Serena Cocciolo
Selene Ghisolfi
Ahsan Habib
SMA Rashid
Anna Tompsett

Access to safe drinking water

Experimental evidence from new water sources in Bangladesh

March 2020

Impact Evaluation Report 109

Environment

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) promotes evidence-informed equitable, inclusive and sustainable development. We support the generation and effective use of high-quality evidence to inform decision-making and improve the lives of people living in poverty in low- and middle-income countries. We provide guidance and support to produce, synthesise and quality-assure evidence of what works, for whom, how, why and at what cost.

3ie impact evaluations

3ie-supported impact evaluations assess the difference a development intervention has made to social and economic outcomes. 3ie is committed to funding rigorous evaluations that include a theory-based design, use the most appropriate mix of methods to capture outcomes and are useful in complex development contexts.

About this report

3ie accepted the final version of the report *Access to safe drinking water: experimental evidence from new water sources in Bangladesh,* as partial fulfilment of requirements under grant DPW1.1006, awarded through Development Priorities Window 1. The content has been copy-edited and formatted for publication by 3ie.

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. The 3ie editorial production team for this report comprises Pallavi Duggal, Anushruti Ganguly and Akarsh Gupta, with Beryl Leach David de Ferranti providing overall editorial supervision.

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, 2020. *Access to safe drinking water: experimental evidence from new water sources in Bangladesh*, 3ie Impact Evaluation Report 109. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: https://doi.org/10.23846/DPW1IE109

Cover photo: Sonia Hoque / REACH

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

Ahsan Habib
NGO Forum for Public Health

SMA Rashid
NGO Forum for Public Health

Anna Tompsett
Stockholm University

Impact Evaluation Report 109

March 2020

International Initiative for Impact Evaluation

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. 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.

Summary

We evaluated the impact of a programme to provide safe sources of drinking water in rural Bangladesh. The programme consisted of a package of subsidies and technical advice to enable the installation of deep tubewells. Deep tubewells access water from deep aquifers that are free from faecal and arsenic contamination, a major naturally occurring problem in much of rural Bangladesh.

The programme installed wells that provide water that is almost free from arsenic contamination, but not from faecal contamination. The programme successfully installed a total of 107 tubewells in 129 communities throughout 2016 and 2017. In communities where the programme has successfully installed a new water source, the median household is 1.6 minutes' walk from the new source. The wells installed provide arsenic-safe drinking water, thereby essentially eliminating arsenic contamination at source for those who use them.

However, 34 per cent of installed tubewells unexpectedly tested positive for faecal contamination, compared with 46 per cent of other tubewells in the same communities, suggesting that the programme's wells only reduce exposure to faecal contamination by approximately 26 per cent. This result is unexpected because the source water is isolated from faecal contamination, meaning that exposure must occur either through the pump body or as a result of shallow groundwater entering the tubewell system.

The test for faecal contamination is coarse, however, meaning that we cannot evaluate whether levels are lower in the installed wells, but only whether faecal contamination is present or not. We therefore may have not fully captured the reduction in exposure to faecal contamination at source.

The programme reduced arsenic contamination in household drinking water, but not faecal contamination. Each tubewell installed under the programme led to a reduction in arsenic contamination of household drinking water that is equivalent to its elimination at the World Health Organization level for about five households. However, each of these tubewells also led to an increase in faecal contamination that is equivalent to introducing faecal contamination into the drinking water of about two households (although we cannot reject a small reduction or no effect on faecal contamination in household drinking water).

Modest improvements in source water quality, with respect to faecal contamination, are offset by an increase in travel time and possibly by changes in storage behaviour. The programme somewhat improved faecal contamination at the source level, but also slightly increased travel time and induced small changes in storage behaviour, both of which increase the risk of faecal contamination in drinking water.

Our best estimates suggest that walking an extra minute to collect drinking water increases the risk of faecal contamination by approximately 1.7 per cent, while storing drinking water in the house increases the risk of faecal contamination by approximately 7 per cent. The consequences of these negative effects are modest because few households walk more than a minute to collect drinking water, and the majority of households did 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 programmes may have a limited effect on faecal contamination. These results allay fears that deep tubewell construction programmes may substantially increase exposure to faecal contamination. However, they also suggest that construction of deep tubewells, in the absence of improvements to tubewell design or maintenance practices, is insufficient to resolve the faecal contamination problem in villages in rural Bangladesh. Further research is needed to quantify the extent of faecal contamination in deep tubewells and to understand the contamination channels.

Contents

List of figures and tables

Abbreviations and acronyms

IGC	International Growth Centre
IV	Instrumental variable
MICS	Multiple Indicator Cluster Survey
ppb	Parts per billion
SRC	Swedish Research Council
WHO	World Health Organization

1. Introduction

The United Nations Sustainable Development Goal 6 sets out the challenge of ensuring the 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 faecally contaminated water and 1,000 children a day died from diarrhoeal disease associated with poor water quality and sanitation (United Nations 2016).

In Bangladesh, which was 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 due to high levels of diarrhoeal disease, resulting from faecal contamination of surface water used for drinking (Caldwell et al. 2003). Educational campaigns encouraged people to shift to obtaining drinking water from groundwater sources, and were 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 contamination problem was discovered, an epidemic of diseases associated with arsenic exposure was already established, described as 'the largest poisoning of a population in history' (Smith et al. 2000).

Despite years of effort by the Bangladeshi government, NGOs and international aid agencies, progress on safe drinking water in Bangladesh remains elusive (Human Rights Watch 2016). Today, almost 100 million people still drink faecally contaminated water, and 39 million people drink water that is designated as contaminated with arsenic by international standards (BBS and UNICEF 2015).

The magnitude of the problem in 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 faecal contamination, there is far less consensus regarding 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 different channels of contamination empirically is difficult, because households that are located closer to safe water sources are also likely to differ from households that are further away in other respects, which may 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 raised 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 policymakers must design policy to reduce exposure to arsenic contamination without increasing exposure to faecal bacteria contamination. Our evaluation measures the impact of a programme that constructed new safe drinking water sources 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.

1.1 Overview

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

2. Intervention, theory of change and research hypotheses

2.1 The intervention

We evaluated the effects of a programme designed to improve access to safe drinking water in rural Bangladesh. The programme consisted of a package of subsidies and technical advice on building new sources of water, which were intended to provide drinking water that is free from both arsenic and bacterial contamination. The programme was fully implemented by a Bangladeshi NGO called NGO Forum for Public Health, which is also our 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 and arsenic contamination. In rural Bangladesh, deep tubewells are the most commonly proposed and implemented solution to the arsenic contamination problem. After installation, we tested all sources to confirm that the water was indeed arsenic free.[1] Table 1 describes the programme in a logical framework.

The subsidies ranged in value from 90 to 100 per cent of the cost of installing a new water source. Communities decided the location of new water sources by unanimous consensus in community meetings. We carried 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 offered to install one new water source in smaller treatment units, and two new water sources in larger treatment units.[2]

[1] We also tested to confirm that the sources were 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: (1) we assigned tubewells to villages as a function of village size, then divided these among the designated treatment units within each village; or (2) we assigned tubewells to treatment units to keep the ratio of households to tubewells as close as possible to 125:1.

Table 1: Logical framework

	Summary	Indicators	Means of verification	Assumptions
Impact	Improvement in drinking water quality	Arsenic contamination in household drinking water Faecal contamination in household drinking water	Water testing programme	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
Outputs	Construction of new drinking water sources	Number of sources constructed	Project records	
Activities	Safe drinking water programme (Section 2.1)			
Inputs	Subsidies, technical advice and support, community engagement/participation			

Installation costs were, on average, BDT60,000 per deep tubewell.[3] We assigned communities to one of three contribution requirements: cash contribution, labour contribution and contribution waiver. Communities assigned to the **cash contribution** treatment arm were required to raise BDT6,000 per installed water source, and the decision on how to divide this amount among households was delegated entirely to the community.

Communities assigned to the **labour contribution** arm were responsible for providing a total of 18 person-days of labour to assist the mason group in the installation work. Each person-day corresponded to a six-hour shift, consistent with local norms for unskilled labour, and was valued at BDT300,[4] or a total of BDT5,400, similar to the cash contribution requirement.[5] The implementation of the cash and labour contributions rule was designed to maximise comparability between the two treatments.[6] Communities assigned to the **contribution waiver** treatment arm received the programme without a required contribution.

A key feature of the programme delivery was the active involvement of targeted communities in the decision-making process regarding: (a) how many water sources to install in the community; (b) where to construct them; (c) how to divide the required contributions between households and (d) which households should take responsibility for

[3] Exchange rate approximately BDT80 to USD1
[4] The average daily unskilled wage in rural Bangladesh
[5] The contractor was paid BDT3,000 less per deep tubewell under the labour contribution requirement. The unskilled labour provided by the community did 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 compensated households contributing labour with BDT300 per person per shift.

the management and maintenance of each new water source. Communities took all decisions at meetings organised by the project staff, who played a strong facilitatory role. We imposed minimum participation requirements to hold a community meeting and required that all decisions were taken by unanimous consensus during the meeting in the presence of project staff. We did not implement the project in communities where an agreement was not found after a maximum of three meetings.[7]

The rules and procedures imposed on the decision-making process were 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 their voice, at least de jure.

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

2.2 Theory of change

Many programmes aim to improve access to safe drinking water by providing new and safe sources. A simple theory of change underlies these programmes: new sources are built; households adopt the new sources; source water quality improves; and thereby household water quality also improves. Figure 1 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 illustrates this more complete theory of change, which acknowledges that: (a) not all households will adopt a new source; (b) among those who do adopt a new source, many will also alter their transport and storage practices and (c) 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 that 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 in quality. These households will increase their transport distance and likely store water for longer.[10]

[7] In practice, only one community failed to reach an agreement. They declined to hold further meetings after a second meeting was unsuccessful in reaching agreement.
[8] For example, Wright and colleagues (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 diarrhoea are those that reduce bacterial contamination at the 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 per cent 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.

Providing new sources of safe drinking water may therefore have unintended consequences 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 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 theory of change, in the closest previous randomised evaluation Kremer and colleagues (2011) found that reducing source contamination by 66 per cent only reduced household water contamination by 24 per cent, interpreting these findings as evidence for recontamination via transport and storage. In the Bangladeshi context, Field and colleagues (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), as well as 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 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 favour use by elites, rather than the community as a whole, or because elites or landowners explicitly restrict use of the source.

Figure 1: Theory of change: simple model

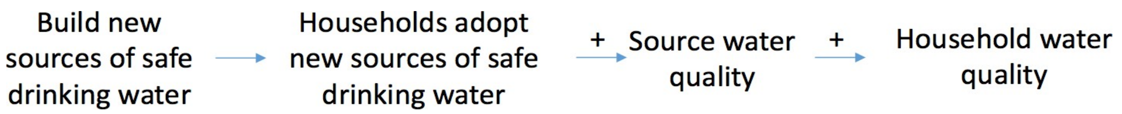
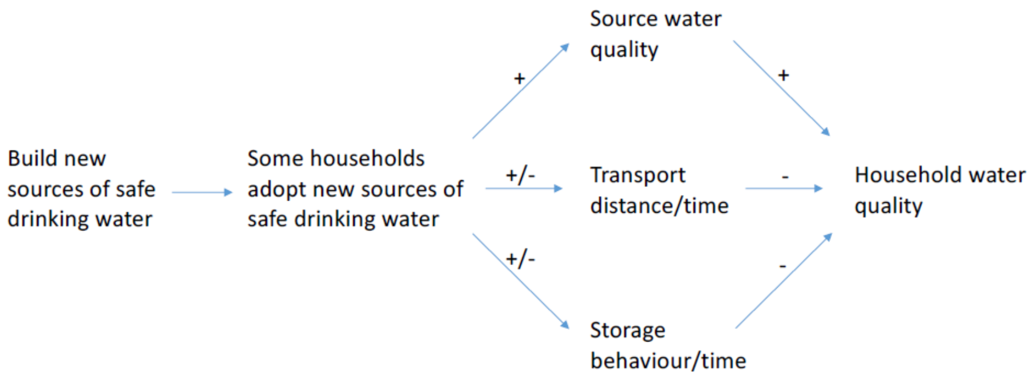


Figure 2: Theory of change: more complete model



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

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 on household contamination in this study for two reasons. First, our measures of drinking water quality have the benefit of being largely objective measures of project impact and are 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 lifetime exposure.

2.3 Key research questions

The research 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 programme on household water quality, measured by:
   (a) arsenic contamination in drinking water?
   (b) faecal contamination in drinking water?
2. How does the programme change behaviour with respect to obtaining water for drinking and cooking?
   (a) What is the average effect of the programme on the water quality of the source used by the household, measured by source arsenic contamination?
   (b) What is the average effect of the programme on water quality of the source used by the household, measured by source faecal contamination?
   (c) What is the average effect of the programme on distance walked to collect water?
   (d) What is the average effect of the programme on household water storage practices?
3. What is the causal effect of the behavioural 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.

3. Context

The context for this study is rural Bangladesh, where access to safe drinking water remains elusive (Human Rights Watch 2016), despite large existing 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’s rural population, consisting of more than 100 million individuals according to World Bank statistics,[12] now relies on drinking water obtained from approximately 10 million shallow, hand-pumped tubewells (Human Rights Watch 2016).

The use of drinking water from these shallow tubewells was originally extensively promoted via educational campaigns as a safe alternative to the use of surface water. The

[12] Available at: https://data.worldbank.org/indicator/SP.RUR.TOTL?locations=BD.

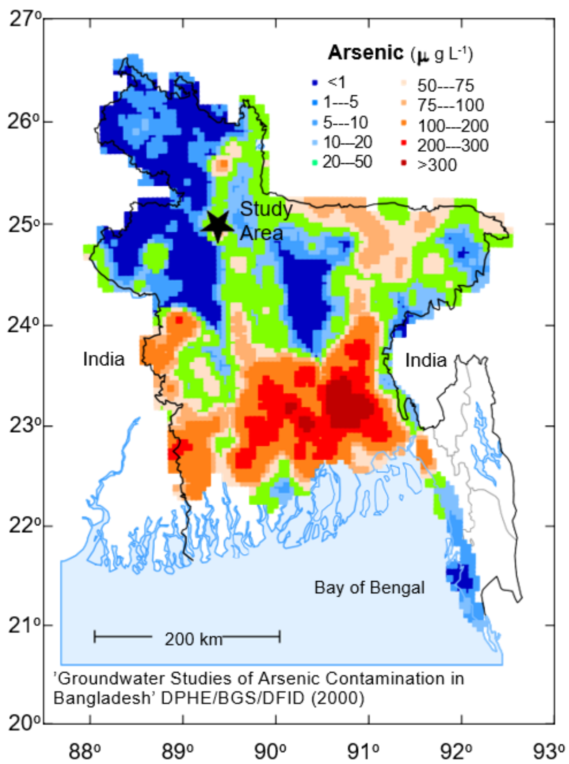
switch from surface water to tubewell water was followed by a sharp decline in child mortality (Caldwell et al. 2003), with its under-five mortality rate dropping from more than 200 per 1,000 live births in the 1960s and 1970s to roughly 30 per 1,000 today.

However, shallow groundwater in Bangladesh is contaminated with naturally occurring arsenic, a fact that was unknown when shallow tubewells were promoted as it is undetectable without water quality tests. Long-term exposure to arsenic eventually leads 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 World Health Organization (WHO) standard of 10 ppb – 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, described as ‘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, NGOs and international aid agencies, approximately 39 million people drink water that is defined as contaminated with arsenic by international standards (BBS and UNICEF 2015). Source locations may be chosen for political purposes, rather than targeted to those areas with the greatest need, and few sources are monitored after installation, meaning that some become unknowingly re-contaminated (Human Rights Watch 2016).

Figure 3: Arsenic contamination in Bangladesh and study site location



Source: This figure is reproduced from https://www.bgs.ac.uk/research/groundwater/health/arsenic/Bangladesh/mapsnhs.html.

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

Our intervention was located in north-west Bangladesh, in Shibganj *Upazila* (sub-district) and Sonatala *Upazila* in Bogra District, and Gobindaganj *Upazila* in Gaibandha District, as shown in Figure 3. The study area is not in the epicentre of the arsenic contamination problem. However, government officials and national media reported high levels of arsenic contamination in the specific study region (Akhtaruzzaman 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 had previously been installed.

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 preselect candidate communities. We then screened these candidate communities using water source testing. The final criteria for selection into the project was that either more than 25 per cent of community water sources were contaminated with arsenic, or more than 15 per cent were contaminated, and these sources were spatially clustered. We provide further details on recruitment to the study in Section 6.

The study population consisted 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 2[13] 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 presents baseline characteristics of the household sample with respect to their access to safe drinking water, and both Tables 2 and 3 also show comparable statistics from the national rural population, obtained from the Multiple Indicator Cluster Survey for 2012–2013 (BBS and UNICEF 2015). These comparisons allow us to evaluate how representative the study communities are of the rural population in Bangladesh, and thus the extent to which the results are likely to generalise.

Table 2 shows that the study population is largely representative of the national rural population, although households are somewhat smaller, more likely to be Muslim, and may be slightly poorer than the rural average (given that they are less likely to own a mobile phone, although they are slightly more likely to own a motorised vehicle).

[13] Appendixes E.1 and E.2 provide a detailed description of how each variable reported in Tables 2 and 3 is constructed.

Table 3 describes access to safe drinking water at baseline. We include measures of the household water quality (variables labelled as ‘HH test’) and the 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 ppb and the more conservative WHO threshold of 10 ppb. There is increasing evidence that the risks of exposure to between 10 and 50 ppb 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 approximately two 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.[14]

The rate of faecal contamination in water sources is higher than the national average, although the rate of faecal 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 recruited communities who face arsenic contamination problems.

Table 2: Socio-economic characteristics – descriptive statistics

	Study sample	**National population (rural)**
Household size	3.9 (0.022)	4.60 (0.015)
The household head is Muslim	0.94 (0.012)	0.87 (0.006)
The household head has no education	0.42 (0.009)	0.46 (0.004)
The household owns livestock	0.76 (0.009)	0.74 (0.004)
The household owns land for cultivation	0.53 (0.011)	0.48 (0.004)
Land owned by the household (acres)	1.00 (0.049)	1.20 (0.053)
The household has some toilet facility	0.84 (0.008)	0.94 (0.003)
Number of rooms to sleep	1.90 (0.016)	2.00 (0.009)
The floor is made of earth or sand	0.84 (0.008)	0.85 (0.004)
The roof is made of metal	0.96 (0.005)	0.92 (0.003)
Mobile phone ownership	0.60 (0.017)	0.83 (0.003)
Ownership of a motorised vehicle	0.065 (0.004)	0.051 (0.001)

Note: 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 the primary sampling unit level (‘treatment unit’ for the study sample and ‘cluster’ for the nationally representative sample).

[14] However, there were differences between how we measured this variable and how the Multiple Indicator Cluster Survey measured it, which may also account for the differences. We calculated the number of minutes it takes to walk to the primary water source from the respondent’s house. In MICS, on the contrary, the interviewer asked the question ‘How long does it take to go to the water source, get water, and come back?’. To improve comparability, we calculated *Total time* = *Walking time* $*$ 2 + *Queueing time* + 0.5, which is the value we report in the table. The distribution in MICS remains positively skewed compared with our data, partially accounting for the large difference in means.

Table 3: Water-related characteristics – descriptive statistics

	Study sample	National population (rural)
Arsenic contamination (WHO) (HH test)	0.63 (0.017)	0.61 (0.009)
Arsenic contamination (BD) (HH test)	0.24 (0.016)	0.17 (0.006)
Bacteria contamination (HH test)	0.65 (0.016)	0.63 (0.010)
Arsenic contamination (primary WS)	37 (2.248)	34 (1.586)
Arsenic contamination (WHO) (primary WS)	0.69 (0.018)	0.59 (0.011)
Arsenic contamination (BD) (primary WS)	0.31 (0.017)	0.19 (0.008)
Bacteria contamination (primary WS)	0.54 (0.010)	0.39 (0.011)
Storage dummy (observed)	0.73 (0.011)	0.19 (0.002)
Water is treated before drinking (primary WS)	0.087 (0.008)	0.035 (0.002)
Time needed to collect water (minutes)	2.2 (0.038)	15 (0.268)
Water collected per day (litres)	59 (1.199)	79 (1.160)

Note: HH = household; BD = Bangladesh; WS = water source. 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 the primary sampling unit level ('treatment unit' for the study sample and 'cluster' for the nationally representative sample).

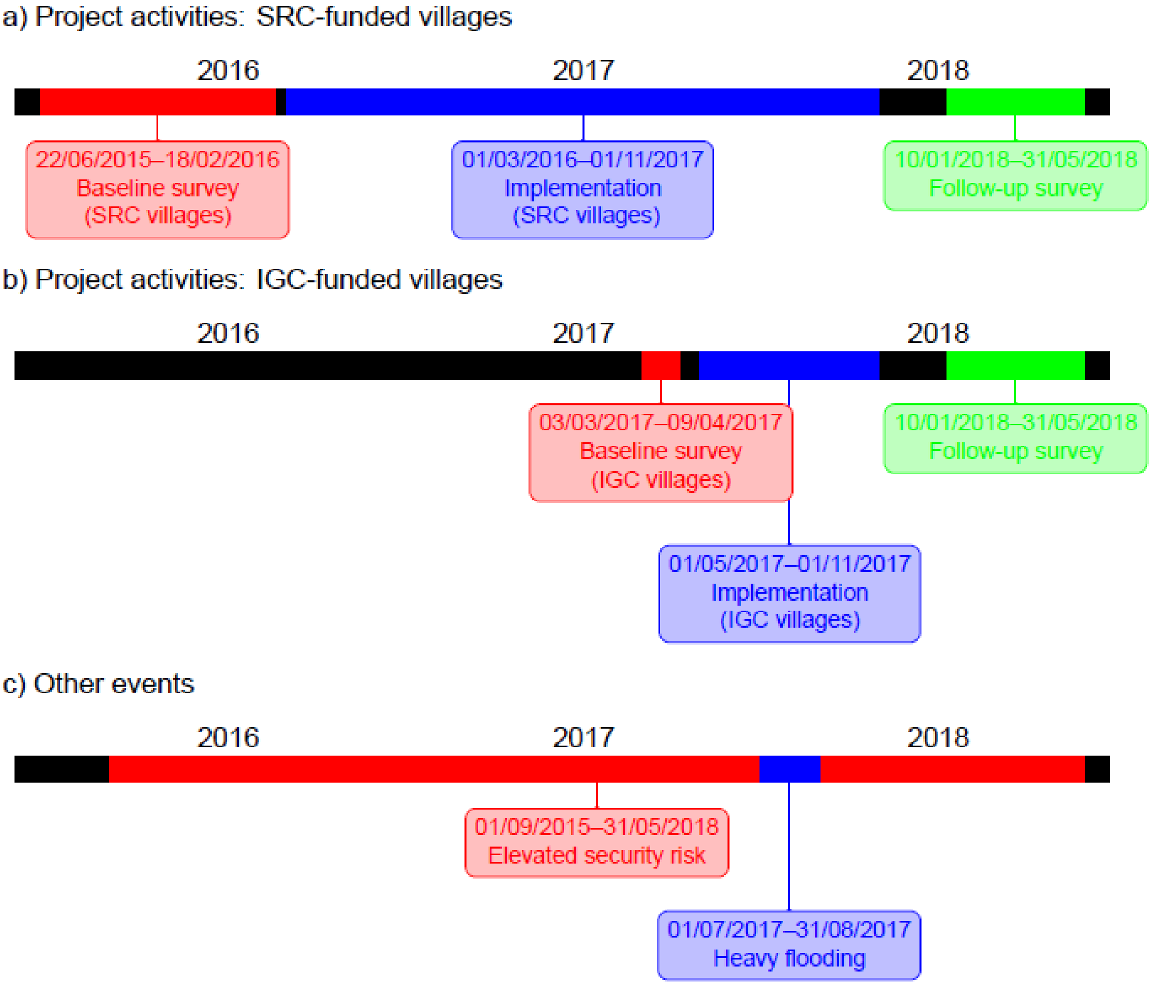
4. Timeline

Figure 4 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 International Growth Centre 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 that 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 around rural Bangladesh, as they 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 that in which implementation took place in specific unions. Finally, there was extreme flooding in the rainy season of 2017 immediately preceding our follow-up survey. It is possible that this flooding resulted in changes to water composition in tubewells, potentially affecting some of the patterns of contamination we observed at follow-up.

Figure 4: Evaluation timeline



Note: SRC = Swedish Research Council; IGC = International Growth Centre.

5. Evaluation: design, methods and implementation

In this section, we outline the study design including 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 an 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 about two thirds of study participants to have very limited literacy. All recruitment and consent procedures and study materials were translated into Bengali; informed oral consent was obtained in Bengali; and all survey data were 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 minimise the risk by asking permission and by asking where the preferred place would be for an interview, if one occurred.

We preserved the confidentiality of the information provided to us. Households were assigned identification numbers, which were used to store and to organise the data, rendering the data anonymous. We stored 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 these data to anyone other than co-investigators. The data are 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 the 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 programme 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 have easily outweighed the risks for those who participated in the survey.

Additionally, the programme required participation in a community meeting and agreement over where to locate any water sources installed. The community meetings were open to the entire community, not only to survey participants. No distinction was 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.

There was a possibility that the community decision-making process might exacerbate any pre-existing community tensions. However, the risks of participating in the intervention were no greater than those associated with participating in any community or NGO-led programme to improve access to safe drinking water, or more broadly, improve local public services.

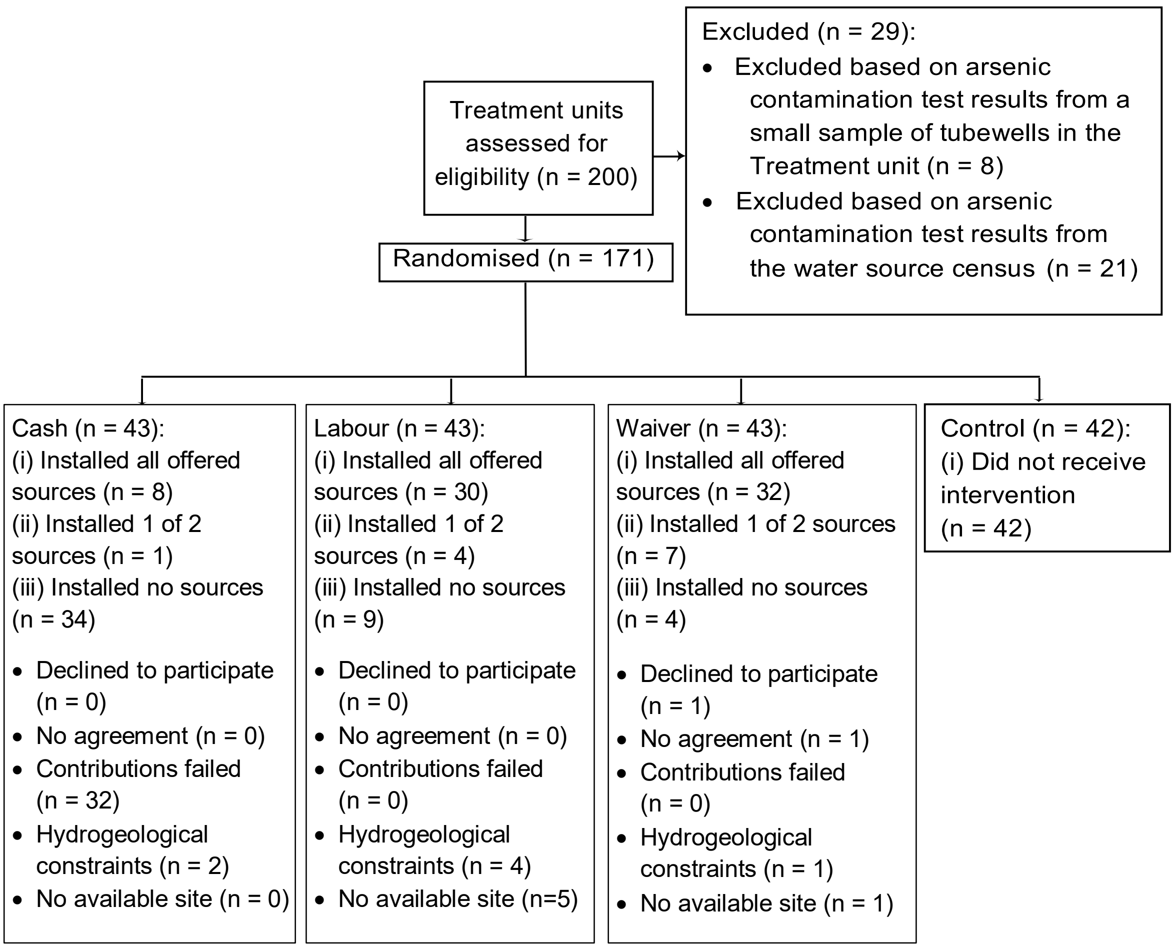
5.2 Evaluation design

The evaluation design is a randomised controlled trial, augmented by an analysis of mechanisms, which allows us to elucidate how the reduced-form effect of the programme arises. The randomised 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 programme.

Figure 5 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 programme under three different contribution requirements: one third of treated villages were required to raise a cash contribution before installation; one third of treated villages were required to contribute labour; and one third of treated villages received the programme 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: Evaluation design flow chart



Our primary interest in this study is the average effect across the three contribution requirements. However, take-up varies under the three contribution arms, 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 analyse mechanisms (key research question 3), we originally proposed two approaches. Our first analysis of mechanisms is a difference-in-differences approach, wherein we evaluate how changes in household bacterial contamination vary with changes in source contamination, transport distance and storage. The difference-in-differences 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-differences estimates are very stable across a range of specifications. However, although assignment to the safe drinking water programme was random, the selection of locations for water source installation was 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 (IV) analysis exploiting the experimental design of the safe drinking water programme. The IV 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 faecal 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-differences analysis. However, the cost, as we noted in our pre-analysis plan (Online appendix A), is substantially decreased precision. Our empirical results indicate that the IV estimates take the same sign as the difference-in-differences estimates, but the confidence intervals are extremely wide. Because it turns out that the IV analyses provide little additional information beyond the difference-in-differences analyses, we discuss the details of the IV method and the results only in Online appendix D.2.

5.3 Sample size

Our study is implemented in geographically defined treatment units comprising between 50 and 250 households. This 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 in order to exclude from the study those with less than 50 households, and to 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 predetermined 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.[15] The balance between treated and control units was intended to maximise 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 Online appendix B. We calculated minimum detectable effects at the 5 per cent level as 2.8 times the estimated standard deviation of coefficients. Table 4 summarises the results of our power calculations. Note that an average change of 2.2 metres in walking distance corresponds to 7 per cent of median distance to a water source at baseline. These minimum detectable effects compared favourably with expected treatment effects. We compare our results to these power calculations in Section 7.

Table 4: Summary of calculations

Research question	Outcome variable	Minimum detectable effect (5% level)
1a	Arsenic contamination in household drinking water	3.5%
1b	Faecal contamination in household drinking water	3.8%
2a	Arsenic contamination in source water	2.4%
2b	Faecal contamination in source water	2.8%
2c	Distance to water source	2.2 metres
2d	Reported storage	3%

In addition, we simulated the difference-in-differences analysis and the IV analysis. The estimated effect sizes for the difference-in-differences analyses also compared favourably with 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 per cent of simulations. However, we anticipated that the IV approach would sacrifice considerable power: we expected the IV approach to have minimum detectable effects approximately 10 times larger than the difference-in-differences 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-differences results for the analysis of mechanisms and discuss the IV results in detail only in Online appendix D.2.

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 wherein 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 their household list, and recorded

[15] Our original budget covered 155 treatment units, but an additional grant from the International Growth Centre allowed us to extend our sample size by a further 16 treatment units.

outcomes in the survey form, as they were required to fill in a form for all households 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 stated as: the household was not found (in 33% of cases); no one was at home during the visit from our enumerator (in 65% of cases); or the respondent refused to participate in the survey (in 2% of cases).

Enumerators conducted the interview with the household head, their spouse or another adult representative of the household. They always asked for informed consent, both for the interview and, separately, for the water testing. Overall, 99.8 per cent of households agreed to the interview and 99.6 per cent agreed to water testing. At baseline, we successfully conducted the household survey in a total of 6,529 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 household locations: in some cases, the administrative units had misassigned households to clusters. We reassigned the households so that each treatment unit retained geographical consistency.

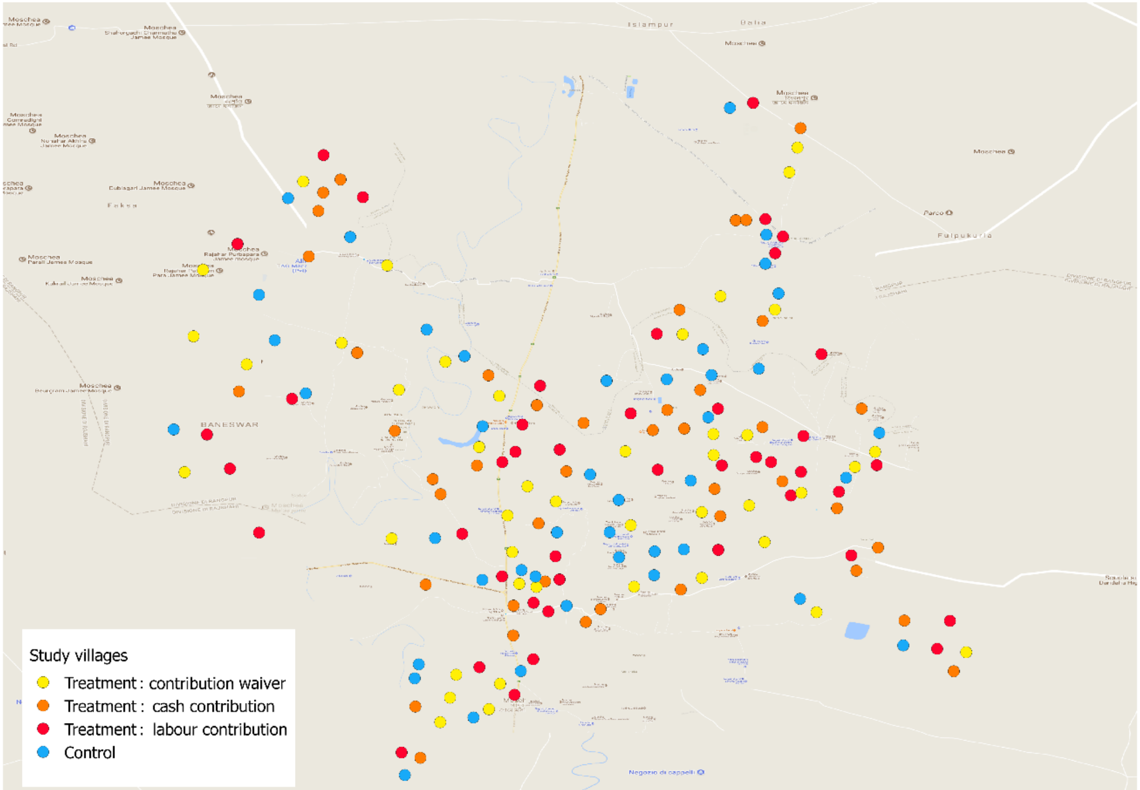
At follow-up, we were able to improve these statistics. The enumerators completed the interview with 99.85% of the households randomly selected to participate in the follow-up household survey. We were unable to complete the interview with four households that migrated, two households with no surviving household member and five 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% agreed to 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 selected 129 to receive the intervention. Of these 129 treatment units, 43 were randomly assigned to each of the three contribution requirements: (1) under the cash contribution requirement, communities are required to co-fund the installation costs; (2) under the labour contribution requirement, communities are required to provide labour to help with the installation work and (3) under the contribution waiver arm, the new water source is installed for free. We assigned 42 treatment units to the control group, which received no intervention.

We conducted the randomisation at public lottery meetings, to which we invited representatives from each eligible community. The randomisation was stratified by union *parishads* (councils) to make it feasible for representatives of the study communities to attend. The decision to use public randomisation was motivated by concerns about transparency, especially given that we offered the same programme 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 6 shows the resulting map of treatment units assigned to the control group and the three treatment arms.

Figure 6: Map of treatment and control areas



Source: Google Street Maps.

We also implemented the project sequentially by union *parishad*. As a result, households differed 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 programme. Additionally, the International Growth Centre 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.[16]

Large treatment units were offered two tubewells and smaller treatment units were offered one. This was determined using an algorithm to assign the number of tubewells as a function of the original village size or of the treatment unit size.[17]

5.6 Data collection

5.6.1 Survey design

We collected data through a combination of surveys and a water-quality testing programme. All data collection activities were carried out by a team of enumerators

[16] The International Growth Centre-funded villages are in different union *parishads* to the Swedish Research Council-funded villages.

[17] 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: (1) we assigned tubewells to villages as a function of village size, then divided these among the designated treatment units within each village; and (2) we assigned tubewells to treatment units to keep the ratio of households-to-tubewells as close as possible to 125:1.

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 Ahsan 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 matched households to the water sources they use. Our procedures for determining these matches are novel, because the problem of linking households to decentralised 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 used different approaches to match households to water sources at baseline and at follow-up.

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 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 a 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 they identified, to ensure that we correctly matched households to water sources. In case the respondent reported using a water source not included in the water source census data, we collected relevant information about this new source. This happened in only 2 per cent of the household surveys, indicating good coverage of the existing water sources from the census.

At follow-up, we did not repeat the water source census from baseline because of the cost of this exercise. Instead, we first conducted the household survey, and then collected data from all water sources the households described using. To avoid resurveying water sources multiple times, we tagged each with a zip tie. If an enumerator visited a source that had already been surveyed, they took a photograph and recorded GPS coordinates, enabling us to confirm its match with the water source data already collected by another enumerator.

5.6.2 Water quality tests

The water quality testing programme consisted of three types of tests: (1) a faecal contamination test; (2) a field arsenic test and (3) a laboratory arsenic test. For faecal 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 used are standard in the literature and have been used in previous studies of water quality.

Faecal 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.[18] The water testing procedure for bacteria contamination used hydrogen sulphide 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 faecal contamination.

The vials should be kept at room temperature for 48 hours, and the test is read as positive if the colour changes from clear to black. The hydrogen sulphide 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.[19]

Project staff entered the bacteria test results two days after the water sample collection, on average. However, in some cases, particularly at baseline, tests were left for more than two days; in some cases, results were entered after one day. To ensure that data are comparable across rounds, we applied a correction to the data, which accounts for any variation in how long each test was left before entering the data.

The correction we applied uses information on the specificity and sensitivity of the faecal contamination field test from a similar set of samples, also from Bangladesh, reported by Gupta and colleagues (2008).[20] We used the mean rate of positive tests, plus the sensitivity and specificity data, to back out the probability that positive and negative test results truly reflect faecal contamination. Intuitively, the correction implies that a sample that turns black in a very short time period has a near 100 per cent 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 households surveyed, provided that the survey respondent agreed to the testing procedure. This testing procedure was implemented in the field using the Hach EZ arsenic high range test kit, which provides results in 20 minutes and measures arsenic levels within the range of 0–500 ppb with the following increments: 0, 10, 25, 50, 250 and 500. Test results are immediately available, so we informed respondents about the results at the end of the survey.

---

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

[19] During the water source and household survey, we asked respondents to provide us with a phone number to be used for sending the results from the bacteria test by SMS. At baseline, 99% of 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 a phone number for 99% of households participating in the household survey.

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

Enumerators also gave a report card (in Bengali) to the owner or caretaker of the water source and to the households participating in the household survey, which reported 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 decreases the longer the water is stored. Arsenic begins to oxidise once the water is stored in a container that is open to the air, and the field test does not detect oxidised arsenic. We collected and tested water samples 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 – either from storage or directly from the source, using the same containers for transport that they normally use. This gave 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 procedure for a subset of households using a test conducted at the water quality testing laboratory of NGO Forum for Public Health, using an atomic absorption spectrophotometer. At baseline, we randomly selected 10 households for the arsenic laboratory test, 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 (by approximately 100 times) than the field tests. In total, we tested 897 water samples in the laboratory 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 of the two sets of tests are compared at baseline, they are highly correlated.[21]

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 Section 8.

5.7.1 Spillover effects
Our programme targeted 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 6). Moreover, because we divided large villages into several treatment units (Section 5.4), it was not uncommon that control and treated communities were adjacent to each other or very nearby.

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

Indeed, 120 out of 171 communities enrolled in the study were obtained by splitting a large village into two or more treatment units. The average distance between control communities and the closest treated community is 650 metres (about 8 minutes' walking time). As a result, the average distance between households in control communities and the closest project tubewell is 575 metres (7 minutes' walking time). For comparison, this distance is 175 metres (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 36 metres from their house or less than half a minute's 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 five minutes' walking distance, which is approximately 400 metres.[22] Moreover, households often stated during community meetings that they were unwilling to use a water source from a different cluster, even if the cluster was located within the same community.

5.7.2 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 behaviour with respect to obtaining drinking water – in other words, households under-report using unsafe water sources. We constructed our questions to reduce the effect of social desirability bias; namely, we initially simply asked respondents to list the sources they used, and asked them about the water sources they used before discussing knowledge about water safety.

5.7.3 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. washing 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 would be stronger in the treatment group, which has also participated in the safe drinking water programme, 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 bias in our comparison of the aggregate effects of the intervention on faecal contamination in household drinking water, and (possibly) bias in 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.

---

[22] At baseline, 94 per cent of households reported that they would switch to a new source if it was within a one-minute walk of their compound, but only 3 per cent report that they would switch if it was seven minutes' walk from their compound.

5.7.4 *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 programme to improve access to safe drinking water) exert more effort to reduce contamination through other channels, for example by improving household hygiene or 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 it would tend to attenuate differences between treatment and control households.

5.8 Quality checks

All survey instruments underwent a rigorous 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 Bengali, and enumerators were free to select the version in the language with which they felt more comfortable. The Bengali version was verified by back translation.

We collected survey data using tablet devices, based on 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 respondents' answers, and to prevent enumerators from accidentally skipping questions.

At follow-up, we incorporated project monitoring data[23] into our checks and constraints. For example, we added verification questions where responses diverged from project records[24] and automatically triggered 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 finalise 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, providing a more comprehensive assessment of each enumerator's work. This 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 owner of the water source, and the number of household members for which detailed demographic data were recorded in the household roster. We also created an incentive structure for the enumerators by paying a weekly salary bonus to the five best-performing enumerators.

---

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

[24] We did this to ensure that we did 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.

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 questions. We provided field supervisors with the weekly summary of the results and used these data to assess the accuracy of the information collected by 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 Ahsan Habib discussed these 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 recorded attendance at project meetings directly and also asked survey participants to verify whether or not they participated in meetings.

6. Programme design, methods and implementation

6.1 Key programme elements

The programme we evaluated was developed jointly by NGO Forum for Public Health and the research team, drawing on experience implementing similar projects. Table 5 summarises the key implementation activities that constitute the safe drinking water programme.

Table 5: Implementation activities

- Preparatory visits, information-gathering, community mobilisation
- Community decision-making
- Collection of cash contributions (if cash contribution arm)
- Installation of the pump body
- Water testing
- Construction of the platform
- Selection of caretakers
- Caretaker training
- Monitoring visits

Before organising community meetings, field staff visited the treatment unit; collected basic information on the geography of the village and the main socio-economic characteristics of the clusters grouped in the treatment unit; informed households about the scope of the intervention; organised information meetings within each cluster; and agreed on a date for the first community meeting. These preparatory activities were usually carried out over the space of a week and were crucial in order to guarantee that project staff were familiar with the specific circumstances within each treatment unit and could mobilise the community.

Field staff then organised 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 participates actively in the community meeting.[25]

Following these initial information meetings, the field staff then organised the main community meeting, in which communities took key decisions about whether to participate in the project and where to locate the water sources offered by the project. All households were invited to the meeting and encouraged to participate.[26]

As discussed in Section 2, decisions taken at the community meeting must have been agreed upon by consensus in the presence of project staff, and both women and poor households must have been represented at the meeting in which decisions were taken. The meetings were only carried out upon fulfilment of minimum participation requirements; field staff sometimes had to reschedule for another date if the minimum participation requirements were not fulfilled.

Community meetings were usually around one-hour long. The meetings began with a short introductory briefing by project staff on water safety issues and project implementation rules. The information provided on water safety issues primarily focused on source safety, explaining how arsenic and faecal contamination at source arises, which sources are at risk, which sources can provide safe water, and the health consequences of exposure to arsenic and faecal contamination. The main activity at the community meetings was a longer discussion session during which the communities took decisions, by consensus, on key aspects of the project. If decisions were not reached, we offered to organise another meeting, with a maximum of three meetings per treatment unit.[27] During the meetings, field staff displayed large-scale maps of the community showing all community water sources and their contamination status; these were developed using the baseline water-source census data.

Communities assigned to the cash contribution treatment arm were given a maximum of 12 weeks to raise the required amount,[28] during which time project staff visited the community several times in order to remind them of the deadline and establish progress. If assigned to the labour contribution treatment arm, communities had to sign a contract committing to provide the labour contribution and coordinate with project staff and contractors to agree on a time to provide the labour contribution.

In practice, the timing was mostly determined by contractor availability. Communities knew in advance approximately when the labour contribution would be required, but there was

---

[25] Although all households were invited to the information meeting, participation was voluntary. The field staff worked exhaustively to involve women in these activities, stressing the importance of their awareness and participation for the safety of the water consumed in the household.
[26] On average, 50% of households attended the community meeting, and 41% of participants were women. Poor and very poor households (by self-reported status) were less likely than middle-income households to attend the meetings, but only slightly: 44% of very poor households attended the meetings, compared with 53% of middle-income households. Households with high baseline arsenic contamination in household drinking water were also more likely to attend the meetings.
[27] In practice, few communities organised more than one meeting, and no communities organised more than two meetings.
[28] Field staff initially gave them a six-week deadline, which could be extended twice for an additional three weeks on each occasion.

some uncertainty until a few days before installation due to variation in how long the wells took to drill (a stage scheduled immediately beforehand).

Installation of the wells involved the use of local technologies, primarily manpower, to manually turn a drill bit approximately 60 millimetres in diameter. The technology can penetrate layers of weak or fractured rock, but not solid rock. Project staff, including the field engineer, supervised the installation in order to guarantee that the tubewell depth was adequate to reach an arsenic-free aquifer.

The goal was to reach a safe layer, meaning one that is permeable (through which water can flow relatively freely), but separated from the arsenic-contaminated layers at the surface by an impermeable layer (through which the contaminated water cannot pass). If such a layer was reached, the drill was lifted and withdrawn, and a PVC pipe was inserted into the hole.

Pumps maintained pressure in the excavation to reduce the likelihood of its collapse. If the underlying geology was very sandy, there was also a risk that the excavation could collapse and the PVC pipe could not be inserted. Communities assigned to the labour contribution arm were 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 conducted laboratory water tests for arsenic, iron and manganese. If these test results were satisfactory, we finalised the 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 pump body and the platform, project staff organised 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 were trained by our field engineer in maintaining the water source and keeping the site clean. Project staff conducted 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 (within 6 weeks, 8 weeks and 12 weeks after the construction of the tubewell pump body and platform).

The installation procedure, from the first preparatory visits to the community to the completion of the construction of the pump body, took on average two months. In most treatment units it was completed within four months. We conducted laboratory water tests for arsenic, iron and manganese contamination, and installation was finalised with the construction of the platform; this occurred, on average, two months after the installation of the pump body, and in the majority of cases within three months.

6.2 Monitoring

The implementation roll-out of the intervention was closely monitored via a systematic and comprehensive process of data collection on most project activities. Most of the information used for implementation monitoring was 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 the 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 accidentally being skipped. These elements minimised the risk that the intervention was not carried out according to the treatment assignments or that information was misrecorded.

We complemented the electronic data collection system with a range of additional project documentations: a record of all staff visits carried out in each treatment unit (activity report); an extensive qualitative narrative by project staff of all implementation stages per treatment unit (project staff report); a record of attendance data and participation in decision-making per community meeting, or caretaker selection meeting, using predefined household lists (attendance sheet); and other office records, including those related to installation processes, key dates for the implementation of the intervention, and caretaker training.

Additionally, we required project staff to record audio of all information meetings and community meetings organised, which we have transcribed, translated and coded. This comprehensive monitoring plan resulted in the list of indicators to monitor the implementation of the intervention, which is described in detail in Online 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.[29] There was limited scope for implementers to innovate, although the process of facilitating the community meetings, in particular, required some learning: the only community that failed to reach an agreement was the very first meeting our team organised. 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 programme 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 with 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 preselect villages and then refined the selection using water source testing. We preselected the 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: (1) data from the Bangladesh Arsenic Mitigation Water Supply Project, which included a large screening

[29] 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 each other 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.

programme for tubewells, conducted between 1999 and 2006; (2) an assessment from the Department of Public Health Engineering on the most arsenic-contaminated villages in Bogra District and (3) data collected in 2008 from the Bangladesh Social Development Services.

We preselected as candidate villages for receiving our intervention all villages indicated by the Department of Public Health Engineering, or for which data from the Arsenic Mitigation Water Supply Project or Social Development Services reported a share of arsenic-contaminated tubewells equal to or higher than 30 per cent. 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 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 into 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 finalise the selection of treatment units eligible for receiving the arsenic mitigation programme. In particular, we excluded from the study all treatment units with less than 15 per cent of arsenic-contaminated water sources.

We further screened treatment units with less than 25 per cent of arsenic-contaminated water sources, including them in the programme only if they presented a well-defined cluster of contaminated water sources.[30] 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 implemented the same recruitment policy until we achieved our target recruitment levels. The final study population consisted of 171 treatment units, all of which had arsenic contamination levels greater than 25 per cent 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 regarding the fairness of the approach for selecting treated villages. Treatment units assigned to the control group understandably expressed disappointment. In addition, a few communities expressed disappointment with their 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 programme. The first step in the programme was organising 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 organise a meeting. In the meetings, communities were asked to take a decision about where the proposed well(s) should be located. Only one community failed to reach consensus during their meeting.

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

6.4 Targeting

We successfully recruited communities to the programme with significant arsenic contamination issues, and communities randomly assigned to receive the safe drinking water programme do not differ significantly from those assigned to the control group (Section 7). In the 129 communities in which we implemented the programme, we offered to construct a total of 179 tubewells. As discussed in Section 2, we either offered one or two tubewells to each community, depending on the treatment unit size. Table 6 summarises the result of each attempted installation.

Of the 179 tubewells we offered, we successfully installed 107. One community declined to hold a meeting and another could not agree on a location. For 13 of the communities that were offered wells, no suitable land could be identified. For 44 of those 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, in 13 sites, the communities successfully completed all stages of the project and we attempted installation. However, we could not complete it because of hydrogeological conditions, namely: 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: 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 7 shows how communities who successfully completed the programme (resulting in attempted installation) differed from communities who did not (resulting in no attempted installation), as well as how communities in which we successfully installed wells differed from those in which we did not. Where we did not successfully install wells, this implies that we either did not attempt installation, because the community did not successfully complete the programme, or that we attempted installation and failed.

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

Table 7: Selection into successful installation

	Attempted installations	Successful installations	Obs
	(1)	(2)	(3)
Arsenic contamination (WHO) (HH test)	0.06	0.02	4,917
	(0.04)	(0.04)	
Arsenic contamination (BD) (HH test)	0.09***	0.06*	4,917
	(0.03)	(0.03)	
Bacteria contamination (HH test)	0.00	0.00	4,899
	(0.03)	(0.03)	
Household size	-0.08	-0.10*	4,918
	(0.05)	(0.05)	
Poverty score – USD2	-0.00	-0.00	4,889
	(0.01)	(0.01)	
Not-educated household members (%)	-0.00	-0.00	4,918
	(0.02)	(0.02)	
Literacy rate in the household	0.01	0.00	4,911
	(0.02)	(0.02)	
Network nominations	-0.06	0.02	4,918
	(0.08)	(0.07)	
Network size	-0.06	0.03	4,918
	(0.08)	(0.08)	
Muslim household	0.03	0.03	4,913
	(0.03)	(0.02)	
High trust towards community	0.03	0.02	4,914
	(0.02)	(0.03)	
Know association	0.01	0.02	4,863
	(0.02)	(0.02)	

Note: HH = household; BD = Bangladesh; Obs = observed. Column 1 summarises differences in listed characteristics between: (1) households living in communities which completed all stages of the programme and in which we attempted installation; and (2) 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 summarises differences between: (1) communities in which we successfully installed at least one water source; and (2) communities in which we did not install any water sources. Results are obtained 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 centred controls for union-level stratification. Standard errors clustered by treatment unit. *** $p < 0.01$; * $p < 0.10$.

6.5 Evidence on implementation procedure

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

Among households that knew about the programme, 87 per cent of households in treated communities and 83 per cent of households in control communities knew that communities were selected to receive the programme by lottery. A very similar percentages of households[31] knew that all selected communities were assigned to receive the programme under different terms and implementation rules, and that assignment to treatment was done by lottery.[32]

Among households in treated communities that knew that their community received the water safety programme implemented by NGO Forum for Public Health, 96% remembered that some meetings were held in their community in relation to the programme[33] and 77% reported that at least one household member attended the community meeting. Our programme records suggest that approximately 50% of households participated in the meeting. Of these, 95% correctly remembered the number of offered tubewells[34] and 98% correctly remembered the contribution requirement (cash, labour or waiver).

Table 8 summarises water quality statistics in the tubewells installed by the project, compared with 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 programme successfully reduce arsenic contamination to minimal levels, although 6 per cent of project-installed wells tested positive for arsenic at the WHO threshold, and 1 per cent did so at the higher Bangladeshi threshold. The arsenic field test we used 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 faecal contamination than non-project wells in the same communities, a considerable proportion of these water sources (34%) still tested positive for faecal contamination. This rate of contamination is unexpected, because the source of water that 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 in the rainy season before our follow-up survey. Another recent study (ICDDRB and UNICEF 2018) also finds substantial levels of faecal 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.

The communities in which we implemented the programme are small and relatively compact. As a result, half of the households in communities where we successfully installed 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.

[31] 87 per cent (treated communities) and 82 per cent (control communities)
[32] 85 per cent (treated communities) and 81 per cent (control communities)
[33] This rises to 97 per cent after prompting.
[34] This rises to 99 per cent after prompting.

Table 8: Comparison of project tubewells with other tubewells

	Faecal contamination (predicted)	Arsenic contamination (WHO)	Arsenic contamination (BD)
	(1)	(2)	(3)
Project tubewells	-0.13***	-0.58***	-0.35***
	(0.05)	(0.04)	(0.03)
Mean (project tubewells)	0.34	0.06	0.01
Mean (other tubewells)	0.46	0.63	0.34
N	3,394	3,510	3,510

Note: BD = Bangladesh. The table reports the regression estimated difference in contamination in project tubewells compared with 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 reports the mean contamination levels in project tubewells and non-project tubewells in the same communities. The sample consists of water sources that 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 programme with similar cash contribution requirements and successfully installed about 83 per cent of the tubewells that we offered (Madajewicz et al. 2018). We were therefore surprised by the negative response to a cash contribution requirement 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 have been 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. Qualitative evidence from that study also suggests that the payment was associated with a perceived right to control use of the source. In our current 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 was 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 programme 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 water source quality with respect to faecal 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 required contributions under the cash contribution 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 programme is sensitive to both the context and the programme 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 faecal contamination, but they do not eliminate faecal contamination at source. However, we do not have measures of intensity of faecal 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 faecal 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.

7. Impact analysis and results of the key research questions

7.1 Methodology

7.1.1 Programme 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 programme to treatment units:

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

where Δy_{ic} is the change in outcome variable y between baseline and follow-up in household i in community c,[35] T_c is an indicator which takes the value 1 if community c is assigned to treatment, and η_d is a union *parishad* fixed effect. The estimated effects are the average intent-to-treat effects of the programme – regardless of whether the programme successfully installs water sources or not – so they are not contaminated by selection into successful installation.

[35] We departed from the pre-specified approach in one minor respect, and analysed data at the household level, applying weights so that each treatment unit counted 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.

The variables η_d are controls that reflect stratification in the original randomisation. We include controls for each lottery in which treatment was assigned.[36] Following Lin (2013), Imbens and Rubin (2015) and Gibbons and colleagues (2018), we demean the lottery fixed effects and include the interaction term between the lottery controls and the treatment dummies, meaning that β_T estimates the average difference between treated and control villages.

We report the pre-specified analyses for all pre-specified variables of interest, as summarised in Table 9. 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.[37]

In our original power calculations, we modelled take-up of the overall intervention at 70 per cent 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.

Table 9: Variables of interest

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

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

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

Additional analyses (not pre-specified)
We report one additional set of results that 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 programme effects pool results across all three treatment arms. However, the effects are very small under the cash contribution arm, attenuating the overall results.[38] We therefore also estimated the impact of the programme using a second approach that scales the effect by the size of the investment.

To implement this second approach, we used 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 programme 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, normalised 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 that successfully installed water sources.

For brevity, we largely report these secondary estimates only for the main outcome variables, as listed in Table 9. 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 Mechanisms

To analyse 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-differences strategy and an IV approach.

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

$$\begin{aligned} FC^h_{if} - FC^h_{ib} &= b_0 + b_1(FC^w_{if} - FC^w_{jb}) + b_2(DIST^w_{if} - DIST^w_{ib}) \\ &\quad + b_3(STORAGE_{if} - STORAGE_{ib}) + \eta_c + \epsilon_i \end{aligned} \qquad (2)$$

where all variables are measured at baseline *b* and follow-up *f*; FC^h is faecal contamination in household *i*'s drinking water and FC^w is faecal 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).

η_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 2 with and without these community-level dummy variables, as there was no clear ex-ante

[38] We report effects separately for each contribution arm in Online 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.

reason to prefer one approach over the other.[39] When we include the community-level dummy variables, Equation 2 only exploits within-community variation in changes in the right-hand side variables to estimate causal effects.

The difference-in-differences yields causal estimates under the assumption that changes in the right-hand side variables are uncorrelated with other changes in household drinking water contamination – for example, through changes in household hygiene practices. Such an assumption may be reasonable, since the programme did not provide extensive or differential information on other types of health or hygiene behaviour.

Additionally, as we show, the difference-in-differences 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 programme 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-differences 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 actual variables we constructed for the IV analysis were only weak predictors of behaviour change. The main reasons for the unexpectedly poor performance of the instruments concerns the unexpectedly small improvement in faecal contamination in source drinking water (Section 6.6) and unexpected measurement issues with the GPS data (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 Online appendix D.2.

7.2 Sample for analysis

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

7.2.1 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. Of these 6,484 households, 6,431 gave consent to household water

[39] Either approach might increase precision, depending on the exact structure of ε_i.

testing at both baseline and follow-up.[40] Of these, we have household drinking-water-quality data (arsenic and faecal contamination) at both baseline and follow-up 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: (1) we could not locate a matching record in the water-source survey data; or (2) we could not uniquely match the faecal contamination test identifier with a result in our test result database.[41] The final panel sample consisted of 6,051 households interviewed at both baseline and follow-up for which we have household and source water quality data from both rounds.

We focus on this sample to avoid changing samples between the main analyses. However, in some analyses we have fewer observations. This is 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).

7.2.2 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 1.24 and 1.12 respectively, and this difference is statistically significant. At baseline, households obtained 99 per cent of their water from the primary source, in both treated and control communities. At follow-up, this share decreases to 93 and 97 per cent in treated and control communities respectively, 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 that 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 it.

7.3 Balance

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

7.3.1 Balance between treatment and control

Table 10 shows that the treated and control groups are similar in terms of 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 communities with control communities, only two individual tests reject the null hypothesis that the means in the two groups are equal at the 10 per cent level.

Neither joint test rejects this null hypothesis. We also compare each contribution treatment arm to the control group separately (resulting in 36 individual tests and 6 joint tests). Of the

[40] A total of 6,481 gave consent to household water testing at baseline and 6,434 gave consent at follow-up.
[41] We used locally produced barcodes, which occasionally contained duplicate IDs.

individual tests, 4 out of 36 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 11 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 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 10: Socio-economic characteristics – balance check

	Control	Treated	Cash	Labour	Waiver
Household size	3.8	3.9	3.9	3.9	3.8
	(0.081)	(0.077)	(0.089)	(0.080)	(0.088)
The household head is Muslim	0.99	0.98	0.98	0.97	0.99
	(0.022)	(0.012)	(0.023)	(0.023)	(0.025)
The household head has no education	0.46	0.47	0.45	0.47	0.49
	(0.034)	(0.031)	(0.036)	(0.038)	(0.034)
The household owns livestock	0.78	0.81	0.82*	0.79	0.81
	(0.033)	(0.023)	(0.026)	(0.025)	(0.027)
The household owns land for cultivation	0.58	0.59	0.58	0.61	0.57
	(0.035)	(0.027)	(0.030)	(0.029)	(0.032)
Land owned by the household (acres)	1.1	1	0.98	0.95	1.1
	(0.118)	(0.068)	(0.091)	(0.093)	(0.118)
The household has some toilet facility	0.85	0.85	0.83	0.86	0.85
	(0.025)	(0.020)	(0.023)	(0.025)	(0.023)
Number of rooms to sleep in	1.9	1.9	1.9	1.9	1.8*
	(0.049)	(0.041)	(0.046)	(0.049)	(0.043)
The floor is made of earth or sand	0.86	0.86	0.87	0.86	0.85
	(0.035)	(0.031)	(0.031)	(0.036)	(0.033)
The roof is made of metal	0.95	0.97	0.97	0.98**	0.96
	(0.017)	(0.011)	(0.013)	(0.012)	(0.013)
Mobile phone ownership	0.65	0.7*	0.69	0.72*	0.7
	(0.053)	(0.040)	(0.043)	(0.046)	(0.044)
Ownership of a motorised vehicle	0.042	0.055*	0.057	0.063**	0.043
	(0.011)	(0.010)	(0.011)	(0.012)	(0.011)
p-value of F-test for joint significance		0.449	0.316	0.005	0.501
p-value of Hotelling's T-squared test		0.309	0.559	0.099	0.805

Note: 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 11: Water-related characteristics – balance check

	Control	Treated	Cash	Labour	Waiver
Arsenic contamination (WHO) (HH test)	0.53	0.6**	0.61*	0.61*	0.58
	(0.064)	(0.063)	(0.070)	(0.069)	(0.070)
Arsenic contamination (BD) (HH test)	0.15	0.22**	0.21	0.21	0.25**
	(0.046)	(0.044)	(0.048)	(0.046)	(0.053)
Faecal contamination (HH test) (predicted)	0.59	0.56	0.56	0.57	0.56
	(0.026)	(0.024)	(0.028)	(0.027)	(0.027)
WS arsenic contamination (WHO)	0.62	0.71**	0.71**	0.72**	0.69
	(0.070)	(0.070)	(0.078)	(0.076)	(0.074)
WS arsenic contamination (BD)	0.23	0.32***	0.32**	0.3	0.35***
	(0.050)	(0.049)	(0.054)	(0.054)	(0.059)
WS faecal contamination (predicted)	0.56	0.55	0.54	0.55	0.55
	(0.022)	(0.016)	(0.020)	(0.021)	(0.020)
The water is treated to make it safe for drinking	0.16	0.17	0.16	0.2	0.15
	(0.039)	(0.043)	(0.038)	(0.048)	(0.039)
Storage dummy (observed)	0.64	0.65	0.65	0.67	0.62
	(0.031)	(0.024)	(0.026)	(0.028)	(0.028)
Time needed to collect water (minutes)	2.1	2.1	2.1	2.1	2.1
	(0.094)	(0.061)	(0.077)	(0.074)	(0.092)
Water collected per day (litres)	53	53	53	51	55
	(4.406)	(3.773)	(4.350)	(4.415)	(4.153)
WTP for a new WS in a socially optimal location	88	98	103	103	85
	(12.311)	(12.073)	(13.384)	(14.206)	(12.035)
p-value of F-test for joint significance		0.153	0.063	0.109	0.007
p-value of Hotelling's T-squared test		0.612	0.536	0.442	0.234

Note: HH = household; BD = Bangladesh; WS = water source; WTP = willingness to pay. 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 Programme effects

Pre-specified analyses

Tables 12, 13 and 14 present the mean programme effects on key research questions 1, 2a and 2b, and 2c and 2d, 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 programme on household water quality? Table 12 reports mean effects of the programme on household water quality. Column 1 shows the main results for arsenic contamination in household water. The average programme 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 show 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 rise on average, albeit insignificantly (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 contamination above 250 ppb and 0.3 % have 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 illustrates that the effects are primarily concentrated in a larger proportion of households that experience relatively small reductions in arsenic concentration, and a smaller proportion of households that experience relatively small increases. There is no effect in the sparsely populated tails of the distribution, which receive the greatest weight in the analysis that uses the test results.

Column 4 in Table 12 shows the main average effects on faecal contamination. Ex ante, it was ambiguous whether the programme would increase or decrease exposure to faecal contamination in drinking water. The results in Column 4 suggest essentially no effect on faecal contamination in household drinking water: the point estimate is a 0.2 percentage point increase in contamination, but the 95% confidence interval spans both modest increases (4.1 percentage points) and modest decreases (3.7 percentage points) in contamination.

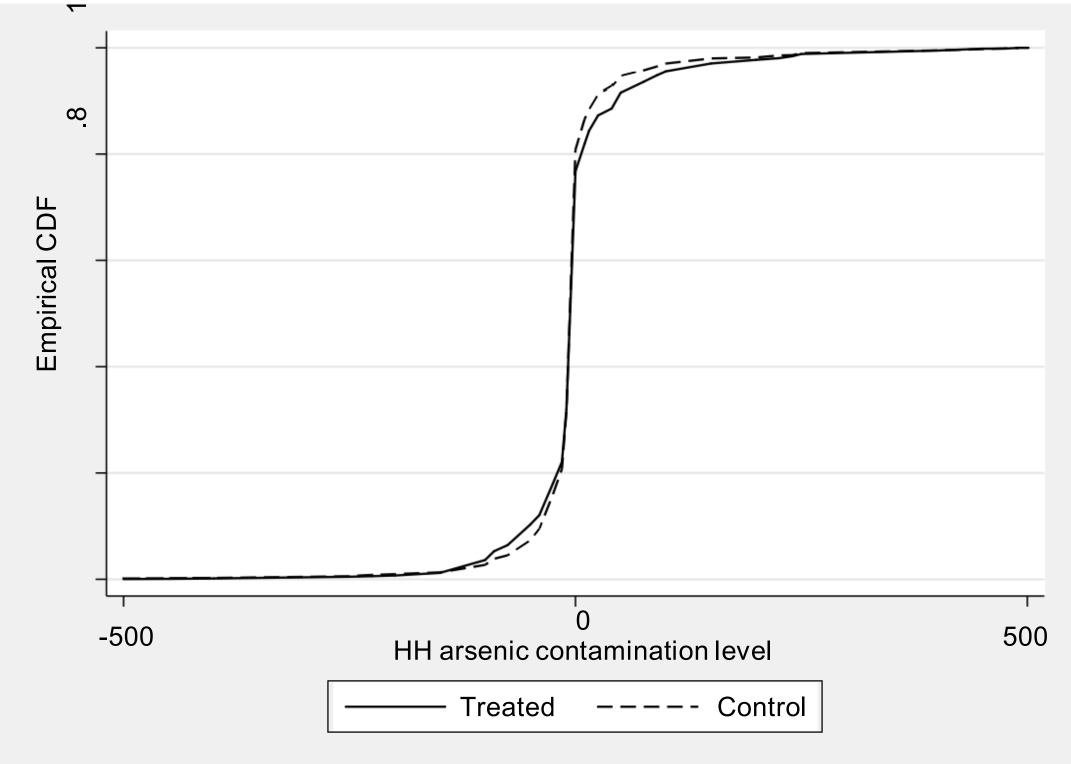
Note that the control group experiences improvements in household water quality, particularly with respect to arsenic contamination. There are a number of potential explanations for these changes, including: changes in behaviour as a result of the information we provided about water source quality at 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 12: Effect of the programme on household water quality

	Arsenic contamination (WHO)	Arsenic contamination (BD)	Arsenic contamination level	Faecal 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	6,051	6,051	6,051	6,048

Note: BD = Bangladesh; Obs = observed; R2 = round two. Table shows estimated average programme 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. * p < 0.1, ** p < 0.05, *** p < 0.01.

Figure 7: Cumulative Distribution Function[42] of arsenic field test result in household sample



Note: HH = household; CDF = cumulative distribution function.

Key research question 2: How did the programme change behaviour with respect to obtaining water for drinking and cooking? Table 13 shows the estimated effect on source water quality. Column 1 shows that the average programme impact is a 5.6 percentage point reduction in the volume of water obtained from sources above the WHO

[42] Cumulative distribution function is correct.

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 8 visualises these results.

Column 4 in Table 12 shows a modest decrease in the share of water obtained from sources with faecal contamination of 1.5 percentage points. However, the confidence interval does not exclude zero. The effect on source faecal contamination is less than 30 per cent of the effect on arsenic contamination. This probably reflects the fact that faecal contamination in project sources was only 28 per cent (13 percentage points) lower than other sources in the same communities. In contrast, the reduction in arsenic contamination in water sources was 92 per cent (58 percentage points).[43]

As in Table 13, Table 14 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 as for the adoption of new sources. However, if we analyse the change in contamination rates in sources for which we have readings at both baseline and follow-up, we see that sources experienced, on average, a 7 percentage point fall in arsenic contamination rates at the WHO standard and 1 percentage point fall at the Bangladeshi standard.[44]

These results suggest that the changes in contamination seen in the control group are likely to be 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.

Table 13: Effect of the programme on source water quality

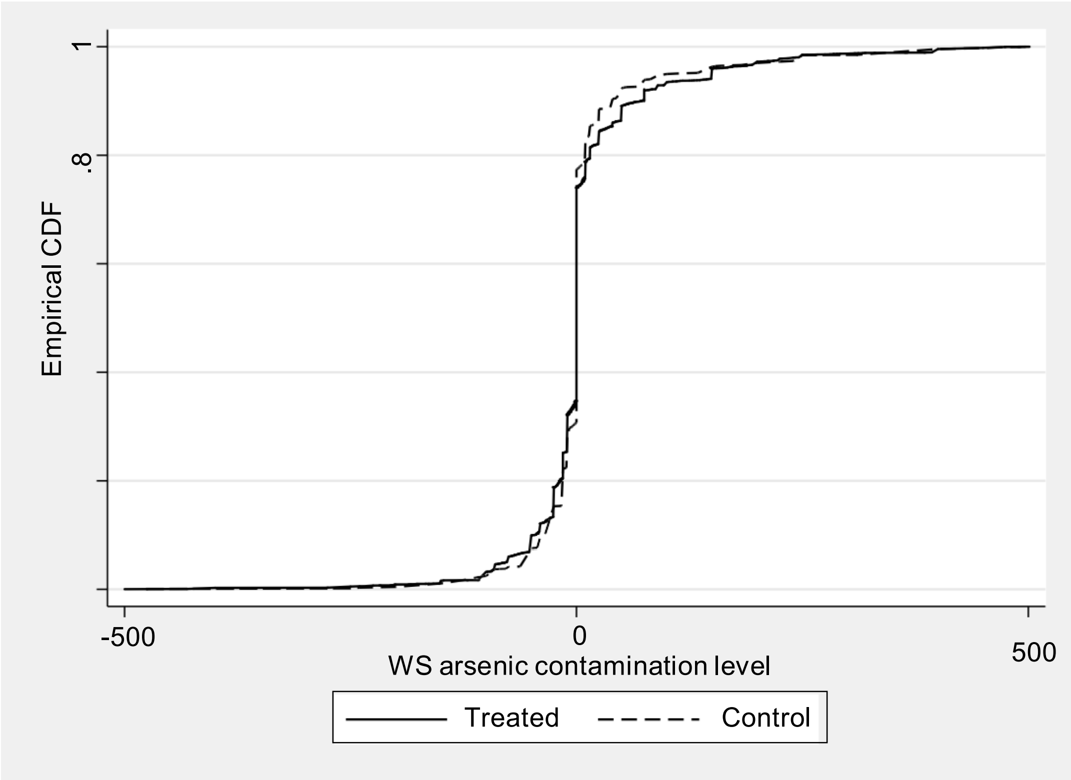
	Arsenic contamination (WHO)	Arsenic contamination (BD)	Arsenic contamination level	Faecal contamination
	(1)	(2)	(3)	(4)
Treated	-0.056**	-0.027	-0.155	-0.015
	(0.023)	(0.019)	(3.648)	(0.019)
Constant	-0.068***	-0.008	7.642**	-0.001
	(0.018)	(0.015)	(3.167)	(0.015)
R2	0.03	0.02	0.07	0.04
Obs	6,051	6,051	6,051	5,993

Note: BD = Bangladesh; R2 = round two; Obs = observed. Table shows estimated average programme 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[43] See Table 8.

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

Figure 8: Cumulative Distribution Function of arsenic field test result in household sample



Note: WS = water source; CDF = cumulative distribution function.

Table 14 also shows the effects of the programme 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 0.1 metres in treated communities relative to control communities, while Column 2, which uses reported data, shows an increase of 0.066 minutes, equivalent to about 5 metres in walking distance.

Table 14: Effect of the programme 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	5,832	5,729	6,050	6,051

Note: HH-WS = household water source; min = minutes; R2 = round two; Obs = observed. Table shows estimated average programme 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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 fact that the two points we measure distance between may reverse in orientation – as well as the fact 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 measured the location of our installed water sources with greater accuracy than other sources, because we drew on multiple measures of location and we verified the locations by inspection. As a result, if a household adopted our water sources, we overestimated the distance between the household and our water source to a lesser extent than we overestimated the distance between that household and the water source they used at baseline, and consequently, we underestimated 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 in the distance travelled, on average, between baseline and follow-up for all study households, while Column 2 does not find any reported change in distance travelled. This is because the tablet devices we used at follow-up recorded locations with more precision than the tablet devices we used at baseline. We therefore overestimated the distances between households and sources that were only measured at baseline relatively more than we overestimated the distances between households and sources that were only measured at follow-up.[45]

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 15 provides additional measures of water storage practice. Note that not all of 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 per cent level. These results suggest that the programme 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 15: Effect of the programme on storage practices

	Containers are uncovered	**Containers are on the floor**	**Containers are uncovered (observed)**	**Containers are on the floor (observed)**	**Water is scooped (observed)**
	(1)	(2)	(3)	(4)	(5)
Treated	0.023	0.008	0.040*	0.025	-0.002
	(0.039)	(0.030)	(0.021)	(0.019)	(0.022)
Constant	-0.351***	-0.092***	0.383***	0.484***	0.439***
	(0.033)	(0.027)	(0.018)	(0.016)	(0.020)
Only follow-up data			✓	✓	✓
R2	0.04	0.02	0.02	0.01	0.01
Obs	6,051	6,051	5,902	5,997	6,029

Note: R2 = round two; Obs = observed. Table shows estimated average programme 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

IV estimates of impact of wells (not pre-specified)

The mean programme estimates reported so far are intent-to-treat estimates, in that they gauge the average effect of the programme, regardless of whether any wells were 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 treated households, under the assumption that programme effects only operate via the provision of new wells.

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

The effects reported in Table 16 imply that installing one well in a community of 100 households would decrease household arsenic contamination by about 4.6 percentage points but would increase household faecal contamination by about 1.6 percentage points. However, in both cases, we also cannot reject either null effects or small effects in the opposite direction.

Table 16: Effect of water sources installed per household on household water quality – IV estimates

	Arsenic contamination (WHO)	Faecal 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	6,051	6,048

Note: R2 = round two; Obs = observed. 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 17 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 on 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 contamination of household water; however, the effect might not be sufficiently large to bring household arsenic contamination below the threshold at which we measure it. The estimated effect on faecal 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 them to other estimates available in the literature.

IV analyses of the transport variables, in this case, both suggest positive effects on distance to safe drinking water. However, 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 IV 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 17: Effect of water sources installed per household on water source quality – IV estimates

	Arsenic contamination (WHO)	Faecal 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	6,051	5,993

Note: R2 = round two; Obs = observed. 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. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 18: Effect of water sources installed per household on transport and storage practice – IV 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	5,832	5,729	6,050	6,051

Note: HH-WS = household water source; min = minutes; R2 = round two; Obs = observed. 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. * p < 0.1, ** p < 0.05, *** p < 0.01.

7.4.2 Mechanism

Table 19 shows the results of the difference-in-differences analysis using the measure of reported travel time to water sources. Table 20 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 metres per minute.[46]

[46] 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.

In Columns 2 and 4, in both Table 19 and 20, we include treatment unit fixed effects, meaning that we exploit only variation in outcomes within treatment units. 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 faecal 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 6 to 8 percentage points, on average. Again, this effect is extremely stable across specifications.

In contrast, the results in Tables 19 and 20 yield quite different conclusions regarding the effects of distance. The results in Columns 1 and 2 of Table 20 suggest that increasing travel time by one minute increases the risk of contamination by at most 0.5 percentage points. However, 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 19 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 in the table below, leads to fractionally larger estimated effects of transport time on drinking water contamination. This is because increasing travel times are associated (weakly)[47] 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 IV 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-differences analyses. We therefore discuss the results of these analyses only in Online appendix D.2.

[47] Results available on request.

Table 19: Mechanism – analysis using reported travel time

	Drinking water faecal contamination			
	(1)	(2)	(3)	(4)
Source faecal contamination	0.238***	0.219***	0.241***	0.220***
	(0.015)	(0.015)	(0.015)	(0.015)
Travel time HH-WS (minutes, reported)	0.015*	0.018**	0.016**	0.018**
	(0.008)	(0.008)	(0.008)	(0.008)
Observed storage	0.080***	0.069***		
	(0.010)	(0.009)		
Constant	0.001	-0.001	-0.010	-0.011***
	(0.010)	(0.001)	(0.010)	(0.001)
Treatment unit fixed effects	No	Yes	No	Yes
R2	0.07	0.14	0.06	0.13
Obs	5,673	5,673	5,674	5,674

Note: HH-WS = household water source; R2 = round two; Obs = observed. 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. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 20: Mechanism – analysis using travel time measured using GPS coordinates

	Drinking water faecal contamination			
	(1)	(2)	(3)	(4)
Source faecal contamination	0.238***	0.219***	0.242***	0.221***
	(0.016)	(0.016)	(0.016)	(0.016)
Travel time HH-WS (minutes, measured)	-0.001	0.005	0.003	0.009
	(0.015)	(0.013)	(0.015)	(0.013)
Observed storage	0.073***	0.063***		
	(0.010)	(0.009)		
Constant	0.001	-0.000	-0.008	-0.008***
	(0.010)	(0.001)	(0.010)	(0.000)
Treatment-unit fixed effects	No	Yes	No	Yes
R2	0.07	0.14	0.06	0.13
Obs	5,774	5,774	5,775	5,775

Note: HH-WS = household water source; R2 = round two; Obs = observed. 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. * p < 0.1, ** p < 0.05, *** p < 0.01.

7.5 Heterogeneous impacts

In this section, we summarise the results of prespecified heterogeneity analyses. We report the details of these analyses in Online appendix D.3.

7.5.1 By use of safe or 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 faecal 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 reductions in household arsenic and faecal contamination are greater among those closer to constructed sources in successful treatment units. We observe similar patterns for changes in source water arsenic contamination, although not for source faecal 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 programme 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 their absence. 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 faecal 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 experienced substantial reductions in source faecal contamination, while the poor experience substantial increases. The differences are sufficiently large that they would probably 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 programme. Each source we install eliminates arsenic contamination for 5–10 households, containing on average 3.9 individuals, so 20–40 individuals in total. The average cost of well installation is BDT60,000 or approximately USD720 at current exchange rates. Including *only* the installation costs, the cost of avoiding arsenic contamination is between BDT6,000 and BDT12,000 per household (between USD72 and USD144).

In per capita terms, these ranges are BDT1,540 to BDT3,080 or USD18.50 to USD37. 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 programme may be higher, because we used a baseline water testing programme to target communities with arsenic contamination. Communities also used this information to select locations for installation. Lacking this information, the programme 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 a large scale.

For comparison, Jamil and colleagues (2019) estimate the cost per person with reduced exposure via deep tubewell installation to be between USD9 (under the best possible siting conditions and assuming 60 per cent uptake within a 100-metre radius, higher than the uptake we see in this context) and USD142 (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.

8. Discussion

8.1 Internal validity

8.1.1 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, our view is that when interpreting the results, it is more conservative to place more weight on the results using reported distance than those using calculated 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, 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.

8.1.2 Spillover

Our programme targets communities that are highly arsenic-contaminated in 10 union *parishads*. Villages enrolled in our project lie in relatively small geographical areas (Figure 6). Moreover, because we divide large villages in several treatment units (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 spillover* 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.[48] Given the local context, the absence of spillover was largely expected. As discussed in Section 5.7, households have a strong preference for local water sources. Indeed, only 0.8 per cent of households interviewed at follow-up reported using a water source in a different cluster than their own.

8.1.3 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 control than by systematic compensatory source switching in response to information about arsenic contamination.

8.1.4 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, differences between household and source contamination are 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 programme alters reporting of behaviour, as well as the behaviour itself – an important question for future evaluation programmes.

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 faecal contamination between treated and control villages.[49] These results provide reassurance that our findings are unlikely to be influenced by patterns of differential reporting bias.

---

[48] 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.

[49] Detailed results available on request.

8.2 External validity

While our specific findings are most applicable to the context of rural Bangladesh, our findings are potentially generalisable to other settings. Like Kremer and colleagues (2011), we find that source water quality only partially explains household water quality. In our case, our results suggest that eliminating source faecal contamination would only reduce household contamination by at most 25 per cent. These results confirm that faecal 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 faecal contamination, although we cannot determine whether there are improvements on the intensive margin (i.e. lower concentrations of faecal bacteria) due to the limitations of the faecal contamination test we used, for budgetary reasons, in this study.

We find limited support for the hypothesis that households walking increasing distances, or storing water for longer, as a result of switching to more distant arsenic-safe sources can explain the results in Field and colleagues (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 faecal contamination and arsenic contamination in shallow tubewells. The results of Field and colleagues (2011) could therefore still be explained by switching to sources with higher faecal 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, between changes in source arsenic contamination and changes in source faecal 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 programme 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 programme is probably smaller than it would have been had we implemented it under one of the other two contribution requirements in all communities. These findings are important, however, because programme take-up is a key determinant of the impact of safe drinking water programmes, and our findings confirm that the impacts of these programmes vary by key aspects of their design.

8.4 Key lessons for researchers

A key finding from this study is the difficulty of using GPS coordinates measured with mobile phones or tablet devices to calculate small distances, especially when the extent of measurement error varies between different objects. To our knowledge, the extent of measurement error in GPS coordinates calculated with tablet devices 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 tablet devices in measuring GPS is likely to change rapidly with time.

9. Findings for policy and practice

9.1 Policy

The key findings for policymakers 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, programme 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 USD18.50 per capita, so the cost of well installation for the rural population affected by arsenic contamination would be more than USD700 million. In comparison, installing local piped water supply systems could cost USD150 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 USD1 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 programmes alone have little impact on faecal contamination**
  Deep tubewells reduce, but do not eliminate source faecal 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 durations increase the risk of faecal contamination in household drinking water. Both effects are small, in part because households rarely walk for more than four to five minutes, at most, 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 faecal contamination must therefore adopt 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 with regard to community contribution requirements. In other work, we previously showed that approaches to decision-making are also important determinants of impact in programmes 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 programme design in order to maximise the impact of safe drinking water programmes.

Online appendixes

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

References

Ahuja, A, Kremer, M and Zwane, AP, 2010. Providing safe water: evidence from randomized evaluations. *Annual Review of Resource Economics*, 2, pp.237–256.

Bangladesh Bureau of Statistics (BBS) and UNICEF, 2015. *Bangladesh Multiple Indicator Cluster Survey 2012–2013*. Dhaka: BBS and UNICEF.

British Geological Society (BGS) and Department of Public Health Engineering (DPHE), 2001. Arsenic contamination of groundwater in Bangladesh. Kinniburgh, DG and Smedley, PL (eds). *British geological survey technical report WC/00/19*. British Geological Survey: Keyworth.

Caldwell, BK, Caldwell, JC, Mitra, SN and Smith, W, 2003. Tubewells and arsenic in Bangladesh: challenging a public health success story. *International Journal of Population Geography*, 9, pp. 23–38.

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.

Akhtaruzzaman, 2014. 'Tube-well water arsenicated in Bogra: 161 villages in eight upazilas affected.' *Daily Observer,* [online] (Last updated at 12.00 a.m. on 18 December 2014) Available at: https://www.observerbd.com/2014/12/18/61337.php.

Fewtrell, L, Kaufmann, RB, Kay, D, Enanoria, W, Haller, L and Colford Jr, JM, 2005. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, 5, pp.42–52.

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*. Cambridge, MA: Department of Economics, Harvard University.

Food and Agriculture Organization of the United Nations (FAO), 2016. *AQUASTAT main database*. Available at: <http://www.fao.org/land-water/databases-and-software/aquastat/en/> [Accessed 30 December 2019].

Gibbons, CE, Serrato, JCS and Urbancic, MB, 2018. Broken or fixed effects? *Journal of Econometric Methods*, 8(1), pp.1–12.

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(2), pp.388–395.

Human Rights Watch, 2016. Nepotism and Neglect: The Failing Response to Arsenic in the Drinking Water of Bangladesh's Rural Poor. New York City: Human Rights Watch.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB) and UNICEF, 2018. A bacteriological water quality issue: Rohingya 'forcefully displaced Myanmar citizen' camp in Cox's Bazar, Bangladesh. Fact sheet 8.

Imbens, GW and Rubin, DB, 2015. *Causal inference for statistics, social, and biomedical sciences: an Introduction*. Cambridge, UK: Cambridge University Press.

Jamil, NB, Feng, H, Ahmed, KM, 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. *Environmental Science & Technology 53*(10), pp.5596–5604

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

Levitt, SD and List, JA, 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), pp.153–174.

Lin, W, 2013. Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique. *Annals of Applied Statistics*, 7(1), pp.295–318.

Madajewicz, M, Tompsett, A and Habib, A, n.d. *Community participation in decision-making: evidence from a field experiment in Bangladesh.*

Ranacher, P, Brunauer, R, Trutschnig, W, der Spek, SV and Reich, S, 2016. Why GPS makes distances bigger than they are. *International Journal of Geographical Information Science*, 30(2), pp.316–333.

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

United Nations, 2016. *Sustainable Development Goals. Goal 6: Facts and Figures*. Available at: <https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-6-clean-water-and-sanitation/targets.html> [Accessed 30 December 2019].

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

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), pp.106–117.

Wu, J, van Geen, A, Ahmed, KM, Alam, YAJ, Culligan, PJ, Escamilla, V, Feighery, J, Ferguson, AS, Knappett, P, Mailloux, BJ, McKay, LD, Serre, ML, Streatfield, PK, Yunus, M and Emch, M, 2011. Increase in diarrheal disease associated with arsenic mitigation in Bangladesh. *PLOS One*, 6(12), pp.e29593.

Other publications in the 3ie Impact Evaluation Report Series

The following reports are available from http://3ieimpact.org/evidence-hub/publications/impact-evaluations

Impact of alternate wetting and drying on farm incomes and water savings in Bangladesh, 3ie Impact Evaluation Report 108. Chakravorty, U, Dar, MH, Emerick, K, 2020.

The effects of vouchers for essential household items on child health, mental health, resilience and social cohesion among internally displaced persons in the Democratic Republic of Congo, 3ie Impact Evaluation Report 107. Quattrochi, J, Bisimwa, G, Thompson, T, van der Windt, P and Voors, M, 2020.

Measuring impacts of conservation interventions on human well-being and the environment in Northern Cambodia, 3ie Impact Evaluation Report 106. Clements, T, Neang, M, Milner-Gulland, EJ and Travers, H, 2020.

The 5 Star Toilet Campaign: improving toilet use in rural Gujarat, 3ie Impact Evaluation Report 105. Chauhan, K, Schmidt, WP, Aunger, R, Gopalan, B, Saxena, D, Yashobant, S, Patwardhan, V, Bhavsar, P, Mavalankar, D and Curtis, V, 2020.

How education about maternal health risk can change the gender gap in the demand for family planning in Zambia, 3ie Impact Evaluation Report 104. Ashraf, N, Field, E, Voena, A and Ziparo, R, 2019.

In search of the holy grail: can unconditional cash transfers graduate households out of poverty in Zambia?, Impact Evaluation Report 103. Handa, S, Tembo, G, Natali, L, Angeles, G and Spektor, G, 2019.

Increasing HIV self-testing and linkage to care for partners of women in antenatal care in Uganda, Impact Evaluation Report 102. Wanyenze, R, Buregyeya, E, Matovu, J, Kisa, R, Kagaayi, J, Vrana-Diaz, C, Malek, A, Musoke, W, Chemusto, H, Mukama, S and Korte, J, 2019.

Improving the quality of care for children with acute malnutrition in Uganda, 3ie Impact Evaluation Report 101. Marzia, L, Wanzira, H, Lochoro, P and Putoto, G, 2019.

Impacts of increasing community resilience through humanitarian aid in Pakistan, 3ie Impact Evaluation Report 100. Avdeenko, A and Frölich, M, 2019.

Impacts of community monitoring of socio-environmental liabilities in the Ecuadorian and Peruvian Amazon, 3ie Impact Evaluation Report 99. Pellegrini, L, 2019.

Increasing HIV testing demand among Kenyan truck drivers and female sex workers, 3ie Impact Evaluation Report 98. Kelvin, E, George, G, Mwai, E, Kinyanjui, S, Inoti, S, Chetty, T, Strauss, M, Romo, M, Oruko, F, Odhiambo J, Nyaga, E, Mantell, J and Govender, K, 2019.

Impacts of community stakeholder engagement interventions in Ugandan oil extractives, 3ie Impact Evaluation Report 97. Parker, R, Coleman, E, Manyindo, J, Schultz, B and Mukuru, E, 2019.

The impacts of formal registration of businesses in Malawi, 3ie Impact Evaluation Report 96. Campos, F, Goldstein, M and McKenzie, D, 2019.

Unpacking the determinants of entrepreneurship development and economic empowerment for women in Kenya, 3ie Impact Evaluation Report 95. McKenzie, D, Puerto, S and Odhiambo, F, 2019.

Impacts of key provisions in Ghana's Petroleum Revenue Management Act, 3ie Impact Evaluation Report 94. Edjekumhene, I, Voors, M, Lujala, P, Brunnschweiler, C, Owusu, CK and Nyamekye, A, 2019.

Using information to break the political resource curse in natural gas management in Mozambique, 3ie Impact Evaluation Report 93. Armand, A, Costa, AI, Coutts, A, Vicente, P and Vilela, I, 2019.

Harnessing transparency initiatives to improve India's environmental clearance process for the mineral mining sector, 3ie Impact Evaluation Report 92. Pande, R and Sudarshan, A, 2019.

Impacts of removing user fees for maternal health services on universal health coverage in Kenya, 3ie Impact Evaluation Report 91. Abuya, T, Dennis, M, Matanda, D, Obare, F and Bellows, B, 2018.

Impact of voice reminders to reinforce harvest aggregation services training for farmers in Mali, 3ie Impact Evaluation Report 90. Osei, RD, Dzanku, FM, Osei-Akoto, I, Asante, F, Hodey, LS, Adu, PN, Adu-Ababio, K and Coulibaly, M, 2018.

Impacts of Breakthrough's school-based gender attitude change programme in Haryana, India, 3ie Impact Evaluation Report 89. Jayachandran, S, Jain, T and Dhar, D, 2018.

Hotspot interventions at scale: the effects of policing and city services on crime in Bogotá, Colombia, 3ie Impact Evaluation Report 88. Blattman, C, Green, D, Ortega, D and Tobón, S, 2018.

Impact evaluation of the Philippine Special Program for Employment of Students, 3ie Impact Evaluation Report 87. Beam, E, Linden, L, Quimbo, S and Richmond, H, 2018.

Community-based distribution of oral HIV self-testing kits: experimental evidence from Zambia, 3ie Impact Evaluation Report 86. Hensen, B, Ayles, H, Mulubwa, C, Floyd, S, Schaap, A, Chiti, B, Phiri, M, Mwenge, L, Simwinga, M, Fidler S, Hayes, R, Bond, V and Mwinga, A, 2018.

Evaluating the economic impacts of rural banking: experimental evidence from southern India, 3ie Impact Evaluation Report 85. Field, E and Pande, R, 2018.

Direct provision versus facility collection of HIV tests: impacts of self-testing among female sex workers in Uganda. 3ie Impact Evaluation Report 84. Ortblad, K, Musoke, DK, Ngabirano, T, Oldenburg, C and Bärnighausen, T, 2018.

Increasing female sex worker HIV testing: effects of peer educators and HIV self-tests in Zambia, 3ie Impact Evaluation Report 83. Chanda, MM, Ortblad, KF, Mwale, M, Chongo, S, Kanchele, C, Kamungoma, N, Fullem, A, Bärnighausen, T and Oldenburg, CE, 2018.

Community delivery of antiretroviral drugs: a non-inferiority matched-pair pragmatic cluster-randomized trial in Dar es Salaam, Tanzania, 3ie Impact Evaluation Report 82. Francis, JM, Geldsetzer, P, Asmus, G, Ulenga, N, Ambikapathi, R, Sando, D, Fawzi, W and Bärnighausen, T, 2018.

Nourishing the future: targeting infants and their caregivers to reduce undernutrition in rural China, 3ie Impact Evaluation Report 81. Cai, J, Luo, R, Li, H, Lien, J, Medina, A, Zhou, H and Zhang, L, 2018.

Impacts of the World Food Programme's interventions to treat malnutrition in Niger. 3ie Impact Evaluation Report 80. Brück, T, Ferguson, NTN, Ouédraogo, J and Ziegelhöfer, Z, 2018.

Impact evaluation of the World Food Programme's moderate acute malnutrition treatment and prevention programmes in Sudan. 3ie Impact Evaluation Report 79. Guevarra, E, Mandalazi, E, Balegamire, S, Albrektsen, K, Sadler, K, Abdelsalam, K, Urrea, G and Alawad, S, 2018.

Impact evaluation of WFP's programs targeting moderate acute malnutrition in humanitarian situations in Chad. 3ie Impact Evaluation Report 78. Saboya, M, Rudiger, J, Frize, J, Ruegenberg, D, Rodríguez Seco, A and McMillon, C, 2018.

Improving midday meal delivery and encouraging micronutrient fortification among children in India, 3ie Impact Evaluation Report 77. Shastry, GK, Berry, J, Mukherjee, P, Mehta, S and Ruebeck, H, 2018.

Evaluation of infant development centres: an early years intervention in Colombia, 3ie Impact Evaluation Report 76. Andrew, A, Attanasio, O, Bernal, R, Cordona, L, Krutikova, S, Heredia, DM, Medina, C, Peña, X, Rubio-Codina, M and Vera-Hernandez, M, 2018.

Can the wounds of war be healed? Experimental evidence on reconciliation in Sierra Leone. 3ie Impact Evaluation Report 75. Cilliers, J, Dube, O and Siddiqi, B, 2018.

Impact evaluation of the Menabe and Melaky development programme in Madagascar, 3ie Impact Evaluation Report 74. Ring, H, Morey, M, Kavanagh, E, Kamto, K, McCarthy, N, Brubaker, J and Rakotondrafara, C, 2018.

Impact evaluation of the Smallholder Dairy Commercialization Programme in Kenya, 3ie Impact Evaluation Report 73. Bonilla, J, McCarthy, N, Mugatha, S, Rai, N, Coombes, A and Brubaker, J, 2018.

Impact and adoption of risk-reducing drought-tolerant rice in India, 3ie Impact Evaluation Report 72. Yamano, T, Dar, MH, Panda, A, Gupta, I, Malabayabas, ML and Kelly, E, 2018.

Poverty and empowerment impacts of the Bihar Rural Livelihoods Project in India, 3ie Impact Evaluation Report 71. Hoffmann, V, Rao, V, Datta, U, Sanyal, P, Surendra, V and Majumdar, S 2018.

How should Tanzania use its natural gas? Citizens' views from a nationwide Deliberative Poll, 3ie Impact Evaluation Report 70. Birdsall, N, Fishkin, J, Haqqi, F, Kinyondo, A, Moyo, M, Richmond, J and Sandefur, J, 2018.

Impact evaluation of the conditional cash transfer program for secondary school attendance in Macedonia, 3ie Impact Evaluation Report 69. Armand, A and Carneiro, P, 2018.

Age at marriage, women's education, and mother and child outcomes in Bangladesh, 3ie Impact Evaluation Report 68. Field, E, Glennerster, R, Nazneen, S, Pimkina, S, Sen, I and Buchmann, N, 2018.

Evaluating agricultural information dissemination in western Kenya, 3ie Impact Evaluation Report 67. Fabregas, R, Kremer, M, Robinson, J and Schilbach, F, 2017.

General equilibrium impact assessment of the Productive Safety Net Program in Ethiopia, 3ie Impact Evaluation Report 66. Filipski, M, Taylor, JE, Abegaz, GA, Ferede, T, Taffesse, AS and Diao, X, 2017.

Impact of the Uddeepan programme on child health and nutrition in India, 3ie Impact Evaluation Report 65. Kochar, A, Sharma, A and Sharma, A, 2017.

Evaluating oral HIV self-testing to increase HIV testing uptake among truck drivers in Kenya, 3ie Impact Evaluation Report 64. Kelvin, EA, Mwai, E, Romo, ML, George, G, Govender, K, Mantell, JE, Strauss, M, Nyaga, EN and Odhiambo, JO, 2017.

Integration of EPI and paediatric HIV services for improved ART initiation in Zimbabwe, 3ie Impact Evaluation Report 63. Prescott, M, Boeke, C, Gotora, T, Mafaune, HW, Motsi, W, Graves, J, Mangwiro, A and McCarthy, E, 2017.

Increasing male partner HIV testing using self-test kits in Kenya, 3ie Impact Evaluation Report 62. Gichangi, A, Korte, JE, Wambua, J, Vrana, C and Stevens, D, 2017.

Evaluating the impact of community health worker integration into prevention of mother-to-child transmission of HIV services in Tanzania, 3ie Impact Evaluation Report 61. Nance, N, McCoy, S, Ngilangwa, D, Masanja, J, Njau, P and Noronha, R, 2017.

Using HIV self-testing to promote male partner and couples testing in Kenya, 3ie Impact Evaluation Report 60. Thirumurthy, H, Omanga, E, Obonyo, B, Masters, S and Agot, K, 2017.

Increasing male partner HIV self-testing at antenatal care clinics in Kenya, 3ie Impact Evaluation Report 59. Gichangi, A, Korte, JE, Wambua, J, Vrana, C and Stevens, D, 2017.

Impact of free availability of public childcare on labour supply and child development in Brazil, 3ie Impact Evaluation Report 58. Attanasio, O, Paes de Barros, R, Carneiro, P, Evans, D, Lima, L, Olinto, P and Schady, N, 2017.

Estimating the effects of a low-cost early stimulation and parenting education programme in Mexico, 3ie Impact Evaluation Report 57. Cardenas, S, Evans, D and Holland, P, 2017.

The Better Obstetrics in Rural Nigeria study: an impact evaluation of the Nigerian Midwives Service Scheme, 3ie Impact Evaluation Report 56. Okeke, E, Glick, P, Abubakar, IS, Chari, AV, Pitchforth, E, Exley, J, Bashir, U, Setodji, C, Gu, K and Onwujekwe, O, 2017.

*The Productive Safety Net Programme in Ethiopia: impacts on children's schooling, labour and nutritional statu*s, 3ie Impact Evaluation Report 55*.* Berhane, G, Hoddinott, J, Kumar, N and Margolies, A, 2016.

The impact of youth skills training on the financial behaviour, employability and educational choice in Morocco, 3ie Impact Evaluation Report 54*.* Bausch, J, Dyer, P, Gardiner, D, Kluve, J and Mizrokhi, E, 2016.

Using advertisements to create demand for voluntary medical male circumcision in South Africa, 3ie Impact Evaluation Report 53*.* Frade, S, Friedman, W, Rech, D and Wilson, N, 2016.

The use of peer referral incentives to increase demand for voluntary medical male circumcision in Zambia, 3ie Impact Evaluation Report 52. Zanolini, A, Bolton, C, Lyabola, LL, Phiri, G, Samona, A, Kaonga, A and Harsha Thirumurthy, H, 2016.

Using smartphone raffles to increase demand for voluntary medical male circumcision in Tanzania, 3ie Impact Evaluation Report 51. Mahler, H and Bazant, E, 2016.

Voluntary medical male circumcision uptake through soccer in Zimbabwe, 3ie Impact Evaluation Report 50. DeCelles, J, Kaufman, Z, Bhauti, K, Hershow, R, Weiss, H, Chaibva, C, Moyo, N, Braunschweig, E, Mantula, F, Hatzold, K and Ross, D, 2016.

Measuring the impact of SMS-based interventions on uptake of voluntary medical male circumcision in Zambia, 3ie Impact Evaluation Report 49*.* Leiby, K, Connor, A, Tsague, L, Sapele, C, Koanga, A, Kakaire, J and Wang, P, 2016.

Assessing the impact of delivering messages through intimate partners to create demand for voluntary medical male circumcision in Uganda, 3ie Impact Evaluation Report 48. Semeere, AS, Bbaale, DS, Castelnuovo, B, Kiragga, A, Kigozi, J, Muganzi, A, Kambugu, A and Coutinho, AG, 2016.

Optimising the use of economic interventions to increase demand for voluntary medical male circumcision in Kenya, 3ie Impact Evaluation Report 47. Thirumurthy, H, Omanga, E, Rao, SO, Murray, K, Masters, S and Agot, K, 2016.

The impact of earned and windfall cash transfers on livelihoods and conservation in Sierra Leone, 3ie Impact Evaluation Report 46. Bulte, E, Conteh, B, Kontoleon, A, List, J, Mokuwa, E, Richards, P, Turley, T and Voors, M, 2016.

Property tax experiment in Pakistan: Incentivising tax collection and improving performance, 3ie Impact Evaluation Report 45. Khan, A, Khwaja, A and Olken, B, 2016.

Impact of mobile message reminders on tuberculosis treatment outcomes in Pakistan, 3ie Impact Evaluation Report 44. Mohammed, S, Glennerster, R and Khan, A, 2016.

Making networks work for policy: Evidence from agricultural technology adoption in Malawi, 3ie Impact Evaluation Report 43*.* Beaman, L, BenYishay, A, Fatch, P, Magruder, J and Mobarak, AM, 2016.

Estimating the impact and cost-effectiveness of expanding access to secondary education in Ghana, 3ie Impact Evaluation Report 42. Dupas, P, Duflo, E and Kremer, M, 2016.

Evaluating the effectiveness of computers as tutors in China, 3ie Impact Evaluation Report 41. Mo, D, Bai, Y, Boswell, M and Rozelle, S, 2016.

Micro entrepreneurship support programme in Chile, 3ie Impact Evaluation Report 40. Martínez, CA, Puentes, EE and Ruiz-Tagle, JV, 2016.

Thirty-five years later: evaluating the impacts of a child health and family planning programme in Bangladesh, 3ie Impact Evaluation Report 39. Barham, T, Kuhn, R, Menken, J and Razzaque, A, 2016.

Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection and malnutrition in India, 3ie Impact Evaluation Report 38. Clasen, T, Boisson, S, Routray, P, Torondel, B, Bell, M, Cumming, O, Ensink, J, Freeman, M and Jenkins, M, 2016.

Evaluating the impact of vocational education vouchers on out-of-school youth in Kenya, 3ie Impact Evaluation Report 37. Hicks, JH, Kremer, M, Mbiti, I and Miguel, E, 2016.

Removing barriers to higher education in Chile: evaluation of peer effects and scholarships for test preparation, 3ie Impact Evaluation Report 36. Banerjee, A, Duflo E and Gallego, F, 2016.

Sustainability of impact: dimensions of decline and persistence in adopting a biofortified crop in Uganda, 3ie Impact Evaluation Report 35. McNiven, S, Gilligan, DO and Hotz, C 2016.

A triple win? The impact of Tanzania's Joint Forest Management programme on livelihoods, governance and forests, 3ie Impact Evaluation Report 34. Persha, L and Meshack, C, 2016.

The effect of conditional transfers on intimate partner violence: evidence from Northern Ecuador, 3ie Impact Evaluation Report 33. Hidrobo, M, Peterman, A and Heise, L, 2016.

The effect of transfers and preschool on children's cognitive development in Uganda, 3ie Impact Evaluation Report 32. Gillian, DO and Roy, S, 2016.

Can egovernance reduce capture of public programmes? Experimental evidence from India's employment guarantee, 3ie Impact Evaluation Report 31. Banerjee, A, Duflo, E, Imbert, C, Mathew, S and Pande, R, 2015.

Improving maternal and child health in India: evaluating demand and supply strategies, 3ie Impact Evaluation Report 30. Mohanan, M, Miller, G, Forgia, GL, Shekhar, S and Singh, K, 2016.

Smallholder access to weather securities in India: demand and impact on production decisions, 3ie Impact Evaluation Report 28. Ceballos, F, Manuel, I, Robles, M and Butler, A, 2015.

What happens once the intervention ends? The medium-term impacts of a cash transfer programme in Malawi, 3ie Impact Evaluation Report 27. Baird, S, Chirwa, E, McIntosh, C and Özler, B, 2015.

Validation of hearing screening procedures in Ecuadorian schools, 3ie Impact Evaluation Report 26. Muñoz, K, White, K, Callow-Heusser, C and Ortiz, E, 2015.

Assessing the impact of farmer field schools on fertilizer use in China, 3ie Impact Evaluation Report 25. Burger, N, Fu, M, Gu, K, Jia, X, Kumar, KB and Mingliang, G, 2015.

The SASA! study: a cluster randomised trial to assess the impact of a violence and HIV prevention programme in Kampala, Uganda, 3ie Impact Evaluation Report 24. Watts, C, Devries, K, Kiss, L, Abramsky, T, Kyegombe, N and Michau, L, 2014.

Enhancing food production and food security through improved inputs: an evaluation of Tanzania's National Agricultural Input Voucher Scheme with a focus on gender impacts, 3ie Impact Evaluation Report 23. Gine, X, Patel, S, Cuellar-Martinez, C, McCoy, S and Lauren, R, 2015.

A wide angle view of learning: evaluation of the CCE and LEP programmes in Haryana, 3ie Impact Evaluation Report 22. Duflo, E, Berry, J, Mukerji, S and Shotland, M, 2015.

Shelter from the storm: upgrading housing infrastructure in Latin American slums, 3ie Impact Evaluation Report 21. Galiani, S, Gertler, P, Cooper, R, Martinez, S, Ross, A and Undurraga, R, 2015.

Environmental and socioeconomic impacts of Mexico's payments for ecosystem services programme, 3ie Impact Evaluation Report 20. Alix-Garcia, J, Aronson, G, Radeloff, V, Ramirez-Reyes, C, Shapiro, E, Sims, K and Yañez-Pagans, P, 2015.

A randomised evaluation of the effects of an agricultural insurance programme on rural households' behaviour: evidence from China, 3ie Impact Evaluation Report 19. Cai, J, de Janvry, A and Sadoulet, E, 2014.

Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya: a multi-sectoral, prospective, randomised evaluation, 3ie Impact Evaluation Report 18. Brooker, S and Halliday, K, 2015.

Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua, 3ie Impact Evaluation Report 17. Barham, T, Macours, K, Maluccio, JA, Regalia, F, Aguilera, V and Moncada, ME, 2014.

The impact of mother literacy and participation programmes on child learning: evidence from a randomised evaluation in India, 3ie Impact Evaluation Report 16. Banerji, R, Berry, J and Shortland, M, 2014.

A youth wage subsidy experiment for South Africa, 3ie Impact Evaluation Report 15. Levinsohn, J, Rankin, N, Roberts, G and Schöer, V, 2014.

Providing collateral and improving product market access for smallholder farmers: a randomised evaluation of inventory credit in Sierra Leone, 3ie Impact Evaluation Report 14. Casaburi, L, Glennerster, R, Suri, T and Kamara, S, 2014.

Scaling up male circumcision service provision: results from a randomised evaluation in Malawi, 3ie Impact Evaluation Report 13. Thornton, R, Chinkhumba, J, Godlonton, S and Pierotti, R, 2014.

Targeting the poor: evidence from a field experiment in Indonesia, 3ie Impact Evaluation Report 12. Atlas, V, Banerjee, A, Hanna, R, Olken, B, Wai-poi, M and Purnamasari, R, 2014.

An impact evaluation of information disclosure on elected representatives' performance: evidence from rural and urban India, 3ie Impact Evaluation Report 11. Banerjee, A, Duflo, E, Imbert, C, Pande, R, Walton, M and Mahapatra, B, 2014.

Truth-telling by third-party audits and the response of polluting firms: Experimental evidence from India, 3ie Impact Evaluation Report 10. Duflo, E, Greenstone, M, Pande, R and Ryan, N, 2013.

No margin, no mission? Evaluating the role of incentives in the distribution of public goods in Zambia, 3ie Impact Evaluation Report 9. Ashraf, N, Bandiera, O and Jack, K, 2013.

Paying for performance in China's battle against anaemia, 3ie Impact Evaluation Report 8. Zhang, L, Rozelle, S and Shi, Y, 2013.

Social and economic impacts of Tuungane: final report on the effects of a community-driven reconstruction programme in the Democratic Republic of Congo, 3ie Impact Evaluation Report 7. Humphreys, M, Sanchez de la Sierra, R and van der Windt, P, 2013.

The impact of daycare on maternal labour supply and child development in Mexico, 3ie Impact Evaluation Report 6. Angeles, G, Gadsden, P, Galiani, S, Gertler, P, Herrera, A, Kariger, P and Seira, E, 2014.

Impact evaluation of the non-contributory social pension programme 70 y más in Mexico, 3ie Impact Evaluation Report 5. Rodríguez, A, Espinoza, B, Tamayo, K, Pereda, P, Góngora, V, Tagliaferro, G and Solís, M, 2014.

Does marginal cost pricing of electricity affect groundwater pumping behaviour of farmers? Evidence from India, 3ie Impact Evaluation Report 4. Meenakshi, JV, Banerji, A, Mukherji, A and Gupta, A, 2013.

The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone, 3ie Impact Evaluation Report 3. Casey, K, Glennerster, R and Miguel, E, 2013.

A rapid assessment randomised-controlled trial of improved cookstoves in rural Ghana, 3ie Impact Evaluation Report 2. Burwen, J and Levine, DI, 2012.

The promise of preschool in Africa: A randomised impact evaluation of early childhood development in rural Mozambique, 3ie Impact Evaluation Report 1. Martinez, S, Naudeau, S and Pereira, V, 2012.

In Bangladesh, the problem of access to safe drinking water is particularly acute. Despite large existing volumes of renewable freshwater, even relative to its high population density, people in Bangladesh are exposed to drinking water with bacterial and arsenic contamination. This impact evaluation measures the impact of a programme introduced to improve access to safe drinking water in rural Bangladesh by constructing safe drinking water sources like deep tubewells and providing technical guidance on building, designing and implementing these new sources.

Impact Evaluation Series

International Initiative for Impact Evaluation
202-203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org

UKaid from the British people