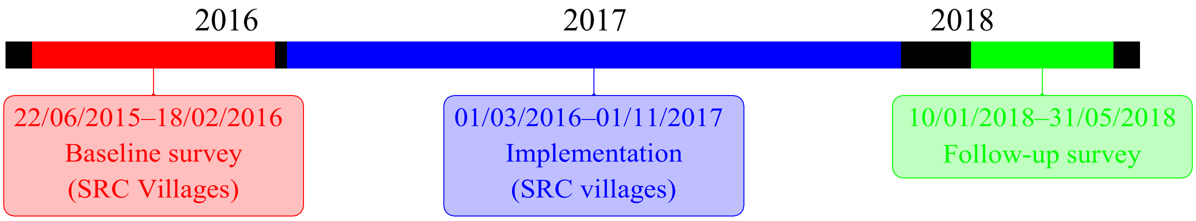
# Timeline

Figure 4.1 provides an evaluation timeline. We carried out baseline data collection in late 2015 and early 2016. Implementation took place throughout 2016 and 2017. An additional grant from the IGC allowed us to extend our sample size and recruit an additional 16 treatment units. We carried out baseline data collection for these treatment units in spring 2017, before implementation at the end of 2017. We carried out follow-up data collection in 2018.

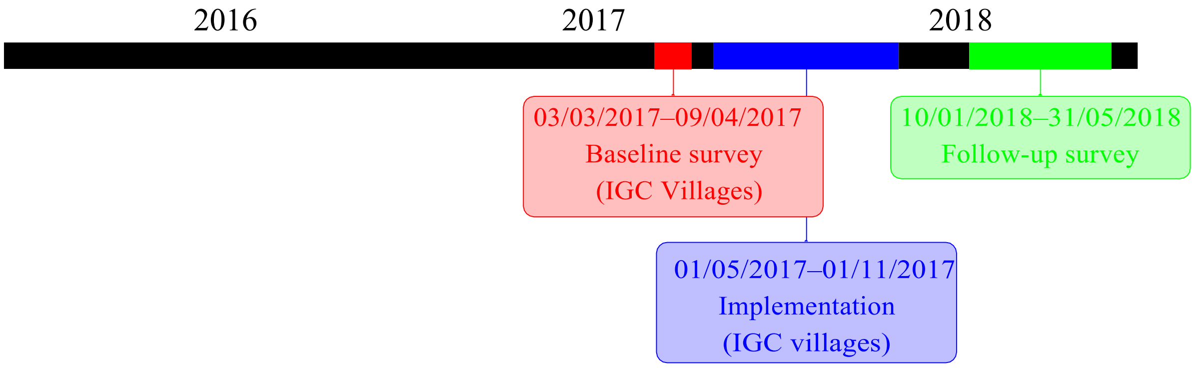
Notable events which took place during the study included an unprecedented rise in security concerns in Bangladesh, particularly associated with the murders of an Italian aid worker and a Japanese farmer in September and October 2015 and the attack on the Holey Artisan Bakery in July 2016. These security concerns did not materially affect the timeline of the project, although they affected the ability of the research team to move freely and discreetly in rural Bangladesh, since the team required a police escort. Local elections also created temporary insecurity, leading us to change our planned implementation schedule to avoid working in districts approaching elections. However, these changes did not alter the overall timeframe, only the timeframe in which implementation took place in specific unions. Finally, there was extreme flooding in the rain season of 2017, immediately preceding our follow-up survey. It is possible that the extreme flooding resulted in changes in water composition in tubewells, potentially affecting some of the patterns of contamination we observed at follow-up.

Figure 4.1: Evaluation timeline

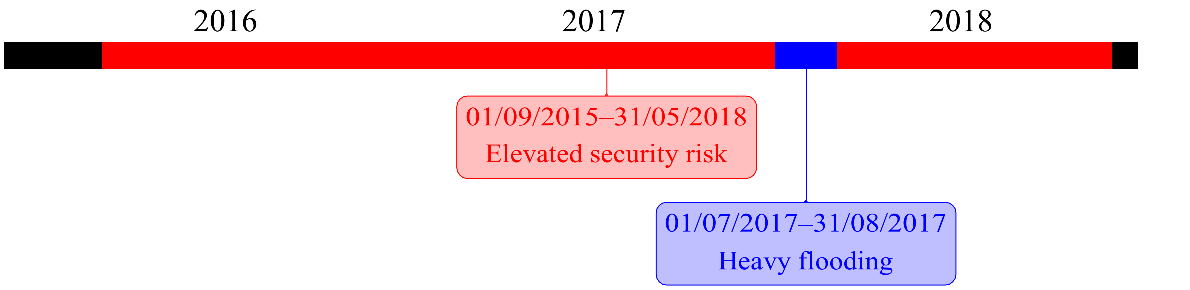
a) Project activities: SRC-funded villages




b) Project activities: IGC-funded villages




c) Other events



# Chapter 5

# Evaluation: Design, methods and implementation

In this chapter, we outline the study design, including the collection of data, assignment to treatment, our identification strategy and the measures we took to ensure data quality.

## 5.1 Ethical concerns

Before implementation, we developed a study protocol complying with all international human subject research standards. NGO Forum for Public Health obtained permission from the NGO Affairs Bureau in Bangladesh. There is currently no Swedish body to formally evaluate social sciences research overseas, and there is no independent Bangladeshi body to evaluate social science research. We therefore obtained independent review of our study protocol and follow-up data collection procedure from Ethical and Independent Review Services, an independent institutional review board based in the United States.

We obtained informed consent before enrolling any subject into the study. We obtained oral consent since we expected around 2/3 of study participants to have very limited literacy. All recruitment and consent procedures and study materials were translated into Bengali; informed oral consent is obtained in Bengali; and all survey data is collected in Bengali.

The risks associated with the study were minimal. The questions asked in the interviews were not sensitive. Participants could refuse to answer any question and interviewers were trained to conduct the interviews according to these rules. There was a risk of invasion of privacy, since we went to potential subjects' homes to ask for permission to interview them. The interview could then take place in their home. We strove to minimize the risk by asking permission and by asking where would be the preferred place for an interview if one occurred.

We preserved the confidentiality of the information provided to us. Households were assigned identification numbers, which are used to store and to organize the data, rendering the data anonymous. We store information linking identification numbers to names, addresses and GPS data securely either on a password protected server or in a locked office. We did not distribute this data to anyone other than co-investigators. The data is necessary to locate households who agree to participate in follow-up surveys and will be stored for the duration of this study and follow-up studies.

The overall benefits of this study are the potential improvements to projects designed to extend access to safe drinking water and potential reductions in the unintended consequences of such projects. Therefore, the potential benefits of the study are quite significant. These benefits are available to all people who lack access to safe drinking water, not just those who agreed to participate in the study.

Households or tubewell caretakers who participated in the water testing program could also acquire information about water source and drinking water quality, which they could use to reduce exposure to unsafe water. Households who participated in the study also had the opportunity to benefit directly from the safe drinking water intervention. The benefits from the safe drinking water intervention are available to all community members, not only survey participants, and will remain available as long as the community maintains and repairs the installed water source(s).

The alternative to participation was simply not to participate in the study. Subjects who chose to participate could withdraw at any time without any penalty. Also, those who did not withdraw could choose not to answer any particular question. Since the risks were minimal, the benefits should easily outweigh the risks for those who participated in the survey.

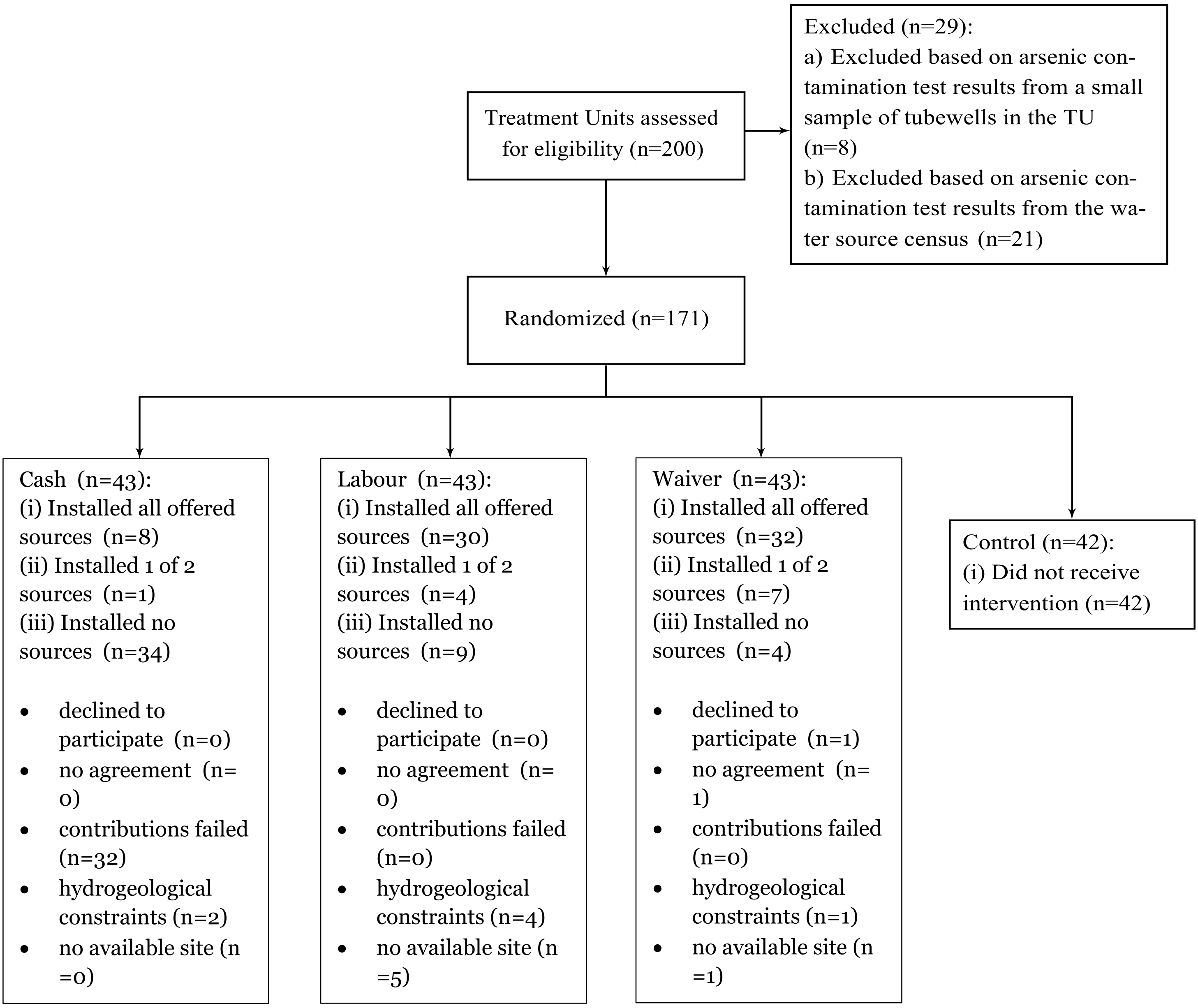
Additionally, the program requires participation in a community meeting and agreement over where to locate any water sources installed. The community meetings are open to the entire community, not only to survey participants, and no distinction is made in the decision-making process between those who participated in the survey and those who did not, either because they chose not to or because they were not randomly sampled for inclusion. The community decision-making process might exacerbate any pre-existing community tensions. However, the risks of participating in the intervention are no greater than those associated with participating in any community or NGO-led program to improve access to safe drinking water, or more broadly, improve local public services.

## 5.2 Evaluation design

The evaluation design is a randomized controlled trial, augmented by an analysis of mechanisms which allow us to elucidate how the reduced form effect of the program arises. The randomized controlled trial allows us to make simple comparisons of mean outcomes or changes in outcomes between treated and control communities. Since the treated and control communities are statistically indistinguishable at baseline, any differences between the two that arise after the intervention can be attributed to the causal effects of the program.

Figure 5.1 illustrates the evaluation design using a flow chart. We first evaluated treatment units for eligibility, and excluded treatment units with low arsenic contamination. We then randomly assigned the eligible treatment units to one of three treatment arms or to a control group. Communities assigned to treatment would all be offered the safe drinking water program, under three different contribution rules: a third of treated villages were required to raise a cash contribution before installation; a third of treated villages were required to contribute labour; and a third of treated villages received the program under a contribution waiver. The control group did not receive any intervention, although we did not prevent them from receiving any other interventions, or from installing their own safe water sources if they wished to do so.

Figure 5.1: Evaluation design flow chart



Our primary interest in this study is the average effect across the three contribution rules. However, take-up varies under the three contribution rules, and in particular, is much lower under the cash contribution arm. We discuss the impact of heterogeneous take-up on the results in Section 7.

To analyze mechanisms (Key Research Question 3), we originally proposed two approaches. Our first analysis of mechanisms is a difference-in-difference approach where we evaluate how changes in household bacterial contamination vary with changes in source contamination, transport distance and storage. The difference-in-difference approach yields causal estimates under the assumption that changes in the right hand side variables are uncorrelated with other changes in household drinking water contamination
e.g. through changes in household hygiene practices. Such an assumption might be reasonable, given that (as we will show) the difference in difference estimates are very stable across a range of specifications. However, although assignment to the safe drinking water program is random, selection of locations for water source installation is determined, by consensus, at a community meeting. As a result, it remains possible that changes in distance to collect drinking water, or source water contamination, may be correlated with other changes that also affect household drinking water contamination, through other channels. These confounding factors might, in principle, bias the above analysis.

To address this concern, we originally proposed a second, instrumental variables analysis exploiting the experimental design of the safe drinking water program. The instrumental variables approach uses baseline data to predict where in a village a community will decide to install a water source. Then, using these predicted locations and baseline household characteristics, we in turn predict changes in behaviour, in particular changes in source fecal contamination and changes in distance to a source. The advantage of this approach, in principle, is to eliminate any potential bias in the difference-in-difference analysis. However, the cost, as we noted in our pre-analyis plan, is substantially decreased precision. Our empirical results indicate that the IV estimates take the same sign as the difference-in-difference estimates, but the confidence intervals are extremely wide. Because it turns out that the instrumental variables analyses provide little additional information beyond the difference-in-difference analyses, we discuss the details of the IV method and the results only in Appendix D.2.

## 5.3 Sample size

Our study is implemented in geographically defined treatment units of between 50 and 250 households. The approach was motivated by a previous study where we found limited evidence for detectable treatment effects of well construction in larger villages. To define treatment units, we obtained the most up-to-date available lists of resident households from administrative sources. We used these lists to obtain village sizes, exclude from the study villages with less than 50 households and divide larger villages into several smaller treatment units along natural boundaries. In each treatment unit, we aimed to survey 40 households.

Several features of the study were pre-determined by the original study design, which was designed to compare treatment effects under the three contribution arms. The total number of treatment units we were able to recruit to the study was limited by budget constraints.[1] The balance between treated and control units was intended to maximize power to detect differences in effects between treatment arms. The number of households sampled was also predetermined.

When we planned this study, we carried out power calculations using simulations. The details of this process are in Appendix B. We calculated minimum detectable effects at the 5% level as 2.8 × the estimated standard deviation of coefficients. Table 5.1 summarizes the results of our power calculations. Note that an average change of 2.2m in walking distance corresponds to 7% of median distance to a water source at baseline. These minimum detectable effects compared favorably to expected treatment effects. We compare our results to these power calculations in Section 7.

Table 5.1: Summary of calculations

| Evaluation question | Outcome variable | Minimum Detectable Effect (5% level) |
|---|---|---|
| 1a | Arsenic contamination in household drinking water | 3.5% |
| 1b | Fecal contamination in household drinking water | 3.8% |
| 2a | Arsenic contamination in source water | 2.4% |
| 2b | Fecal contamination in source water | 2.8% |
| 2c | Distance to water source | 2.2m |
| 2d | Reported storage | 3% |

In addition, we simulated the difference-in-difference analysis and the instrumental variables (IV) analysis. The estimated effect sizes for the difference-in-difference analyses also compared favourably to plausible parameter values. We also simulated Sanderson Windmeijer first stage F-statistics of more than 10 for both instruments used in the IV analysis in about 85% of simulations. However, we anticipated that the IV approach would sacrifice considerable power: we expected the IV approach to have minimum detectable effects approximately ten times larger than the difference in difference approach. The first stage F-statistics are weaker than anticipated, for reasons we discuss in Section 7.1.2. We therefore report only the difference-in-difference results for the analysis of mechanisms and discuss the IV results in detail only in Appendix D.2.

[1]Our original budget covered 155 treatment units, but an additional grant from the IGC allowed us to extend our sample size by an additional 16 treatment units.

## 5.4 Sampling design

Within the study upazilas, we targeted communities with high levels of arsenic contamination. We describe the process of recruiting treatment units to the study in detail in Section 6.3.

We used the available household administrative lists in order to randomly sample households in each treatment unit for the household survey. We accommodated cases when selected households were not available for the interview or refused to participate by providing enumerators with a list of "replacement households", sorted in random order. Enumerators documented this replacement process in the household list used by the enumerators and recorded outcomes in the survey form, as they were required to fill in a form for all household that they tried to locate and conduct the interview with.

In 92% of cases the enumerators were able to conduct the interview with the household originally sampled for participating in the household survey at baseline. When this was not possible, the reason was that the household was not found in 33% of the cases, that noone was at home during the visit from our enumerator in 65% of the cases, or that the respondent refused to participate in the survey in 2% of the cases. Enumerators conducted the interview with the household head, their spouse, or another adult representative of the household. They always asked for their informed consent, both for the interview and, separately, for the water testing. 99.8% of households agreed to the interview, and 99.6% to the water testing. At baseline, we successfully conducted the household survey in a total of 6529 households across 171 eligible treatment units.

Occasionally, the number of households surveyed in a treatment unit was higher or lower than the targeted number. This is because in some cases we had to revise the treatment unit definition after completing the household surveys and reviewing the locations of households: in some cases, the administrative units had misassigned households to clusters. We reassigned the households so that each treatment unit retained geographical consistency.

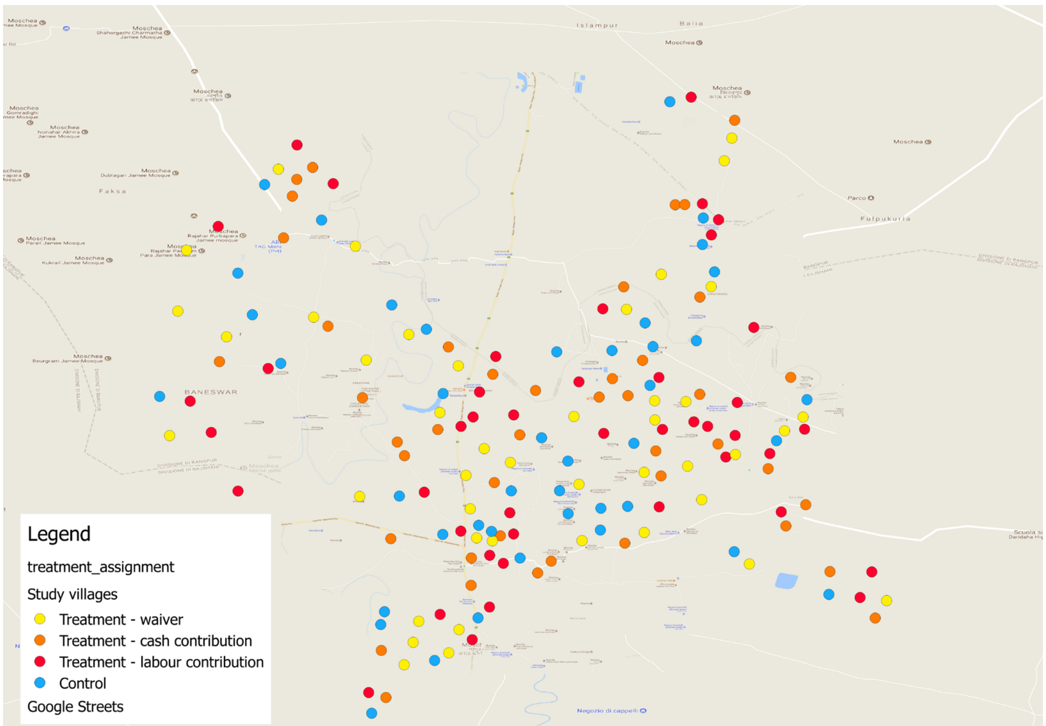
At follow-up, we were able to improve these statistics. The enumerators completed the interview with 99.85% of the households randomly selected to participate in the follow-up household survey. We were unable to complete the interview with 4 households that migrated, 2 households with no surviving household member and 5 households that refused to participate in the follow-up survey. Among households that we were able to successfully contact at follow-up, 99.9% agreed to the interview, and 99.1% to the water testing. The attrition rate between the baseline and the follow-up survey is 0.7%.

## 5.5 Assignment to treatment

Among the 171 treatment units enrolled in the project, we randomly select 129 to receive the intervention. Of these 129 treatment units, 43 treatment units were randomly assigned to each of the three contribution requirements: (i) under the cash contribution approach communities are required to co-fund the installation costs; (ii) under the labour contribution approach communities are required to provide labour to help with the installation work; (iii) under the waiver approach the new water source is installed for free. We assigned 42 treatment units to a control group which received no intervention.

We conducted the randomization at public lottery meetings, to which we invited representatives from each eligible community. The randomization was stratified by Union Parishad to make it feasible for representatives of the study communities to attend. The decision to use public randomization was motivated by concerns about transparency, especially given that we offer the same program under different conditions in different communities. We anticipated that information about the different conditions would spread, and this was indeed the case. The public lottery meetings gave our research staff an important source of legitimacy for project decisions taken. Figure 5.2 shows the resulting map of treatment units assigned to the control group and the three treatment arms.

Figure 5.2: Map of treatment and control areas




Note: Mid-point of study communities, as recorded during our baseline data collection.

Source: Google Street Maps

We also implemented the project sequentially by Union Parishad. As a result, households differ at follow-up in the amount of time they have been exposed to the treatment. The treatment effect we will estimate is therefore a weighted mean of treatment effects over the first two years of exposure to the program. Additionally, the IGC villages were added to the study after funding became available, and the time between baseline and follow-up differs for these households. These differences are absorbed by controls for stratification at the Union Parishad level in the final analysis.[2]

[2]The IGC-funded villages are in different Union Parishads to the SRC-funded villages.

Large treatment units were offered two tubewells; smaller treatment units were offered one tubewell, using an algorithm to assign the number of tubewells as a function of the original village size or of the treatment unit size.[3]

## 5.6 Data collection

### 5.6.1 Survey design

We collect data through a combination of surveys and a water quality testing program. All data collection activities were carried out by a team of enumerators employed by NGO Forum for Public Health and managed by the research team. At baseline and follow-up, the enumerators participated in a three-week training course including field testing, led by Habib in coordination with the Stockholm-based research team. The same enumerators carried out data collection in both treated and control villages. No compensation was provided to survey participants, although all survey participants were given the opportunity to acquire information about household and water source safety, which the vast majority took up.

Our data match households to the water sources they use. Our procedures for matching households to the water sources they use are novel, because the problem of matching households to decentralized infrastructure is not easy to solve. However, we extensively piloted the procedures in the field, and additionally built a number of checks into the process. We use different approaches to match households to water sources at baseline and at followup.

At baseline, we first conducted a full census of existing sources of drinking water. In order to identify all sources of drinking water, enumerators visited all households residing in the treatment unit and asked for an exhaustive list of nearby water sources. We used the existing administrative household list to structure the water source census, and collected information on households missing from that list during the census process. We also included public water sources in the census.

We then conducted the baseline household survey in the randomly selected sample of households. The household survey consisted of a detailed interview on household's composition, health, wealth, network and habits related to water collection and use. Each household identified the water source(s) used to obtain water for drinking or cooking purposes, selecting water sources from the list established during the baseline water source census. We showed the respondent a picture of each water source that he/she identified, to ensure that we correctly match households to water sources. In case the respondent reported using a water source not included in the water source census data, we collected the relevant information from this new source. This happened in only 2% of the household surveys, indicating good coverage of the exi-

[3]We designed the rules to allocate tubewells to achieve the goals of a parallel study regarding the effect of group size on collective action. Specifically, we implemented one of two rules: i) we assigned tubewells to villages as a function of village size, then divided these among the designated treatment units within each village; ii) we assigned tubewells to treatment units to keep the ratio of households to tubewells as close as possible to 125:1.

sting water sources from the census.

At followup, we do not repeat the water source census from baseline, because of the cost of this exercise. Instead, we first conduct the household survey, and then collect data from all the water sources that households describe using. To avoid resurveying water sources multiple times, we tag each water source with a zip tie. If an enumerator visits a source that has already been surveyed, they record a photograph and take GPS coordinates, enabling us to confirm the match to the water source data already collected by another enumerator.

### 5.6.2 Water quality tests

The water quality testing program consisted of three types of tests: (i) fecal contamination test; (ii) field arsenic test; (iii) laboratory arsenic test. For fecal contamination and laboratory arsenic tests we used QR barcodes to identify each water sample and to link the survey data with test results. The tests we use are standard in the literature and have been used in previous studies of water quality.

**Fecal contamination field test** We conducted the bacteria test for all water sources and for all households surveyed, provided that the survey respondent agreed to the testing procedure.[4] The water testing procedure for bacteria contamination used hydrogen sulfide vials produced by NGO Forum for Public Health. These tests detect bacteria that produce hydrogen, which are almost exclusively organisms that live in the gut of warm-blooded animals, and therefore indicate the presence of human or animal fecal contamination. The vials should be kept at room temperature for 48 hours, and the test is read as positive if the colour changed from clear to black. The hydrogen sulfide test has been rigorously evaluated in Bangladesh by NGO Forum for Public Health. We informed respondents about the bacteria test results when the results were ready, on average two days after the water sample collection, by SMS.[5]

Project staff entered the bacteria test results on average after 2 days from the water sample collection. However, in some cases, particularly at baseline, tests were left for more than 2 days, or in some cases, entered after 1 day. To ensure that data is comparable across rounds, we apply a correction to the data which accounts for variation in how long each test was left before entering the data. The correction we apply uses information on the specificity and sensitivity of the fecal contamination field test from a similar set of samples, also from Bangladesh, reported in Gupta et al. (2008). [6] We use the mean rate of positive tests, the sensitivity and the specificity to back

[4]Of the households who consented to participate in the survey, only 3 households did not consent to the testing procedure. For these households, the test results are set to missing.

[5]During the water source and household survey we asked respondents to provide us with a phone number to be used for sending by SMS the results from the bacteria test. At baseline, 99% of the respondents in the water source survey and 94% in the household survey provided us with a phone number for further communications. At follow-up, we were able to obtain the phone number for 99% of households participating in the household survey.

out the probability that positive and negative test results truly reflect fecal contamination. Intuitively, the correction implies that a sample that turns black in a very short time period has a near 100% chance of contamination, while a sample that remains clear after a longer time period has an increasingly small probability of contamination.

**Arsenic field test** We conducted the field arsenic test for all water sources and for all households surveyed, provided that the survey respondent agreed to the testing procedure. This testing procedure is implemented in the field, and it uses the EZ Arsenic High Range Test Kit (Hach), which provides results in 20 minutes and measures arsenic levels within the range of 0-500 ppb (parts per billion) with the following increments: 0, 10, 25, 50, 250, 500. Test results are immediately available, so we informed respondents about the results at the end of the survey. Enumerators also gave a report card (in Bangla) to the owner/caretaker of the water source and to the households participating in the household survey, reporting the date of the test, the result of the arsenic field test and some guidelines on safety actions to take in case of bacteria or arsenic contaminated water.

This procedure for measuring arsenic levels in the field provides reliable results for water freshly obtained from the source, but the ability of the test to detect the presence of arsenic in the water decreases the longer the water is stored. Arsenic begins to oxidize once the water is stored in a container that is open to the air, and the field test does not detect oxidized arsenic. We collect and test samples from water sources directly from the source and are therefore confident about the accuracy of the field test. However, during the household survey we asked respondents for a glass of water obtained in the same way household members would normally obtain a glass of water for drinking i.e. either from storage or direct from the source, using the same containers for transport that they normally use. This gives us a measure of the quality of water normally used by households. However, for stored water, we were concerned that this might underestimate arsenic levels, if the tested water had been stored for a long time.

**Arsenic laboratory test** Because of our concerns about the accuracy of the field test in samples of water that had been stored for some time, we complemented the field testing procedure for a subset of households, using a laboratory test conducted at the Water Quality Testing Laboratory (WQTL) of NGO Forum for Public Health using Atomic Absorption Spectrophotometer (AAS). At baseline, we randomly selected for the arsenic laboratory test 10 households, out of the 40 sampled for the household survey, in 92 treatment units. We stopped laboratory testing after 92 treatment units because of budget constraints, as the lab tests are much more expensive (approximately 100 times) than the field tests. In total, we tested 897 water samples in the labo-

[6]The specificity and sensitivity of the test in reality will vary depending on the extent and not just the presence of contamination in the samples used. Ideally, we would use values of specificity and sensitivity that were specific to the tubewell and household samples separately. However, these values are not available.

ratory at baseline. The field tests are designed to be somewhat more conservative than the laboratory tests, because a false negative has much more serious consequences for health than a false positive. However, when the results for the two sets of tests are compared at baseline, they are highly correlated.[7]

## 5.7 Potential sources of bias

We discuss potential sources of bias in this section, and return to the question of whether bias from these sources affects our results in Chapter 8.

**Spillover effects** Our program targets communities that are highly arsenic contaminated in 10 Union Parishads. Arsenic contamination is also geographically clustered. As a result, villages enrolled in our project lie in relatively small geographical areas (Figure 5.2). Moreover, because we divide large villages in several treatment units (Section 5.4), it is not uncommon that control and treated communities are adjacent or very nearby. Indeed, 120 out of 171 communities enrolled in the study are obtained by splitting a large village in two or more treatment units. The average distance between control communities and the closest treated community is 650m (about 8 minutes walking time). As a result, the average distance between households in control communities and the closest project tubewell is 575m (7 minutes walking time). For comparison, this distance is 175m (2.2 minutes walking time) in treated communities.

Despite this geographical proximity between control and treated communities, we expected minimal spillovers from treatment to control communities in terms of take-up of wells. As reported during the baseline household survey, households use water sources very close to their house, on average 36m from their house or less than half a minute walking distance. Water is most often collected by women and children, so households are unlikely to use a water source not in the proximity of their house or outside their cluster. Take-up rates of new sources decline steeply with distance and are negligible at more than 5 minutes walking distance, which is approximately 400m.[8] Moreover, households often stated during community meetings that they were unwilling to use a water source from a different cluster, even if the cluster were located within the same community.

**Reporting bias** All the analyses rely on household reports of which water source they use, allowing us to match household data to water source data. Previous research (e.g. Ahuja et al., 2010) and our own experience suggest that social desirability bias influences household reports of behavior with respect to obtaining drinking water i.e. households underreport using unsafe

[7]Correlation is much weaker for the same two sets of tests at follow-up, which may reflect a problem with our tracking systems.

[8]At baseline, 94% of households reported that they would switch to a new source if it was within 1 minute of their compound, but only 3% report that they would switch if it was 7 minutes walk from their compound.

water sources. We constructed our questions to reduce the effect of social desirability bias i.e. we initially simply ask them to list the sources they use, and ask them about the water sources they use before discussing knowledge about water safety.

**Hawthorne effects** Our primary concern regarding Hawthorne effects is that people are likely to take more care to avoid water contamination if they know the water is going to be tested e.g. wash their hands and vessels more scrupulously than usual. The data collection process is identical in the treatment and control groups. However, one might suspect that these effects might be stronger in the treatment group, who have also participated in the safe drinking water program, and might therefore experience stronger "experimenter demand effects" (Levitt and List, 2007). In general, this biases us against finding effects on transport and storage contamination. The possible consequences are to bias our comparison of the aggregate effects of the intervention on fecal contamination in household drinking water; and (possibly) to bias our estimates of the effect of increasing transport distance and storage time. However, to the extent that both treatment and control groups experience the same level of observation and scrutiny, and the difference between the two stems only from the intervention to provide safe drinking water, then the effect adjusting for any hygiene response may in fact be the policy effect of interest.

**John Henry effects** In contrast, our study might encounter John Henry effects if households in control villages, who receive information about water contamination but no program to improve access to safe drinking water, exert more effort to reduce contamination through other channels, for example by improving household hygiene or by more proactively seeking access to other safe sources in their communities. This source of bias would have the opposite effect to the Hawthorne effects discussed above, and is perhaps less of a concern, in that this would tend to attenuate differences between treatment and control households.

## 5.8 Quality checks

All survey instruments underwent a rigorous process testing process, including at least two rounds of piloting. Enumerators and research assistants provided extensive feedback which was incorporated into the survey design. All survey forms were available in English and in Bengali, and enumerators were free to select the language version that they felt more comfortable with. The Bengali version was verified by back translation.

We collected survey data using tablets, using the technology platform provided by SurveyCTO. The electronic platform allowed us to introduce checks and constraints on enumerators' entries at the moment of data collection, to automatically trigger the correct modules, depending on respondent answers, and to prevent enumerators from accidentally skipping questions.

At follow-up, we incorporated project monitoring data[9] into checks and constraints, for example adding verification questions where responses diverged from project records[10] and automatically triggering different modules depending on treatment status and project stages.

The data collection process included monitoring tools, quality control measures and incentives for enumerators. First, enumerators were required to finalize and submit at the end of each working day the surveys collected that day. We provided field supervisors with the basic statistics on the number of surveys conducted by each enumerator, disaggregated by date and updated every day after their daily submissions. Second, we complemented this monitoring tool with weekly statistics with a more comprehensive assessment of each enumerators' work, which included quality indicators such as the percentage of non missing answers in the survey, the percentage of water source surveys conducted with the caretaker or the owner of the water source and the number of household members for which detailed demographic data were recorded in the household roster. We created an incentive structure for the enumerators by paying weekly a salary bonus to the best five performing enumerators. Third, we randomly selected five households in each treatment unit for a back-check survey conducted by field supervisors. We selected a set of back-check questions from the main surveys, and each back-check survey consisted of a random subset of these back-check questions. We provided field supervisors with the weekly summary of the back-check results and we can use this data to assess the accuracy of the information collected by the enumerators. Fourth, we exploited the electronic nature of the surveys in order to introduce unannounced audits in the survey forms, which recorded the number of seconds spent on each question, providing us with another indication of data collection accuracy. Finally, we took audio recordings of surveys (with consent from participants but for quality control purposes only) and Habib discussed the audio recordings with the enumerators, providing guidance where necessary.

We also designed our data collection procedures to provide multiples sources of evidence on outcomes. For example, we record attendance at project meetings directly and also ask survey participants to verify whether or not they participated in meetings.

[9] The monitoring strategy for implementation is described in Section 6.2.

[10] We did this to ensure that we did not not miss important data. For example, if households denied all knowledge of the project, we did not ask them any follow-up questions. To avoid missing valuable data, we prompted households in treated communities with a reminder of the project's characteristics, and allowed them to change their answer if they recalled the project after prompting. We always recorded their initial responses, before prompting, and we always allowed respondents to report answers that diverged from our reported records.

# Chapter 6

# Programme design, methods and implementation

## 6.1 Key programme elements

The programme we evaluate was developed jointly by NGO Forum for Public Health and the research team, drawing on past experience of implementing similar projects. Table 6.1 summarizes the key implementation activities that constitute the safe drinking water program.

Table 6.1: Implementation activities

- Preparatory visits, information gathering, community mobilization
- Community decision-making
- (If cash approach) Collection of cash contributions
- Installation of the pumpbody
- Water testing
- Construction of the platform
- Selection of caretakers
- Caretakers traning
- Monitoring visits

Before organizing community meetings, field staff visit the treatment unit, collect basic information on the geography of the village and the main socio-economic characteristics of the clusters grouped in the treatment unit, inform households about the scope of the intervention, organize information meetings within each cluster and agree on a date for the first community meeting. These preparatory activities are usually carried out over the space of a week and are crucial in order to guarantee that project staff are familiar with the specific circumstances within each treatment unit and mobilize the community.

Field staff then organize information meetings in all clusters (or groups of households) in each community, increasing awareness about water safety issues and stressing how important it is that everyone participate actively in the community meeting.[1]

Following these initial information meetings, the field staff then organize the main community meeting, at which communities take key decisions about whether to participate in the project and where to locate the sources offered by the project. All households are invited to the meeting and encouraged to participate.[2] As discussed in Section 2, decisions taken at the community meeting must be agreed upon by consensus in the presence of project staff, and both women and the poor must be represented at the meeting at which decisions are taken. The meetings are only carried out upon fulfilment of minimum participation requirements; field staff sometimes have to reschedule for another date if the minimum participation requirements are not fulfilled.

Community meetings are usually around one hour long. The meetings begin with a short introductory briefing by project staff on water safety issues and project implementation rules. The information provided on water safety issues primarily focuses on source safety, explaining how arsenic and fecal contamination at source arises, which sources are at risk, which sources can provide safe water, and the health consequences of exposure to arsenic and fecal contamination. The main activity at the community meetings is a longer discussion session during which the communities take decisions, by consensus, on key aspects of the project. If decisions are not reached, we offer to organize another meeting, up to a maximum of three meetings per treatment unit.[3] During the meetings, field staff displayed large-scale maps of the community showing all the community water sources and their contamination status, developed using the baseline water source census data.

Communities assigned to the cash approach have a maximum of 12 weeks to raise the required amount,[4] during which time project staff visit the community several times in order to remind the communities of the deadline and establish progress. If assigned to the labour contribution treatment, communities must sign a contract committing to provide the labour contribution and coordinate with the project staff and the contractors to agree on a time to provide the labour contribution. In practice, the timing is mostly determined by contractor availability. Communities know in advance approximately when the labour contribution will be required, but there is some uncertainty until a few days before installation, resulting from variation in how long the wells scheduled immediately beforehand take to drill, which is uncertain ex ante.

[1]Although all households were invited to the information meeting, participation was voluntary. The field staff extensively tried to involve women in these activities, stressing the importance of their awareness and their participation for the safety of the water consumed in the household.

[2]On average, 50% of the households attend the community meeting, and 41% of participants are women. Poor and very poor households (by self-reported status) are less likely than middle income households to attend the meetings, but only slightly less likely to do so: 44% of very poor households attended the meetings, compared to 53% of middle income households. Households with high baseline arsenic contamination in household drinking water are also more likely to attend the meetings.

[3]In practice, few communities organized more than one meeting, and no communities organized more than two meetings.

[4]Field staff initially give them a six week deadline, which can be extended twice for an additional three weeks on each occasion.

Installation of the wells uses local technologies, primarily human manpower, to manually turn a drill bit approximately 60mm in diameter. The technology can penetrate layers of weak or fractured rock, but not solid rock. Project staff, including the field engineer, supervise the installation in order to guarantee that the tubewell depth is adequate to reach an arsenic-free aquifer. The goal is to reach a safe layer, meaning a layer which is permeable (through which water can flow relatively freely) but which is separated from the arsenic-contaminated layers at the surface by an impermeable layer (through which the contaminated water cannot pass). If such a layer is reached, the drill is lifted and withdrawn and a PVC pipe is inserted into the hole. Pumps maintain pressure in the excavation to reduce the likelihood of collapse of the excavation. If the underlying geology is very sandy, there is also a risk that the excavation collapses and the PVC pipe cannot be inserted. Communities assigned to the labour approach are required to provide unskilled labour during the first three days of the installation work, monitored by project staff.

After installation of the PVC pipe and the pump body, we conduct laboratory water tests for arsenic, iron and manganese. If these test results are satisfactory, we finalize installation by adding the pump handle and constructing a platform to protect the pump body and manage drainage around the pump.

After the construction of the pumpbody and the platform, project staff organize a community meeting with users of the well in order to appoint two responsible individuals as caretakers for each tubewell, one man and one woman. The appointed caretakers are trained by our field engineer in how to maintain the water source and keep the site clean.

Project staff conduct three monitoring visits after the completion of the intervention, to assess usage and maintenance of the provided tubewells in the first few months after installation, respectively within 6 weeks, 8 weeks and 12 weeks after the construction of the tubewell pumpbody and platform.

The installation procedure, from the first preparatory visits to the community to the completion of the construction of the pumpbody, takes on average two months. In most treatment units it is completed within four months. We conduct laboratory water tests for arsenic, iron and manganese contamination, and installation is finalized with the construction of the platform, on average two months after the installation of the pumpbody, and in the majority of cases within three months.

## 6.2 Monitoring

The implementation roll-out of the intervention has been closely monitored via a systematic and comprehensive data collection on most project activities. Most of the information used for implementation monitoring has been collected by electronic forms, making it available directly after submission to the management team in Bangladesh as well as the research team. We exploited the electronic nature of this data collection in order to make information collected at previous stages of the project automatically available for project staff and to prevent important data collection procedures from being accidentally skipped. These elements minimized the risk that the intervention was not carried out accordingly to the treatment assignments or that information was mis-recorded.

We complement the electronic data collection system with a range of additional project documentations: a record of all staff visits carried out to each treatment unit (Activity Report); an extensive qualitative narrative by project staff of all implementation stages per treatment unit (Project Staff Report); the record of attendance data and participation in decision-making per community meeting/caretaker selection meeting using predefined household lists (Attendance Sheet); and other office records, including of installation processes, key dates for the implementation of the intervention, and caretakers training. Additionally, we required project staff to record the audio of all information meetings and community meetings organized, and we have transcribed, translated and coded the audio records of the meetings. This comprehensive monitoring plan resulted in the list of indicators to monitor the implementation of the intervention, which is described in detail in Appendix C by implementation stage.

The implementation programme did not change during the study period, and there were only minor deviations from the study protocol.[5] There was limited scope for implementers to innovate, although the process of facilitating the community meetings, in particular, requires some learning: the only community which failed to reach an agreement was the very first meeting our team organized. We note that much of the intervention design was based on prior experience from a similar project, meaning that the implementation procedures were to a large extent "tried and tested".

The implementers were necessarily aware that they were participating in an experiment, since they were required to implement the program under different conditions in different communities. However, only the contribution requirements varied across communities: all other features of the implementation protocol remained the same.

## 6.3 Recruitment

We targeted communities who faced a problem of arsenic contamination and lacked safe sources of drinking water. A major challenge was identifying these communities in a region with relatively limited data on arsenic contamination. We used the limited data available to pre-select villages and then refined selection using water source testing. We pre-selected a list of candidate villages for the intervention on the basis of contamination levels reported in the available sources of arsenic testing data. We had access to village-level data from the following data sources: (i) data from the Bangladesh Arsenic Mitigation Water Supply Project (BAMWSP), which included a large tubewells screening program conducted between 1999 and 2006; (ii) the assessment from the Department of Public Health Engineering (DPHE) on the most arsenic contaminated villages in the Bogra region; (iii) data collected in 2008 from the Bangladesh Social Development Services (BSDS). We pre-selected as candidate villages for receiving our intervention all villages ind-

[5]For example, in one treatment unit, our treatment unit definition protocol was not correctly implemented, resulting in a treatment unit consisting of two clusters too geographically distinct from one another to be treated together in practice. As a result, the field staff only implemented the project in one of the two clusters, not the full treatment unit. These cases were rare.

icated by the DPHE or for which BAMWSP or BSDS data reported a share of arsenic contaminated tubewells equal or higher than 30%. We confirmed this initial selection by testing for arsenic contamination in a small sample of tubewells in the village.

For these candidate villages, we defined treatment units of between between 50 and 250 households, as described in Section 5.3. We identified a total of 192 candidate treatment units in 103 villages, of which 51 were divided in two or more treatment units. We conducted a full census of existing sources of drinking water in these candidate treatment units. We used the water source contamination data in order to finalize the selection of the treatment units eligible for receiving the arsenic mitigation program. In particular, we excluded from the study all treatment units with less than 15% of arsenic-contaminated water sources. We further screened treatment units with less than 25% of arsenic-contaminated water sources, including them in the program only if they presented a well defined cluster of contaminated water sources.[6] We excluded treatment units where arsenic contaminated water sources were geographically scattered, because in these cases all households in the village already had a nearby source of arsenic-safe water. We continued to recruit new unions and communities to the study and implement the same recruitment policy until we achieved our target recruitment levels. The final study population consists of 171 treatment units, all of which had arsenic contamination levels greater than 25% or substantial clusters of arsenic contamination.

We assigned treatment units to one of the treatment arms or the control group at public lottery meetings, as described in Section 5.5. At the public lottery meetings, representatives of the study communities expressed approval of the fairness of the approach of selecting treated villages. Treatment units assigned to the control group understandably expressed disappointment. Few communities expressed disappointment with assignment to a particular treatment arm (e.g. the cash contribution arm) at this stage of the process.

In the 129 communities assigned to treatment, we implemented the program. The first step in the program is organizing community meetings. To do this, field staff made a number of visits to the communities to disseminate information about the project, and to agree on meeting times and locations. Among the 129 communities assigned to treatment, only one community (assigned to the contribution waiver) declined to organize a meeting. At the meeting, communities were asked to take a decision about where the proposed well(s) should be located. Only one community failed to reach consensus during the community meeting.

---

[6]To evaluate these treatment units with between 15% and 25% contamination, we reviewed the maps obtained from the water source census.

## 6.4 Targeting

We successfully recruited communities to the program with significant arsenic contamination issues, and communities randomly assigned to receive the safe drinking water program do not significantly differ from those assigned to the control group (as discussed in Section 7). In the 129 communities in which we implemented the program, we offered to construct a total of 179 tubewells. As discussed in Section 2, we either offered 1 or 2 tubewells to each community, depending on the treatment unit size. Table 6.2 summarizes the result for each attempted installation.

Of the 179 tubewells we offered, we successfully installed 107. One community declined to hold a meeting; one community could not agree on a location. For 13 offered wells, no suitable land could be identified. For 44 offered wells under the cash contribution arm, the community did not raise cash contributions, despite holding a community meeting, agreeing on a site and committing to raise the cash contributions at the time of the meeting. Finally, at 13 sites, the communities successfully completed all stages of the project and we attempted installation, but we could not complete installation because of hydrogeological conditions: either the presence of an impenetrable rocky layer or a sandy layer which caused the excavation to collapse before the PVC pipe could be installed.

Table 6.3 shows how communities who successfully completed the program (resulting in attempted installation) differed from communities who did not successfully complete the program (resulting in no attempted installation), as well as how communities in which we successfully installed wells differ from those in which we did not successfully install wells. Where we did not successfully install wells, this implies that we either did not attempt installation, because the community did not successfully complete the program, or that we attempted installation and failed.

Communities who successfully completed the program are positively selected for arsenic contamination i.e. higher contamination, particularly at the more conservative Bangladeshi threshold, is selected with a greater likelihood of attempted and successful installation. Other characteristics, including the poverty score, are not strongly correlated with attempted and successful installation. These differences suggest that the program was successful in targeting communities with arsenic contamination.

Table 6.2: Project outcomes, by offered tubewells

| Installation outcome | Number of tubewells |
|---|---|
| Successful installations | 107 |
| Failed to raise cash contributions | 44 |
| Installation attempted but failed due to hydrogeological conditions | 13 |
| No suitable land was identified | 13 |
| Community did not agree on location | 1 |
| Community did not hold meeting | 1 |
| Total number of offered tubewells | 179 |

Table 6.3: Selection into successful installation

| | Attempted Installations | Successful Installations | Obs |
|---|---|---|---|
| Arsenic contamination (WHO) (HH test) | 0.06<br>(0.04) | 0.02<br>(0.04) | 4917 |
| Arsenic contamination (BD) (HH test) | .09***<br>(0.03) | .06*<br>(0.03) | 4917 |
| Bacteria contamination (HH test) | 0.00<br>(0.03) | 0.00<br>(0.03) | 4899 |
| Household size | -0.08<br>(0.05) | -.1*<br>(0.05) | 4918 |
| Poverty score - 2 USD | -0.00<br>(0.01) | -0.00<br>(0.01) | 4889 |
| Not educated HH members (%) | -0.00<br>(0.02) | -0.00<br>(0.02) | 4918 |
| Literacy rate in the household | 0.01<br>(0.02) | 0.00<br>(0.02) | 4911 |
| Network nominations | -0.06<br>(0.08) | 0.02<br>(0.07) | 4918 |
| Network size | -0.06<br>(0.08) | 0.03<br>(0.08) | 4918 |
| Muslim household | 0.03<br>(0.03) | 0.03<br>(0.02) | 4913 |
| High trust towards community | 0.03<br>(0.02) | 0.02<br>(0.03) | 4914 |
| Know association | 0.01<br>(0.02) | 0.02<br>(0.02) | 4863 |

Notes: Column 1 summarizes differences on listed characteristics between: i) households living in communities which completed all stages of the program and in which we attempted installation; and ii) households living in communities in which we did not attempt installation, because these communities either did not choose a site, could not identify a suitable piece of land, or did not raise cash contributions. Column 2 summarizes differences between: i) communities in which we successfully installed at least one water source and ii) communities in which we did not install any water sources. Results are obtain from a regression of the listed characteristic, measured at baseline, on the rate of attempted or successful installation. Installation rates can take the value 0, 0.5 or 1. Regression at the household level with weights ensuring that all communities count equally, with centered controls for union-level stratification. Standard errors clustered by treatment unit.

## 6.5 Evidence on implementation procedure

There was widespread awareness of the program. Among treated communities, 87% of households knew that NGO Forum had carried out a program to prove new safe sources of drinking water to communities in their district, while 57% were aware of the program in control communities. In treated communities, 80% of households knew that their community was selected to receive the water safety program implemented by NGO Forum.

Among households that knew about the program: 87% of households in treated communities and 83% of households in control communities knew that communities were selected to receive the program by lottery, and very similar percentages of households[7] knew that communities selected to receive the program were assigned to different conditions and implementation rules and that assignment to treatment was done by lottery[8].

Among households in treated communities that knew that their community received the water safety program implemented by NGO Forum, 96% of households remember that some meetings were held in their community in relation to the program[9] and 77% of households reported that at least one household member attended the community meeting. Our program records suggest that around 50% of households participated in the meeting. 95% of households correctly remembered the number of offered tubewells[10] and 98% of households correctly remember the contribution requirement (cash, labour or waiver).

Table 6.4 summarizes water quality statistics in the tubewells installed by the project, compared to other non-project wells used by other households in the same communities, in which we successfully installed at least one tubewell. The results confirm that the tubewells installed by the program successfully reduce arsenic contamination to minimal levels, although 6% of project-installed wells test positive for arsenic at the WHO threshold and 1% of project-installed wells do so at the higher Bangladeshi threshold. The arsenic field test we use to test tubewells is conservative, so these results likely overstate contamination in these wells.

In contrast, although the project tubewells are substantially less likely (13 percentage points) to test positive for fecal contamination than are non-project wells in the same communities, a considerable fraction of these water sources (34%) still test positive for fecal contamination. This rate of contamination is unexpected, because the source of water these tubewells draw upon is free from contamination. How these wells become contaminated, and via what channel, is an open question: contamination could potentially take place through leakage into the pipe system from shallow groundwater, or within the pump body itself. It is possible that contamination occurred during the floods during the rain season before our follow-up survey. Another recent study (ICDDRB and UNICEF, 2018) also finds substantial levels of fecal contamination in water obtained from tubewells. In that study, a comparison of samples taken before and after decontamination of the mouth of the tubewells pointed to contamination of the tubewell mouth as the mechanism.

[7] 87% (treated communities) and 82% (control communities).

[8] 85% (treated communities) and 81% (control communities)

[9] This rises to 97% after prompting.

[10] This rises to 99% after prompting.

The communities in which we implement the program are small and relatively compact. As a result, half of households in communities where we successfully install at least one water source are less than 1.6 minutes walk from a new source. The mean distance to a new source in these communities is 2.2 minutes walking time.

Table 6.4: Comparison of project tubewells with other tubewells

| | Fecal contamination (predicted) | Arsenic contamination (WHO) | Arsenic contamination (BD) |
|---|---|---|---|
| | (1) | (2) | (3) |
| Project TW | -0.13***<br>(0.05) | -0.58***<br>(0.04) | -0.35***<br>(0.03) |
| Mean (project TW) | 0.34 | 0.06 | 0.01 |
| Mean (other TWs) | 0.46 | 0.63 | 0.34 |
| N | 3394 | 3510 | 3510 |

Notes: The table reports the regression estimated difference in contamination in project tubewells compared to other tubewells in the same communities, in a regression which includes treatment unit fixed effects. Standard errors are clustered by treatment unit and shown in parentheses. The table also results mean contamination levels in project tubewells and non-project tubewells in the same communities. The sample is water sources which at least one sample household reported using for drinking or cooking.

## 6.6 Unexpected events

The primary unexpected response we encountered was the low rate of take-up in the treatment arm assigned to the cash contribution requirement. In another context in rural Bangladesh we implemented a similar program with similar cash contribution requirements and we successfully installed about 83% of the tubewells that we offered (Madajewicz et al., 2018). We were therefore surprised by the negative response to the cash contribution treatment in this context.

A number of differences between the contexts may account for the different responses. First, the context for the previous study had higher average rates of arsenic contamination, meaning that the willingness to pay for safe tubewells may be higher. Second, we implemented the project under slightly different rules that may have eliminated certain kinds of elite capture. In the previous study (Madajewicz et al., 2018), we frequently found that only one household paid the cash contribution for the well and qualitative evidence from that study suggests that payment of the cash contribution was associated with a perceived right to control use of the source. In this study, field staff also reported to us that they frequently received offers from households willing to pay the (full) cash contribution, but only if the source were constructed on their land. If these offers were made during the community meetings, project staff reported to us that communities rejected the offers on the grounds that the well was for the whole community. If these offers were made to staff after the meeting, project staff followed project guidelines and upheld the decisions taken at the meeting as binding. Third, the communities may have been disgruntled about the program being offered for labour contributions or at a contribution waiver in other communities, although communities assigned to the cash contribution arm did go through the process of selecting locations at a community meeting.

As discussed in the previous section, an additional unexpected response was the relatively limited improvements in source quality with respect to fecal contamination.

## 6.7 Weak links

The unexpected developments discussed in the previous section suggest two potential weak links in our posited theory of change. The first weak link concerns the first step: constructing new safe sources of drinking water. In this study, unlike in our previous work, some communities did not successfully raise the necessary contributions under the cash treatment arm, and as a result, the rate of successful well installation was lower than expected. These results confirm that the success rate of a well installation program is sensitive to both the context and the program design.

The second weak link concerns the assumption that new water sources improve drinking water quality at source. In this study, we find that the sources do improve drinking water quality with respect to fecal contamination, but they do not eliminate fecal contamination at source. However, we do not have measures of intensity of fecal contamination, so it is possible that our results are too pessimistic. In any case, these findings suggest the need for more research to specify the extent of fecal contamination in deep tubewells and the channels via which contamination occurs.

Otherwise, the simple theory of change we posited appears to accurately describe the behaviour we observed.

# Chapter 7

# Impact analysis and results of the key evaluation questions

## 7.1 Methodology

### 7.1.1 Program effects

**Pre-specified analyses** To causally estimate changes in average household water quality and in behaviour with respect to obtaining water for drinking and cooking, we primarily estimate reduced form "intent-to-treat" effects that exploit the random assignment of the program to treatment units.

$$\Delta y_{ic} = \alpha + \beta T_c + \eta_d + \epsilon_c \tag{7.1}$$

where $\Delta y_{ic}$ is the change in outcome variable $y$ between baseline and follow-up in household $i$ in community $c$,[1] $T_c$ is an indicator which takes the value 1 if community $c$ is assigned to treatment, and $\eta_d$ is a Union Parishad fixed effect. The estimated effects are the average intent-to-treat effects of the program — regardless of whether or not the program successfully installs water sources or not — so they are not contaminated by selection into successful installation.

The variables $\eta_d$ are controls that reflect stratification in the original randomization. We include controls for each lottery at which treatment was assigned.[2] Following Lin (2013), Imbens and Rubin (2015) and Gibbons et al. (2018), we demean the lottery fixed effects and include the interaction term between the lottery controls and the treatment dummies, meaning that $\beta_T$ estimates the average difference between treated and control villages.

[1]We depart from the pre-specified approach in one minor respect, and analyze data at the household level, applying weights so that each treatment unit counts equally in the analysis, and clustering standard errors at the treatment unit level. Our pre-specified approach was to collapse the data to village-level means. The estimated point effects are mechanically identical when we estimate at the household level, but are slightly more precisely estimated. This results from making less conservative adjustments to standard errors for the stratification controls.

[2]We ran one lottery in most unions, and two lotteries in one of the larger unions.

We report the pre-specified analyses for all pre-specified variables of interest, as summarized in Table 7.1. As noted in the pre-analysis plan, where multiple measures for a single outcome variable are listed, the expected main measure is given in bold, and the variables we anticipated using to provide corroborating evidence are listed in regular text.[3]

In our original power calculations, we modelled take-up of the overall intervention at 70% of communities, and used self-reported rates of intended adoption to model take-up of installed sources at the household level. In practice, average take-up was slightly lower, primarily in the cash contribution arm, and take-up at the household level was also lower than suggested by self-reported intentions at baseline. This means that we have somewhat less power to detect effects than estimated in our original power calculations. For this reason, we estimate an alternative analysis which partially accounts for the lower take-up, particularly in the cash contribution arm.

[3] The exception is the results based on the arsenic lab tests. We encountered some issues with the tracking of lab tests results at follow-up, and at the time of writing this report, these data are not yet available.

Table 7.1: Variables of interest

| Evaluation Question | Variables |
|---|---|
| 1a | **Arsenic fteld test of household water above WHO standard (10ppb)**<br>Arsenic field test of household water above Bangladeshi standard (50ppb)<br>Arsenic lab test of household water above Bangladeshi standard (50ppb)<br>Arsenic lab test of household water above WHO standard (10ppb)<br>Arsenic field test of household water result<br>Arsenic lab test of household water result |
| 1b | **Indicator for fecal contamination of household water** |
| 2a | **Arsenic fteld test of source water above WHO standard (10ppb)**<br>Arsenic field test of source water above Bangladeshi standard (50ppb)<br>Arsenic field test of source water result |
| 2b | **Indicator for fecal contamination of source water** |
| 2c | **Calculated distance between household and primary water source in metres**<br>Reported distance walked to collect safe drinking water in minutes |
| 2d | **Indicator for whether household is observed to obtain drinking water from storage**<br>Indicator for whether household reports regularly storing drinking water<br>Indicator for whether household reports/is observed storing water in an open container<br>Indicator for whether household reports/is observed storing water at floor level<br>Indicator for whether household reports/is observed scooping water from storage container |

**Additional analyses (not pre-specified)** We report one additional set of results were not included in the pre-analysis plan. As discussed in the previous section, take-up was very low under the cash contribution arm. The average program effects pool results across all three treatment arms. However, the effects are very small under the cash contribution arm, attenuating the overall results.[4] We therefore also estimate the impact of the program using a second approach which scales the effect by the size of the investment.

[4]We report effects separately for each contribution arm in Appendix D.1, noting that these results are not the main focus of the present study, and will be discussed in detail in other work.

To implement this second approach, we use the three treatment dummies as instruments to predict the number of installed wells per household in each treatment unit. Under the assumption that the effects of the program are directly proportional to the number of wells installed per household, the coefficients from these analyses can be interpreted as giving the average effect on safe drinking water of each well installed, normalized by the number of households. These results allow us to make first order estimates of how many wells per capita would need to be installed across rural Bangladesh to eliminate arsenic contamination in drinking water. Note, however, that these estimates yield a Local Average Treatment Effect, or effect on the compliers, meaning that they estimate the effect of each well successfully installed in the population of communities who successfully installed water sources.

For brevity, we largely report these secondary estimates only for the main outcome variables, as listed in Table 7.1. We report these results for two measures relevant to Key Research Questions 2c and 2d because, as discussed in Section 7.4, our results suggest that changes in reported distance may be a better measure of changes in distance than changes in calculated distance, in this context.

### 7.1.2 Mechanism

To analyze mechanisms (Key Research Question 3), we evaluate how changes in household bacterial contamination vary with changes in source contamination, transport distance and storage. We originally proposed two empirical approaches, specified in our pre-analysis plan: a difference-in-difference strategy and an instrumental variable (IV) approach.

The difference-in-difference approach is as follows. For household $i$, we estimate:

$$\begin{aligned} FC^h_{if} - FC^h_{ib} = b_0 + b_1(FC^w_{if} - FC^w_{jb}) + b_2(DIST^w_{if} - DIST^w_{ib}) \\ + b_3(STORAGE_{if} - STORAGE_{ib}) + \eta_c + \epsilon_i \end{aligned} \quad (7.2)$$

where all variables are measured at baseline $b$ and follow-up $f$; $FC^h$ is fecal contamination in household $i$'s drinking water and $FC^w$ is fecal contamination in household $i$'s water source; $DIST^w$ is the distance between household $i$ and its drinking water source; and $STORAGE$ is an indicator variable for whether or not household $i$ stores drinking water (as opposed to collecting drinking water on demand). $\eta_c$ is a community-level dummy variable that absorbs village-level average changes in the outcome variables and the right-hand side variables. We estimate versions of Equation 7.2 with and without these community-level dummy variables, as there was no clear ex-ante reason to prefer one approach over the other.[5] When we include the community-level dummy variables, Equation 7.2 only exploits within-community variation in changes in the right-hand side variables to estimate causal effects.

The difference-in-difference yields causal estimates under the assumption that changes in the right hand side variables are uncorrelated with other changes in household drinking water

[5]Either approach might increase precision, depending on the exact structure of $\varepsilon_i$.

contamination e.g. through changes in household hygiene practices. Such an assumption may be reasonable, since the program did not provide extensive or differential information on other types of health or hygiene behaviour. Additionally, as we show, the difference in difference results are stable when we include or exclude additional controls for storage behaviour or community fixed effects, suggesting that the effects of unobserved hygiene behaviour on contamination would have to be several orders of magnitude larger than the combined effect of community-level unobservables *and* storage on contamination to meaningfully affect the results . However, although assignment to the safe drinking water program is random, selection of locations for water source installation is determined, by consensus, at a community meeting. As a result, it remains possible that changes in distance to collect drinking water, or source water contamination, may be correlated with other changes that also affect household drinking water contamination, through other channels. These confounding factors might in principle bias the above analysis.

The IV approach we proposed leveraged our detailed baseline data to predict the locations of wells chosen by communities and in turn to construct instruments for predicted behaviour change that would rely only on variation induced by the experimental assignment to treatment. This approach would allow us to eliminate any potential bias undermining the difference-in-difference analysis, although we anticipated that the approach would have substantially lower precision. In practice, however, although we can successfully predict the locations of water sources using a variety of approaches, the instrumental variables we constructed were only weak predictors of behaviour change. The main reasons for the unexpectedly poor performance of the instruments concerns the unexpectedly small improvement in fecal contamination in source drinking water (discussed in Section 6) and the unexpected measurement issues with the GPS data (discussed in Section 7.4.1). Since the instruments fail standard tests for instrument strength, we discuss the details of the IV method and the results only in Appendix D.2.

## 7.2 Sample for analysis

To obtain the data for analysis, we merge the data from households and water sources at both baseline and follow-up surveys. The final sample is constructed as follows.

**Construction of sample for analysis** At baseline we successfully conducted the household survey with 6,529 households. At follow-up, we were able to locate 6,487 of these households and complete the interview with 6,484 of them. Among these 6,484 households, 6,431 gave consent to household water testing at both baseline and follow-up,[6] among which we have the household drinking water quality data (arsenic and fecal contamination) at both baseline and followup for 6,313 households, and the source water quality data at both baseline and follow-up for 6,162 households. The potential causes of missing observations are: i) we could not locate a matching record in the water source survey data; or ii) we could not uniquely match the fecal cont-

[6]A total of 6,481 gave consent to household water testing at baseline and 6,434 at follow-up.

amination test identifier with a result in our test result database.[7] The final panel sample consists of 6,051 households interviewed at both baseline and follow-up, and for which we have household and source water quality data from both rounds.

We focus on this sample to avoid changing sample between the main analyses. However, in some analyses we have fewer observations, primarily because we cleaned the location data of extreme outliers which largely reflect error in GPS coordinates (in measures of calculated distances) or enumerator error in recording walking times (in measures of reported distance).

**Aggregation of information from multiple water sources** At baseline, households reported using on average 1.03 water sources, in both treated and control communities. The number of water sources used on average by households increases between baseline and follow-up in both treated and control communities, to respectively 1.24 and 1.12, and this difference is statistically significant. At baseline, households obtained 99% of water from the primary source, in both treated and control communities. At follow-up, this share decreases to, respectively, 93% and 97% in treated and control communities, and this difference is statistically significant.

Where the household uses multiple water sources, the values of $FC^w$ and $DIST^w$ are weighted averages across the sources the household reports using, as we pre-specified in the pre-analysis plan. We weight each source by the fraction of drinking and cooking water a household collects from each source.

## 7.3 Balance

In this section, we confirm that the random assignment to treatment was successful in creating groups that are statistically equivalent with respect to baseline characteristics.

**Balance between treatment and control** Table 7.2 shows that the treated and con- trol groups are similar on socio-economic characteristics. We report individual balance checks for 12 variables, along with two tests for joint similarity on all 12 variables. When we compare treated to control communities, only individual test rejects that the means are equal at the 10% level, and neither joint test rejects that the means are equal. We also compare each contribution treatment arm to the control group separately (resulting in 36 individual tests and 6 joint tests). Four out of 36 individual tests reject equality of means at the 10% level, of which 2% also reject equality of means at the 5% level. These differences are approximately consistent with differences that could arise due to chance.

Table 7.3 repeats this exercise for measures of baseline water use. At first glance, these results are less reassuring, as 4 out of 10 tests reject equality of means between treated and control groups at baseline, suggesting that treated communities have higher arsenic contamination than

[7] We used locally produced barcodes, which occasionally contained duplicate ids.

control communities. However, all four tests that fail are for highly correlated variables; the joint tests, which account for correlation between these variables, do not reject equality of means between treated and control groups on all variables.

Table 7.2: Socio-economic characteristics - Balance check

| | Control | Treated | Cash | Labour | Waiver |
|---|---|---|---|---|---|
| Household size | 3.8<br>(0.081) | 3.9<br>(0.077) | 3.9<br>(0.089) | 3.9<br>(0.080) | 3.8<br>(0.088) |
| The household head is muslim | .99<br>(0.022) | .98<br>(0.012) | .98<br>(0.023) | .97<br>(0.023) | .99<br>(0.025) |
| The household head has no education | .46<br>(0.034) | .47<br>(0.031) | .45<br>(0.036) | .47<br>(0.038) | .49<br>(0.034) |
| The household owns livestock | .78<br>(0.033) | .81<br>(0.023) | .82*<br>(0.026) | .79<br>(0.025) | .81<br>(0.027) |
| The household owns land for cultivation | .58<br>(0.035) | .59<br>(0.027) | .58<br>(0.030) | .61<br>(0.029) | .57<br>(0.032) |
| Land owned by the household (acres) | 1.1<br>(0.118) | 1<br>(0.068) | .98<br>(0.091) | .95<br>(0.093) | 1.1<br>(0.118) |
| HH has some toilet facility | .85<br>(0.025) | .85<br>(0.020) | .83<br>(0.023) | .86<br>(0.025) | .85<br>(0.023) |
| Number of rooms to sleep | 1.9<br>(0.049) | 1.9<br>(0.041) | 1.9<br>(0.046) | 1.9<br>(0.049) | 1.8*<br>(0.043) |
| The floor is made of earth or sand | .86<br>(0.035) | .86<br>(0.031) | .87<br>(0.031) | .86<br>(0.036) | .85<br>(0.033) |
| The roof is made of metal | .95<br>(0.017) | .97<br>(0.011) | .97<br>(0.013) | .98**<br>(0.012) | .96<br>(0.013) |
| Mobile phone ownership | .65<br>(0.053) | .7*<br>(0.040) | .69<br>(0.043) | .72*<br>(0.046) | .7<br>(0.044) |
| Ownership of a motorized vehicle | .042<br>(0.011) | .055*<br>(0.010) | .057<br>(0.011) | .063**<br>(0.012) | .043<br>(0.011) |
| Pvalue of F-test for joint significance | | 0.449 | 0.316 | 0.005 | 0.501 |
| Pvalue of Hotelling's T-Squared test | | 0.309 | 0.559 | 0.099 | 0.805 |

Notes: The table reports means and standard errors (in parentheses). Standard errors are clustered at "Treatment Unit" level. Significance levels are obtained from a regression at household level of each outcome variable on indicators for the treatment assignments (with no constant and Union Parishad dummies) and pairwise tests of the difference between the means of each treatment group versus the control group. The F-test is obtained by regressing indicators for treatment status on the full set of controls (including Union Parishad dummies) and testing for joint significance.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7.3: Water-related characteristics - Balance check

| | Control | Treated | Cash | Labour | Waiver |
|---|---|---|---|---|---|
| Arsenic contamination (WHO) (HH test) | .53<br>(0.064) | .6**<br>(0.063) | .61*<br>(0.070) | .61*<br>(0.069) | .58<br>(0.070) |
| Arsenic contamination (BD) (HH test) | .15<br>(0.046) | .22**<br>(0.044) | .21<br>(0.048) | .21<br>(0.046) | .25**<br>(0.053) |
| Fecal contamination (HH test) (predicted) | .59<br>(0.026) | .56<br>(0.024) | .56<br>(0.028) | .57<br>(0.027) | .56<br>(0.027) |
| WS arsenic contamination (WHO) | .62<br>(0.070) | .71**<br>(0.070) | .71**<br>(0.078) | .72**<br>(0.076) | .69<br>(0.074) |
| WS arsenic contamination (BD) | .23<br>(0.050) | .32***<br>(0.049) | .32**<br>(0.054) | .3<br>(0.054) | .35***<br>(0.059) |
| WS fecal contamination (predicted) | .56<br>(0.022) | .55<br>(0.016) | .54<br>(0.020) | .55<br>(0.021) | .55<br>(0.020) |
| The water is treated to make it safe for drinking | .16<br>(0.039) | .17<br>(0.043) | .16<br>(0.038) | .2<br>(0.048) | .15<br>(0.039) |
| Storage dummy (observed) | .64<br>(0.031) | .65<br>(0.024) | .65<br>(0.026) | .67<br>(0.028) | .62<br>(0.028) |
| Time needed to collect water (mins) | 2.1<br>(0.094) | 2.1<br>(0.061) | 2.1<br>(0.077) | 2.1<br>(0.074) | 2.1<br>(0.092) |
| Water collected per day (litres) | 53<br>(4.406) | 53<br>(3.773) | 53<br>(4.350) | 51<br>(4.415) | 55<br>(4.153) |
| WTP for a new WS in a socially optimal location | 88<br>(12.311) | 98<br>(12.073) | 103<br>(13.384) | 103<br>(14.206) | 85<br>(12.035) |
| Pvalue of F-test for joint significance | | 0.153 | 0.063 | 0.109 | 0.007 |
| Pvalue of Hotelling's T-Squared test | | 0.612 | 0.536 | 0.442 | 0.234 |

Notes: The table reports means and standard errors (in parentheses). Standard errors are clustered at "Treatment Unit" level. Significance levels are obtained from a regression at household level of each outcome variable on indicators for the four treatment assignments (with no constant and Union Parishad dummies) and pairwise tests of the difference between the means of each treatment group versus the control group. The F-test is obtained by regressing indicators for treatment status on the full set of controls (including Union Parishad dummies) and testing for joint significance.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

## 7.4 Main Results

### 7.4.1 Program effects

**Pre-specified analyses** Table 7.4, Table 7.5 and Table 7.6 present the mean program effects on Key Research Questions 1, Questions 2a and b, and Questions 2c and d, respectively. In all Tables, the coefficient reported as the constant corresponds to the mean change in the outcome variable between baseline and follow-up in the control group, while the coefficient labelled "Treated" corresponds to the estimated treatment effect.

**Key Research Question 1: What is the average effect of the program on household water quality?** Table 7.4 reports mean effects of the program on household water quality. Column 1 shows the main results for arsenic contamination in household water. The average program impact was a 2.2 percentage point reduction in arsenic contamination at the WHO standard in household drinking water. The effect is imprecisely measured and the confidence interval does not exclude zero.

Columns 2 and 3 shows evidence on alternative measures of arsenic contamination: arsenic contamination at the higher Bangladeshi threshold falls slightly in treated communities (column 2) although arsenic test results actual rise on average, albeit insignificantly so (column 3). The reason for this result is that arsenic contamination is highly skewed: 38% of households have drinking water with no contamination, while 1.5% have arsenic contamination above 250 ppb and 0.3% have arsenic contamination above 500 ppb. The analyses using the test results are therefore sensitive to a small number of outliers and are not very well-equipped to detect small changes in arsenic contamination. Figure 7.1 illustrates: the effects are primarily concentrated in a larger fraction of households who experience relatively small reductions in arsenic concentration and a smaller fraction of households who experience relatively small increases. There is no effect in the sparsely populated tails of the distribution, which receive the greatest weight in the analysis which uses the test results.

Column 4 shows the main average effects on fecal contamination. Ex ante, it was ambiguous whether the program would increase or decrease exposure to fecal contamination in drinking water. The results in column 4 suggest essentially no effect on fecal contamination in household drinking water: the point estimate is a 0.2% increase in contamination, but the 95% confidence interval spans both modest increases (4.1%) and modest decreases (3.7%) in contamination.

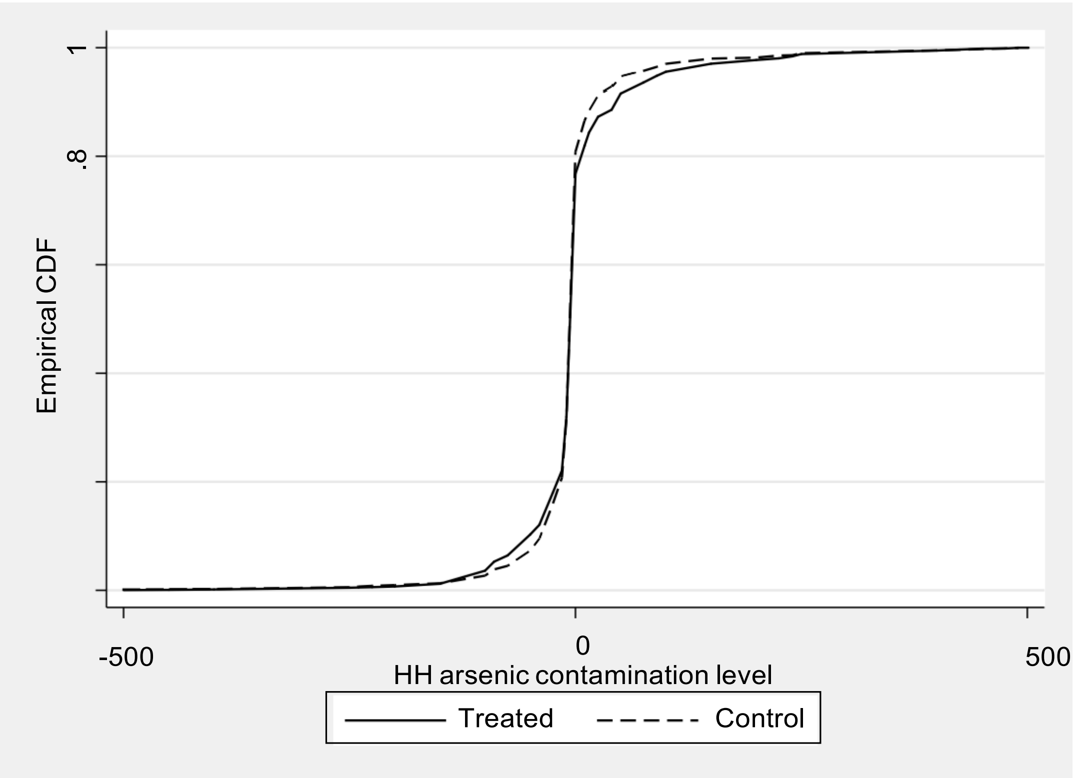
Note that the control group experiences improvements in household water quality, particularly with respect to arsenic contamination. There are a number of potential explanations for these changes, including changes in behaviour as a result of the information we provided about water source quality at baseline, changes in the way we measured contamination, or changes in source contamination, possibly as a result of the floods in the region which occurred just before our follow-up survey. We discuss the likely reasons for these changes further in the next section.

Table 7.4: Effect of the program on household water quality

| | Arsenic contamination (WHO) | Arsenic contamination (BD) | Arsenic contamination level | Fecal contamination |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treated | -0.022 | -0.004 | 3.756 | 0.002 |
| | (0.020) | (0.018) | (2.865) | (0.019) |
| Constant | -0.096*** | -0.023 | 1.218 | -0.010 |
| | (0.016) | (0.015) | (2.414) | (0.017) |
| R2 | 0.02 | 0.01 | 0.04 | 0.03 |
| Obs | 6051 | 6051 | 6051 | 6048 |

Notes: Table shows estimated average program impact on listed household water quality measure. Regression in first differences, including stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

Figure 7.1: CDF of arsenic field test result in household sample



**Key Research Question 2: How did the program change behaviour with respect to obtaining water for drinking and cooking?** Table 7.5 shows the estimated effect on source water quality. Column 1 shows that the average program impact is a 5.6 percentage point reduction in the volume of water obtained from sources above the WHO contamination level. Columns 2 and 3 provide additional evidence on other measures of arsenic contamination in water sources: the volume of water obtained from sources above the Bangladeshi contamination level falls by 2.7 percentage points while the weighted-average arsenic contamination level in sources used falls by just under 0.1 ppb (not statistically different from zero). Figure 7.2 visualizes these results.

Column 4 shows a modest decrease in the share of water obtained from sources with fecal contamination of 1.5 percentage points. However, the confidence interval does not exclude zero. The effect on source fecal contamination is less than 30% of the effect on arsenic contamination. This likely reflects the fact that fecal contamination in project sources was only 28% (13 percentage points) lower than other sources in the same communities. In contrast, the reduction in arsenic contamination in water sources was 92% (58 percentage points).[8]

As in Table 7.4, Table 7.5 shows that households in the control group experienced large reductions in arsenic contamination at source as well as in the household. We do see evidence for slightly increased use of multiple sources in the control group as well and for adoption of new sources. However, if we analyze the change in contamination rates in sources for which we have contamination readings at both baseline and follow-up, we see that sources experienced on average a 7% fall in arsenic contamination rates at the WHO standard and a 1% fall in contamination rates at the Bangladeshi standard.[9] These results suggest that the changes in contamination seen in the control group are likely the consequence of some kind of secular change, either in arsenic contamination in groundwater — possibly because of the flood events — or changes in how our enumerators measured arsenic contamination, rather than the consequence of any systematic response in the control group to information about arsenic.

[8]See Table 6.4.

[9]Curiously they also experienced a rise in arsenic contamination levels, driven by a higher rate of extremely high values of arsenic contamination.

Table 7.5: Effect of the program on source(s) water quality

| | Arsenic contamination (WHO) | Arsenic contamination (BD) | Arsenic contamination level | Fecal contamination |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treated | -0.056** | -0.027 | -0.155 | -0.015 |
| | (0.023) | (0.019) | (3.648) | (0.019) |
| Constant | -0.068*** | -0.008 | 7.642** | -0.001 |
| | (0.018) | (0.015) | (3.167) | (0.015) |
| R2 | 0.03 | 0.02 | 0.07 | 0.04 |
| Obs | 6051 | 6051 | 6051 | 5993 |

Notes: Table shows estimated average program impact on listed water source quality measure. Regression in first differences, including stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

Figure 7.2: CDF of arsenic field test result in household sample

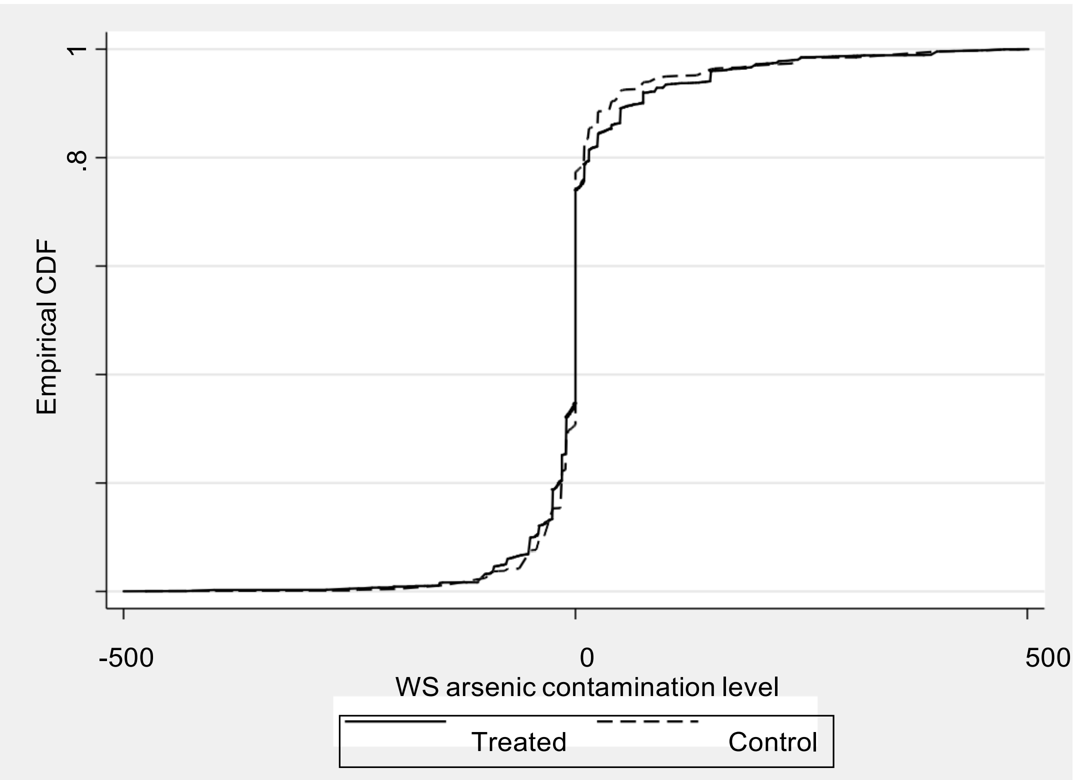


Table 7.6 also shows the effects of the program on practices related to the transport and storage of drinking water. Columns 1 and 2 show changes in distance travelled to collect safe drinking water using measured distances (column 1) and reported distances (column 2). The results are somewhat different. Column 1, which uses calculated data, shows a decrease of 0.1m in treated communities relative to control communities while column 2, which uses reported data, shows an increase of 0.066 minutes, equivalent to about 5m walking distance.

Table 7.6: Effect of the program on water related practices

| | Distance HH-WS (m) | Distance HH-WS (min) | Observed storage | Reported storage |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treated | -0.147 | 0.066** | 0.002 | 0.003 |
| | (1.041) | (0.027) | (0.028) | (0.032) |
| Constant | -2.801*** | 0.012 | -0.134*** | -0.040 |
| | (0.685) | (0.024) | (0.024) | (0.029) |
| R2 | 0.01 | 0.01 | 0.02 | 0.02 |
| Obs | 5832 | 5729 | 6050 | 6051 |

Notes: Table shows estimated average program impact on listed measure of water-related practice. Regression in first differences, including stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

Note, however, that there is a potential concern with our measure of changes using calculated data. Measurement error in GPS coordinates leads to overestimation of distances (Ranacher et al., 2016). The reason is twofold: first, any measurement error that is orthogonal to the true distance leads to an increase in the measured distance. Second, at distances within the support of the distribution of measurement error, the facts that the two points we measure distance between may reverse in orientation and that distance is an absolute measure of proximity means that errors that lead us to overestimate distance are not cancelled out by errors that lead us to underestimate distance. To see this most intuitively, consider two measures of the same location, for which the true distance is obviously zero. Any measurement error in the location measures leads us to estimate a non-zero distance between the points. The second problem increases in magnitude as the measured distances decrease in size relative to the measurement error in the GPS coordinates. In our case, the measured distances are indeed small relative to the measurement error.

This potentially affects our results in the following way. We measure the location of our installed water sources with greater accuracy than other sources, because we draw on multiple measures of location and we verify the locations by inspection. As a result, if households adopt our water sources, we overestimate the distance between their household and our water source less than we overestimate that distance between their household and the water source they used at baseline, and consequently, we underestimate any increase in distance travelled. These biases could be sufficient to cancel out any true increase in distance travelled, explaining the difference in results between columns 1 and 2.

Note that this would also provide an explanation for why column 1 suggests a reduction on average of distance between baseline and follow-up for all study households households, while column 2 does not find any reported change in distance. This is because we used more accurate tablets as follow-up than baseline. We therefore overestimate distances between households and sources that were only measured at baseline relatively more than distances between households and sources that were only measured at followup.[10]

Columns 3 and 4 show that there are unlikely to be large changes in storage practice as a consequence of the intervention, since the differences between treated and control communities are small.

Table 7.7 provides additional measures of water storage practice. Note that not all these measures were recorded at baseline and follow-up, so in some cases these analyses use only follow-up data. Across these measures, there is a weak increase in relatively unsafe storage practices in the treated group relative to the control group, with the exception of the measure of whether or not the water is scooped from its container as opposed to being poured. Only one of these differences is significant at the 10% level. These results suggest that the program leads at most to small changes in unsafe storage behaviour.

The substantial decrease in incidence of uncovered storage, and to a lesser extent whether containers are stored on the floor, that is seen in the control group, is somewhat surprising but may possibly reflect seasonal differences in storage practice.

Table 7.7: Effect of the program on storage practices

| | Containers are uncovered | Containers are on the floor | Containers are uncovered (observed) | Containers are on the floor (observed) | Water is scooped (obserbed) |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Treated | 0.023<br>(0.039) | 0.008<br>(0.030) | 0.040*<br>(0.021) | 0.025<br>(0.019) | -0.002<br>(0.022) |
| Constant | -0.351***<br>(0.033) | -0.092***<br>(0.027) | 0.383***<br>(0.018) | 0.484***<br>(0.016) | 0.439***<br>(0.020) |
| Only endline data | | | ✓ | ✓ | ✓ |
| R2 | 0.04 | 0.02 | 0.02 | 0.01 | 0.01 |
| Obs | 6051 | 6051 | 5902 | 5997 | 6029 |

Table shows estimated average program impact on listed measure of storage practice. Regression in first differences, unless otherwise indicated, and includes stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

[10]For sources where we have locations at both baseline and follow-up, we average the data after excluding outliers.

**Instrumental variables estimates of impact of wells (not pre-specified)** The mean program estimates reported so far are intent to treat estimates, in that they estimate the average effect of the program regardless of whether any wells were actually installed. We can alternatively use assignment to one of the three treatment arms as instruments for the number of wells installed per household. Using this approach, we can estimate the effect of well installation on the treated, under the assumption that program effects only operate via the provision of new wells.

The effects reported in Table 7.8 imply that installing one well in a community of 100 households would decrease household arsenic contamination by about 4.6 percentage points but increase household fecal contamination by about 1.6 percentage points, although, in both cases, we also cannot reject either null effects or small effects in the opposite directions.

Table 7.8: Effect of water sources installed per household on household water quality: Instrumental variables estimates

| | Arsenic contamination (WHO) | Fecal contamination |
|---|---|---|
| | (1) | (2) |
| Wells installed per household | -4.551 | 1.628 |
| | (3.010) | (2.560) |
| First stage F-statistic | 93.5 | 93.4 |
| R2 | 0.01 | 0.03 |
| Obs | 6051 | 6048 |

Notes: Table shows estimated impact of installed wells. Regression in first differences, using dummies for assignment to the three treatment arms to predict number of installed wells per household and including stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

Similarly, the effects reported in Table 7.9 suggest that installing one water source in a community of 100 households would decrease arsenic contamination from water sources by about 10 percentage points. The estimated effect on arsenic in source water might be larger than the effect of arsenic in household water samples for one of at least two reasons. First, the water source measure may be more likely to be affected by reporting bias, because households may report that they use a source installed by the project because of experimenter demand effects. However, as we discuss in Section 8, this appears unlikely to be the case. Second, the pattern of results might be a product of the way we measure outcomes: using a greater fraction of water from an arsenic safe source might reduce arsenic contamination in household water, but the effect might not be sufficiently large to bring household arsenic contamination below the threshold at which we measure arsenic contamination. The estimated effect on fecal contamination suggests that installing one water source in a community of 100 reduces source water contamination by 4 percentage points.

Together, the results suggest that each well installed in a community of 100 households is sufficient to eliminate arsenic contamination for between 5 and 10 households. We will use these estimates in our cost-effectiveness analyses, where we also compare these estimates to other estimates available in the literature.

Instrumental variables analyses of the transport variables, in this case, both suggest positive effects on distance to safe drinking water, although the effects on measured distance are considerably smaller in magnitude than the effects on reported distance, once we adjust the estimated effect in metres to an estimated effect in minutes. Again, this pattern of results is consistent with the hypothesis that measurement error contaminates the results on calculated distances. The instrumental variables estimates on storage also suggest small positive effects on the likelihood that water is stored before drinking, but we cannot rule out null effects or small negative effects.

Table 7.9: Effect of water sources installed per household on water source quality: Instrumental variables estimates

| | Arsenic contamination (WHO) | Fecal contamination |
|---|---|---|
| | (1) | (2) |
| Wells installed per household | -9.654*** | -4.082 |
| | (3.281) | (2.665) |
| First stage F-statistic | 93.5 | 93.9 |
| R2 | 0.03 | 0.04 |
| Obs | 6051 | 5993 |

Notes: Table shows estimated impact of installed wells. Regression in first differences, using dummies for assignment to the three treatment arms to predict number of installed wells per household and including stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

Table 7.10: Effect of water sources installed per household on transport and storage practice: Instrumental variables estimates

| | Distance HH-WS (m) | Distance HH-WS (min) | Observed storage | Reported storage |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Wells installed per household | 89.111 | 12.373*** | 0.878 | 1.962 |
| | (176.038) | (3.273) | (3.725) | (3.979) |
| First stage F-statistic | 89.4 | 91.7 | 93.7 | 93.5 |
| R2 | 0.01 | 0.01 | 0.02 | 0.02 |
| Obs | 5832 | 5729 | 6050 | 6051 |

Notes: Table shows estimated impact of installed wells. Regression in first differences, using dummies for assignment to the three treatment arms to predict number of installed wells per household and including stratification controls. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Standard errors clustered by community.

### 7.4.2 Mechanism

Table 7.11 shows the results of the difference-in-difference analysis using the measure of reported travel time to water sources. Table 7.12 shows the results of the same analysis using the inferred travel time using GPS coordinates of both water sources and households. Throughout this section, we convert distances calculated in metres to travel times in minutes assuming an average walking speed of about 80 m/minute.[11]

In columns 2 and 4, we include treatment unit fixed effects, meaning that we exploit only within treatment unit variation in outcomes. In columns 1 and 2, we include changes in observed storage in the regression; in columns 3 and 4, we omit these variables.

The results suggest that switching to a source with fecal contamination increases the household-level risk of contamination by about 22 to 25 percentage points, an estimate that is extremely stable across specifications. Storing drinking water also increases the risk of contamination by on average 6 to 8 percentage points. Again, this effect is extremely stable across specifications.

In contrast, the results in Tables 7.11 and 7.12 yield quite different conclusions regarding the effects of distance. The results in columns 1 and 2 of Table 7.12 suggest that increasing travel time by one minute increases the risk of contamination by at most 0.5 percentage points, but the estimates reported do not rule out the possibility that increasing travel time has no effect on contamination, and the confidence intervals include both quite substantial positive effects and quite substantial negative effects. The results in columns 1 and 2 of Table 7.11 instead suggest that increasing travel time by one minute increases the risk of contamination by about 1.5 to 1.9 percentage points. It seems likely that the difference between the two sets of results is explained by increased measurement error in the measured distances, biasing the effects towards zero through attenuation bias.

Omitting the controls for changes in storage practice, as shown in columns 3 and 4, leads to fractionally larger estimated effects of transport time on drinking water contamination. This is because increasing travel times are associated (weakly)[12] with increasing storage, and storage is in turn positively correlated with household-level contamination.

As discussed in Section 7.1.2, we originally prespecified an alternative instrumental variables approach. However, the instruments we constructed do not have sufficient predictive power to yield reliable results (a weak instrument problem). As a result, the point estimates are very imprecisely estimated, although they do largely take the same sign as the difference-in-difference analyses. We therefore discuss the results of these analyses only in Appendix D.2.

[11]This is a simple rescaling, so the conclusions are not sensitive to the rescaling factor used, but this helps to give the coefficients a meaningful interpretation.

[12]Results available on request.

Table 7.11: Mechanism: Analysis using reported travel time

| | Drinking water fecal contamination | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Source fecal contamination | 0.238***<br>(0.015) | 0.219***<br>(0.015) | 0.241***<br>(0.015) | 0.220***<br>(0.015) |
| Travel time hh-ws (mins, reported) | 0.015*<br>(0.008) | 0.018**<br>(0.008) | 0.016**<br>(0.008) | 0.018**<br>(0.008) |
| Observed storage | 0.080***<br>(0.010) | 0.069***<br>(0.009) | | |
| Constant | 0.001<br>(0.010) | -0.001<br>(0.001) | -0.010<br>(0.010) | -0.011***<br>(0.001) |
| Treatment unit fixed effects | No | Yes | No | Yes |
| R2 | 0.07 | 0.14 | 0.06 | 0.13 |
| Obs | 5673 | 5673 | 5674 | 5674 |

Notes: Regression in first differences. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Treatment unit fixed effects where specified. Standard errors clustered by community.

Table 7.12: Mechanism: Analysis using travel time measured using GPS coordinates

| | Drinking water fecal contamination | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Source fecal contamination | 0.238***<br>(0.016) | 0.219***<br>(0.016) | 0.242***<br>(0.016) | 0.221***<br>(0.016) |
| Travel time hh-ws (mins, measured) | -0.001<br>(0.015) | 0.005<br>(0.013) | 0.003<br>(0.015) | 0.009<br>(0.013) |
| Observed storage | 0.073***<br>(0.010) | 0.063***<br>(0.009) | | |
| Constant | 0.001<br>(0.010) | -0.000<br>(0.001) | -0.008<br>(0.010) | -0.008***<br>(0.000) |
| Treatment unit fixed effects | No | Yes | No | Yes |
| R2 | 0.07 | 0.14 | 0.06 | 0.13 |
| Obs | 5774 | 5774 | 5775 | 5775 |

Notes: Regression in first differences. Analysis is at the household level with weights applied so that each community counts equally in the analysis. Treatment unit fixed effects where specified. Standard errors clustered by community.

## 7.5 Heterogeneous impacts

In this section we summarize the results of pre-specified heterogeneity analyses. We report the details of these analyses in Appendix D.3

### 7.5.1 By use of safe/unsafe sources at baseline

Households with high arsenic contamination at baseline show the largest reductions in arsenic contamination at source and the largest reductions in fecal contamination at source and at home. These households are also the least likely to increase distance to collect water and decrease their likelihood of being observed storing water before drinking. The differences are small and not very precisely measured, but one explanation for these results is that these houses are closest to the installed sources (an algorithm that predicts chosen locations indeed places most weight on households with high contamination) and therefore least likely to experience the negative effects of increased transport time. However, it is possible that such differences could arise due to chance.

### 7.5.2 By distance to source

Descriptive evidence shows that both reductions in household arsenic contamination and fecal contamination are larger closer to constructed sources in successful treatment units. We observe similar patterns for changes in source arsenic contamination, although not for source fecal contamination.

With respect to changes in transport time, the calculated measures of changes in distance do not vary systematically with distance to a constructed source. The reported measures of changes in distance do vary systematically with distance to constructed wells, with those closer to the well increasing distance more. While this may seem counterintuitive, it probably reflects higher uptake closer to the source dominating the effect of needing to walk further at higher distances. Changes in storage do not exhibit clear patterns with distance to a source.

However, this evidence is simply descriptive, and may not reflect causal relationships, because the relationships we estimate may be confounded by other characteristics that are correlated with distance to selected or successful installation locations. We also evaluate how the effects of the program vary with distance to the predicted locations, relative to households in control communities at the same distance from the predicted location. These analyses provide limited evidence for systematically varying effects, but this may reflect a lack of power to distinguish heterogeneous effects rather than the absence of heterogeneous effects. The exception is a systematic decline in the effects on water source quality for arsenic with increasing distance from the predicted source.

### 7.5.3 By self-reported poverty level at baseline

The effects on arsenic contamination are similar across all income categories. The effects on household fecal contamination are also similar, at least across the three main income categories.

However, the effects on water source contamination do exhibit some striking differences by poverty level: the middle and upper income groups experience substantial reductions in source fecal contamination while the poor experience substantial increases. The differences are sufficiently large that they would likely survive corrections for multiple hypothesis testing. One possible explanation is that water sources used by many poor households become contaminated more quickly, while deep tubewells used by middle and upper income households remain uncontaminated through use patterns and are thus more likely to be able to realise the potential gains in water quality.

## 7.6 Cost-effectiveness

The following is an upper bound on the cost-effectiveness of the program. Each source we install eliminates arsenic contamination for 5-10 households, containing on average 3.9 individuals, so between 20 and 40 individuals in total. The average cost of well installation is 60,000 BDT or approximately US$720 at current exchange rates. Including *only* the installation costs, the cost of avoiding arsenic contamination is between 6000 BDT and 12000 BDT per household (between US$72 and US$144). In per capita terms, these ranges are 1540 BDT to 3080 BDT or US$18.5 and US$37. These costs are quite substantial, even without factoring in the labour costs and overheads of project implementation. However, they provide a useful benchmark for comparison with alternative approaches to providing safe drinking water in rural Bangladesh.

Additionally, the costs of implementing the program may be higher, because we used a baseline water testing program to target communities with arsenic contamination. Communities also used this information to select locations for installation. Lacking this information, the program might have been less successful in targeting communities and communities might have selected locations with lower arsenic contamination. Collecting baseline water source census data is relatively costly. A key question for future research is whether the benefits of this information in improved targeting would justify its costs at large scale.

For comparison, Jamil et al. (2019) estimate the cost per person with reduced exposure via deep tubewell installation to be be between US$9 (under the best possible siting conditions and assuming 60% uptake within a 100m radius, higher than the uptake we see in this context) and US$142 (for poorly-targeted wells installed by governments, with very high levels of elite capture. The cost-effectiveness estimates we find are closer to the “best case scenario” values.

# Chapter 8

# Discussion

## 8.1 Internal validity

**Measurement concerns** A primary concern with the internal validity of our findings is the inconsistency between results based on measured and reported changes in distance. We believe that there are plausible explanations for the differences in results, but the inconsistency in the results remains somewhat disconcerting. However, we note that our view is that in interpreting the results, it is more conservative to place relatively more weight on the results using reported distance. This is because the policy conclusion that we might reach given the results using calculated distance is that travel distance has little effect on contamination. The results using reported distance suggest that the in fact there are negative effects of increasing travel distance, although they are relatively small. In taking policy decisions, it may be cautious to place more weight on the more pessimistic estimates. Taken together, however, both sets of results suggest that large effects of transport on contamination are unlikely.

**Spillover** Our program targets communities that are highly arsenic contaminated in 10 Union Parishads. Villages enrolled in our project lie in relatively small geographical areas (see Figure 5.2). Moreover, because we divide large villages in several treatment units (see Section 6.3), it is not uncommon that control and treated communities are adjacent or very nearby. Despite this geographical proximity between control and treated communities, we observe *no spillovers* from the treatment to the control group in terms of take-up of wells. In control villages, no household interviewed at follow-up reported using any project tubewell.[1] Given the local context, the absence of spillovers was largely expected. As discussed in Section 5.7, households have a strong preference for local sources. Indeed, only 0.8% of households interviewed at follow-up report to use a water source in a different cluster than their own.

[1] This is not a mechanical result. There were no restrictions on data collection to constrain households from selecting water sources in communities other than their own.

**John Henry or Hawthorne effects** We expected John Henry or Hawthorne effects to operate primarily through changes in hygiene behaviour or source selection in control villages. The detailed analysis of storage practices suggests that if anything, treated households have slightly worse hygiene practices at baseline, being more likely to store water, to store water in uncovered containers and to store water at floor level. These effects tend to offset differences between treated and control households. Additionally, the changes in source water contamination we see in the control group appear more likely to be explained by secular changes that are outside of the household's controls than by systematic compensatory source switching in response to information about arsenic contamination.

**Reporting bias** Comparing household and source measures of arsenic contamination provides us with a mechanism for evaluating the extent of response bias. Since arsenic contamination only takes place via source contamination, then differences between household and source contamination is primarily driven by measurement error, potentially including reporting bias. Further, and more importantly, we can compare whether the difference between household and source contamination varies between treated and control groups. This enables us to evaluate whether receiving the safe drinking water program alters reporting of behavior, as well as the behavior itself, an important question for future evaluation programs. We find that there is no difference between treated and control groups in the relationship between source and household contamination, for arsenic contamination at both the Bangladeshi and WHO threshold. We also do not find differences in the correlation between source and household fecal contamination between treated and control villages.[2] These results provide reassurance that our findings are unlikely to influenced by patterns of differential reporting bias.

## 8.2 External validity

While our specific findings are most applicable to the context of rural Bangladesh, our findings are potentially generalizable to other settings. Like Kremer et al. (2011), we find that source water quality only partially explains household water quality. In our case, our results suggest that eliminating source fecal contamination would only reduce household contamination by at most 25%. These results confirm that fecal contamination of household drinking water is difficult to eliminate in contexts where water is collected and stored in the household.

We also find that the sources themselves retain substantial levels of fecal contamination, although we cannot determine whether there are improvements on the intensive margin (i.e. lower concentrations of fecal bacteria) due to the limitations of the fecal contamination test we used, for budgetary reasons, in this study.

We find limited support for the hypothesis that households walking increasing distances or

[2] Detailed results available on request.

storing water for longer as a result of switching to more distant arsenic-safe sources can explain the results in Field et al. (2011). In our context, the effects of distance and storage are relatively modest in size, and households show limited responsiveness to the intervention in terms of changes in storage and transport behaviour. On the other hand, our data do confirm, as others have previously noted, that there is an inverse correlation between fecal contamination and arsenic contamination in shallow tubewells. The results in Field et al. (2011) could therefore still be explained by switching to sources with higher fecal contamination in an effort to avoid arsenic contamination. We note, however, that we find very little correlation in our control group, either positive or negative, in changes in source arsenic contamination and changes in source fecal contamination.

## 8.3 Influence of treatment design on the results

A key aspect of how the study design influences the results is that the impact of the program was much lower under the cash contribution arm than under the labour contribution and contribution waiver arms. The mean intent to treat estimate of the arsenic mitigation program is likely smaller than it would have been had we implemented the program under one of the other two contribution treatments in all communities. These findings are important, however, because program take-up is a key determinant of the impact of safe drinking water programs, and our findings confirm that the impacts of safe drinking water programs vary with key features of program design.

## 8.4 Key lessons for researchers

A key finding from this study is the difficulty of using GPS coordinates measured with phones or tablets to measure small distances, especially when the extent of measurement error varies between different objects. To our knowledge, the extent of measurement error in GPS coordinates measured with tablets has not been comprehensively documented. We will endeavour to provide more systematic documentation of the impacts of measurement error in this context to provide input for researchers in designs relying on these technologies. We also note, however, that the performance of tablets in measuring GPS is likely changing rapidly with time.

# Chapter 9

# Findings for policy and practice

## 9.1 Policy

The key findings for policy-makers designing programmes to provide safe drinking water in rural Bangladesh are the following:

**Each deep tubewell installed provides arsenic safe water to between 5 and 10 households** This finding implies that to resolve the arsenic problem in Bangladesh, program implementers would need to identify highly contaminated communities and budget for at least one new source for every 10 households affected by arsenic contamination. The cost of constructing these wells would be at least US$18.5 per capita, so the cost of well installation for the rural population affected by arsenic contamination would be above US$700 million. For comparison, installing local piped water supply systems could cost US$150 per capita and in some contexts it appears that simply providing information about arsenic contamination could lead to well-switching at a cost of less than US$1 per capita (Jamil et al., 2019). However, we do not find evidence for widespread well-switching in our control group, in which we provide full information about water source quality but no subsidies for well construction or incentives to share sources.

**Deep tubewell programs alone have little impact on fecal contamination** Deep tubewells reduce, but do not eliminate source fecal contamination, at least not with current use and maintenance practices. Households increase transport times and possibly change their storage behaviour to adopt slightly more distant sources. Greater transport times and longer storage increase the risk of fecal contamination in household drinking water. Both effects are small, in part because households rarely walk more than, at most, 4 or 5 minutes, to collect drinking water. The improvements in source contamination and recontamination effects offset each other, so that the net effect on contamination in household drinking water is very small. Interventions to eliminate exposure to fecal contamination must therefore adopt other alternative approaches, possibly including an increased focus on storage practices, tubewell maintenance, or other hygiene measures.

## 9.2 Programme and implementation

Features of project design are important determinants of success, including, in this context, the approach taken to community contribution requirements. In other work, we previously showed that approaches to decision-making are also important determinants of impact in programs to provide safe drinking water in rural Bangladesh (Madajewicz et al., 2018). The results from these studies emphasise the need for rigorous and systematic evaluation of different approaches to program design to maximize the impact of safe drinking water programs.

# Bibliography

Ahuja, A., Kremer, M., and Zwane, A. P. (2010). Providing safe water: Evidence from randomized evaluations. *Annual Review of Resource Economics*, 2.

BBS and UNICEF (2015). Bangladesh Multiple Indicator Cluster Survey 2012-2013: Progotir Pathey. Bangladesh Bureau of Statistics and United Nations Children's Fund.

Caldwell, B. K., Caldwell, J. C., Mitra, S. N., and Smith, W. (2003). Tubewells and arsenic in Bangladesh: Challenging a public health success story. *International Journal of Population Geography*, 9.

Clasen, T., Roberts, I., Rabie, T., Schmidt, W., and Cairncross, S. (2006). Interventions to improve water quality for preventing diarrhoea. *Cochrane Database of Systematic Reviews*, 3.

Daily Observer (2014). Tube-well water arsenicated in Bogra: 161 villages in eight upazilas affected. Newspaper article.

FAO (2016). AQUASTAT main database, Food and Agriculture Organization of the United Nations (FAO). Website.

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 Disease*, 5.

Field, E., Glennerster, R., and Hussam, R. (2011). Throwing the baby out with the drinking water: Unintended consequences of arsenic mitigation efforts in Bangladesh. Working Paper.

Gibbons, C. E., Serrato, J. C. S., and Urbancic, M. B. (2018). Broken or fixed effects? *Journal of Econometric Methods*, Forthcoming.

Gupta, S., Sheikh, M., Islam, M., Rahman, K., Jahan, N., Rahman, M., Hoekstra, R., Johnston, R., Ram, P., and Luby, S. (2008). Usefulness of the hydrogen sulfide test for assessment of water quality in Bangladesh. *Journal of Applied Microbiology*, 104.

Human Rights Watch (2016). *Nepotism and neglect: The Failing Response to Arsenic in the Drinking Water of Bangladesh's Rural Poor*. Human Rights Watch.

ICDDRB and UNICEF (2018). A bacteriological water quality issue: Rohingya “Forcefully Displaced Myanmar Citizen” Camp in Cox’s Bazar, Bangladesh. Fact Sheet 8.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Jamil, N. B., Feng, H., Ahmed, K. M., Choudhury, I., Barnwal, P., and van Geen, A. (2019). Effectiveness of different approaches to arsenic mitigation over 18 years in Bangladesh: Implications for national policy. In progress.

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

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21 (2).

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique. *Annals of Applied Statistics*, 7 (1).

Madajewicz, M., Tompsett, A., and Habib, A. (2018). Community participation in decision-making: Evidence from a field experiment in Bangladesh. In preparation.

Ranacher, P., Brunauer, R., Trutschnig, W., der Spek, S. V., and Reich, S. (2016). Why GPS makes distances bigger than they are. *International Journal of Geographical Information Science*, 30 (2).

Smith, A., Lingas, E., and Rahman, M. (2000). Contamination of drinking water by arsenic in Bangladesh: A public health emergency. *Bulletin WorldHealth Organization*, 78:1093–1103.

United Nations (2016). Sustainable Development Goals. Goal 6: Facts and Figures. Report.

Waddington, H. and Snilstveit, B. (2009). Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *Journal of Development Effectiveness*, 1(3).

Wright, J., Gundry, S., and Conroy, R. (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).

Wu, J., van Geen, A., Ahmed, K. M., Alam, Y. A. J., Culligan, P. J., Escamilla, V., Feighery, J., Ferguson, A. S., Knappett, P., Mailloux, B. J., McKay, L. D., Serre, M. L., Streatfield, P. K., Yunus, M., and Emch, M. (2011). Increase in diarrheal disease associated with arsenic mitigation in Bangladesh. *PLOS One*, 6 (12)

# Online-appendixes

https://www.3ieimpact.org/sites/default/files/2019-10/Online-appendixes-DPW1.1006-Bangladesh-Safe-water.pdf