Access to safe drinking water
Experimental evidence from new water sources in Bangladesh
Pre-Analysis Plan

Anna Tompsett*[1], Serena Cocciolo[1], Selene Ghisolfi[1] and Ahasan Habib[2]

[1]Stockholm University, Sweden
[2]NGO Forum for Public Health, Dhaka, Bangladesh

March 8, 2018

Abstract

This document describes the design and analysis plan for evaluating the impact of a program designed to increase access to safe drinking water on household water quality. The program creates new, safe sources of drinking water in communities in rural Bangladesh. We evaluate the effects of the program on behaviour with respect to access to safe drinking water — source selection, transport distance, and storage practices — and how these changes in behaviour determine household drinking water quality.

The goal of this document is to outline the key research questions and the specifications to be used in the empirical analysis. This document was written before collection of follow-up data. We do not exclude the possibility to conduct additional exploratory analyses. When reporting results we will mark all analyses not planned ex-ante and therefore not included in this document.

1 Motivation

SDG6 sets out the challenge of ensuring availability and sustainable management of water and sanitation for all. However, access to safe drinking water remains limited, particularly in rural areas where safe sources may be few and far between. In 2015, 663 million people worldwide still lacked access to improved sources of drinking water; 1.8 billion people drank fecally-contaminated water; and 1000 children a day died from diarrheal disease, associated with poor water quality and sanitation (United Nations, 2016).

In Bangladesh, the focus of this evaluation, the problem of access to safe drinking water is particularly acute. In the 1970s and 1980s, infant mortality in Bangladesh was extremely high, largely as a result of high levels of diarrheal disease resulting from fecal contamination of surface water, used for drinking. Education campaigns encouraged people to shift to obtaining drinking water from groundwater sources instead, resulting in a decline in child mortality (Caldwell, Caldwell, Mitra, & Smith, 2003). However, in the 1990s high but naturally-occurring levels of arsenic

*anna.tompsett@ne.su.se

were discovered in the groundwater. Arsenic is undetectable without water quality tests. By the time the arsenic contamination problem was discovered, an epidemic of diseases associated with arsenic exposure was already established, called “the largest poisoning of a population in history” by Smith, Lingas, and Rahman (2000). Despite years of effort by the Bangladeshi government, non-governmental organizations and international aid agencies, progress on safe drinking water in Bangladesh remains elusive (Human Rights Watch, 2016). Today, almost 100 million people still drink fecally-contaminated water, and 39 million people drink water that is contaminated with arsenic at international standards (BBS & UNICEF, 2015).

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

These uncertainties make it more difficult to design effective interventions to improve access to safe drinking water. In particular, they raise the risk that providing safer but more distant sources may increase exposure to pathogens via contamination in transport. These questions are particularly salient in Bangladesh, where policy-makers must design policy to reduce exposure to arsenic contamination without increasing exposure to fecal/bacterial contamination. Our evaluation will measure the causal impacts of source water quality and transport time on household water quality in rural Bangladesh. The results from our study will inform policy to improve access to safe drinking water in Bangladesh, and more broadly, to achieve SDG6 worldwide.

1.1 Program description

We evaluate the effects of a program designed to improve access to safe drinking water in rural Bangladesh. The program consists of a package of subsidies and technical advice to build new public sources of water, which provide drinking water that is free of both arsenic and bacterial contamination. The new safe sources of water are deep tubewells, which draw water from aquifers that are sufficiently deep to be safe from both bacterial contamination and arsenic contamination, and we test all sources after installation to confirm that the water is arsenic free. Communities decide the location of new water sources by consensus in community meetings. We offer to install one new water source in smaller communities, and two new water sources in larger communities.

Treated communities are assigned to one of three contribution rules: a third of treated villages were required to raise a cash contribution before installation; a third of treated villages were required to contribute labour; and a third of treated villages received the program under a contribution waiver. Our primary interest in this study is the average effect across the three contribution rules. However, take-up varies under the three contribution rules, and the variation in take-up may be important for statistical power.

1.2 Key Research Questions

The key research questions we will evaluate are as follows:

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

We note that we expect to have stronger causal evidence on Key Research Questions 3a and 3b than regarding Key Research Question 3c, as we discuss further in section 3.3.

2 Research Design

2.1 Sampling

2.1.1 Sampling Frame

Our study is located in north-western Bangladesh, in Shibganj and Sonatala Upazilas in Bogra District and in Gobindaganj Upazila in Gaibandha District. Within these Upazilas, we target villages with high levels of arsenic contamination, using the limited data available to pre-select villages and then refining selection using testing. We pre-selected a list of candidate villages for the intervention on the basis of contamination levels reported in the available sources of arsenic testing data. We had access to village-level data from the following data sources: (i) data from the Bangladesh Arsenic Mitigation Water Supply Project (BAMWSP), which included a large tubewells screening program conducted between 1999 and 2006; (ii) the assessment from the Department of Public Health Engineering (DPHE) on the most arsenic contaminated villages in the Bogra region; (iii) data collected in 2008 from the Bangladesh Social Development Services (BSDS). We pre-selected as candidate villages for receiving our intervention all villages indicated by the DPHE or for which BAMWSP or BSDS data reported a share of arsenic contaminated tubewells equal or higher than 30%. We confirmed this initial selection by testing for arsenic contamination a small sample of tubewells in the village.

For these candidate villages, we obtained the most updated list of resident households from administrative sources. For logistical reasons, we implement the program in geographically defined treatment units of between 50 and 250 households; we use the terms treatment unit and community interchangeably throughout this document. To define treatment units, we used available household administrative lists in order to obtain village sizes, exclude from the study villages with less than 50 households and divide larger villages into several smaller treatment units along natural boundaries. Following this process, we identified 192 candidate treatment units in 103 villages, of which 51 were divided in two or more treatment units

We conducted a full census of existing sources of drinking water in these candidate treatment units. We used the water source contamination data in order to finalize the selection of the treatment units eligible for receiving the arsenic mitigation program. In particular, we excluded from the study all treatment units with less than 15% of arsenic-contaminated water sources. We further screened treatment units with less than 25% of arsenic-contaminated water sources, including them in the program only if they presented a well defined cluster of contaminated water sources. To evaluate these treatment units with between 15% and 25% contamination, we reviewed the maps obtained from the water source census. 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. This process lead us to identify 171 eligible treatment units, which is the final number of treatment units enrolled in the project.

We used the available household administrative lists in order to randomly sample 40 households per treatment unit for the household survey. We accommodated cases when selected households were not available for the interview or refused to participate by providing enumerators with a list of “replacement households”, sorted in random order. Enumerators documented this replacement process in the household list used by the enumerators and recorded outcomes in the survey form, as they were required to fill in a form for all household that they tried to locate and conduct the interview with.

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

2.1.2 Statistical Power

We carry out power calculations by simulating follow-up data using baseline data, project implementation data, and plausible parameter values and assumptions about behavioural change (based on previous studies and our own experience). Intra-cluster correlation is modelled implicitly via the true intra-cluster correlation in the baseline data. Where relevant, we cluster standard errors by treatment unit. Our simulation is based on the following key assumptions: i) that absent our program, approximately 1/3 households would switch water source between baseline and follow-up; ii) that water quality in the new source would be a random draw from the baseline distribution in that treatment unit; iii) that distance to the new source is a random draw between the minimum distance to a source in the treatment unit and a 50% increase in walking distance. These assumptions yield behavior that matches our previous studies, baseline data and qualitative reports from the field.

117 treatment units are assigned to treatment under one of the three contribution conditions.

We successfully installed water sources in 82 (70%) of these treatment units. Based on these installation numbers, we simulate take-up of installed sources based on what households reported to us at baseline about whether they would adopt a new safe source at a given distance from their home. These assumptions yield aggregate take-up rates that are consistent with our previous work (Madajewicz, Tompsett, & Habib, in preparation). We then simulate water source quality and distance to a water source at follow-up. To simulate storage behavior at follow-up, we assume that household storage behavior is correlated across time (using, as an imperfect proxy, the correlation between a measure of habitual storage behavior and a measure of observed storage behavior at the time of the baseline survey), and increases with distance to collect water. We use the correlation between source and household arsenic levels at baseline to predict household arsenic levels at follow-up. Finally, we use plausible effect sizes to simulate household bacteria levels at follow-up i.e. we assume that switching to a bacteria-free source yields a 30% drop in contamination (Kremer et al, 2011); that walking an additional 100m to collect water increases risk of bacterial contamination by 2%; and that storage increases the risk of bacterial contamination by 5%.

For the reduced form analyses, we find that our study has minimum detectable effects at the 5% level (2.8 x estimated standard deviation of coefficients) of: a 3.5% change in household arsenic contamination; 3.8% change in household bacterial contamination; 2.4% change in source arsenic contamination; a 2.8% change in source fecal contamination; an average change of 2.2m in walking distance (7% of median distance to a water source at baseline); and 3% change in the rate of water storage before drinking. These compare favorably to expected treatment effects.

2.1.3 Assignment to treatment

We randomly assigned the 171 candidate treatment units to four study arms. We assigned 42 candidate treatment units to a control group which received no intervention. We assigned 43 candidate treatment units to receive the safe drinking water program under one of the three contribution requirements: (i) cash approach; (ii) labour approach; (iii) waiver approach. We conducted the randomization at public lottery meetings, to which we invited representatives from each eligible community. The randomization was stratified by Union Parishad to make it feasible for representatives of the study communities to attend. The decision to use public randomization was motivated by concerns about transparency, especially given that we offer the same program under different conditions in different communities. We anticipated that information about the different conditions would spread, and this was indeed the case. The public lottery meetings gave our research staff an important source of legitimacy for project decisions taken.

Large treatment units were offered two tubewells; smaller treatment units were offered one tubewell, using an algorithm to assign the number of tubewells as a function of the original village size or of the treatment unit size.[1]

2.1.4 Attrition

Overall, we expect relatively low attrition i.e. we expect to find some members of the same household in the same house in a large (>97%) percentage of cases. Since overall attrition is expected to be low, differential attrition by treatment groups is also less likely. We did not experience differential attrition rates in a previous study with some similarities (Madajewicz et al., in preparation).

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

To minimize attrition, we build flexibility into our follow-up survey to schedule interviews with households at the time of their convenience, and include a short confirmatory telephone survey with any households who are reported to have migrated at the time of follow-up.

2.2 Fieldwork

2.2.1 Instruments

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

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

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

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

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

We conducted the bacteria test for all water sources and for all households surveyed, provided that the survey respondent agreed to the testing procedure.[2] The water testing procedure for bacteria contamination used hydrogen sulfide vials produced by NGO Forum for Public Health. The test detects the presence of Escherichia coli in water. The vials should be kept at room temperature for 48 hours, and the test is read as positive if the colour changed from clear to black. The hydrogen sulfide test has been rigorously evaluated in Bangladesh by NGO Forum for Public Health. We informed respondents about the bacteria test results when the results were ready, on average two days after the water sample collection, by SMS. During the water source and household

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

survey we asked respondents to provide us with a phone number to be used for sending by SMS the results from the bacteria test. 99% of the respondents in the water source survey and 94% in the household survey provided us with a phone number for further communications. Project staff entered the bacteria test results on average after 2 days from the water sample collection, which were promptly communicated to surveys' respondents via automated SMS.

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

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

For this reason, for a subset of households, we complemented this testing procedure with a laboratory test conducted at the Water Quality Testing Laboratory (WQTL) of NGO Forum for Public Health using Atomic Absorption Spectrophotometer (AAS). We randomly selected for the arsenic laboratory test 10 households, out of the 40 sampled for the household survey, in 92 treatment units. We stopped laboratory testing after 92 treatment units because of budget constraints, as the lab tests are much more expensive (approximately 100 times) than the field tests. In total, we tested in the laboratory 897 water samples collected during the household surveys. The field tests are designed to be somewhat more conservative than the laboratory tests, because a false negative has much more serious consequences for health than a false positive. However, when the results for the two sets of tests are compared, they are highly correlated.

3 Empirical Analysis

3.1 Variables of interest

The main variables of interest are summarized in Table 1. Where multiple measures for a single outcome variable are listed, the expected main measure is given in bold, and variables we will use to provide corroborating evidence are listed in regular text.

Households may report using multiple sources. At baseline, the vast majority (96.5%) used only one source of drinking water; 3.4% reported using two sources, and only 6 households reported using three sources. It is possible, however, that the rate of multiple source use will increase as a result of the intervention. The primarily analysis will aggregate the source quality and distance

Table 1: Variables of interest

Evaluation Question	Variables
1a	**Arsenic field test of household water above WHO standard (10ppb)** Arsenic field test of household water above Bangladeshi standard (50ppb) Arsenic lab test of household water above Bangladeshi standard (50ppb) Arsenic lab test of household water above WHO standard (10ppb) Arsenic field test of household water result Arsenic lab test of household water result
1b	**Indicator for fecal contamination of household water**
2a	**Arsenic field test of source water above WHO standard (10ppb)** Arsenic field test of source water above Bangladeshi standard (50ppb) Arsenic field test of source water result
2b	**Indicator for fecal contamination of source water**
2c	**Calculated distance between household and primary water source in metres** 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

measures based on the fraction of drinking water drawn from each source[3] to provide a measure of average source quality and distance travelled. In robustness checks, we will test whether our results are stable to including and excluding these households from the analyses; and to aggregating information from multiple sources in different ways.

3.2 Balancing checks

We verify that treatment and control groups are balanced in terms of baseline characteristics. To carry out these checks, we collapse the data to community-level means and regress the baseline characteristic on a dummy for treatment, and Union Parishads (district) level dummies, reflecting the stratification of randomization by Union Parishad.

$$y_{cb} = \beta T_c + \eta_d + \epsilon_c \tag{1}$$

where y_{cb} is the mean value of characteristic y at baseline in community c, 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. Standard errors are robust.

The variables we include in the balance checks include the main variables listed in bold in Table 1, as well as other socioeconomic factors that may predict access to safe drinking water or hygiene behaviour. For the full set of variables included in the balance checks, we also test simultaneously for joint significance of the differences in baseline characteristics by: i) reversing Equation 1 to include village level characteristics on the right hand side and a treatment dummy on the left hand side, and testing for joint significance of all baseline characteristics; and ii) conducting a Hotelling's T-squared test.

3.3 Treatment effects

3.3.1 Program effects

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

$$\Delta y_c = \alpha + \beta T_c + \eta_d + \epsilon_c \tag{2}$$

where Δy_c is the change in outcome variable y between baseline and follow-up in community c, 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 effects of the program, regardless of whether or not the program successfully installs water sources or not. These are the treatment effects that are relevant to policy-makers. We include Union Parishad fixed effects to reflect stratification in the original randomization. We use this approach to estimate effects for key research questions 1 and 2.

3.3.2 Mechanisms

To analyze mechanisms in key research question 3, we take two approaches. Our first analysis of mechanisms is a difference-in-difference approach where we evaluate how changes in household

[3] We ask households to calculate the number of times they collect water in a week, and the volume of water collected each time.

bacterial contamination vary with changes in source contamination, transport distance and storage. For household i, we estimate:

$$\begin{aligned} FC^h_{if} - FC^h_{ib} = b_0 + b_1(FC^w_{if} - FC^w_{jb}) + b_2(DIST^w_{if} - DIST^w_{ib}) \\ + b_3(STORAGE_{if} - STORAGE_{ib}) + \eta_c + \epsilon_i \end{aligned} \tag{3}$$

where all variables are measured at baseline b and follow-up f; FC^h is fecal contamination in household i's drinking water and FC^w is fecal contamination in household i's water source; $DIST^w$ is the distance between household i and its drinking water source; and $STORAGE$ is an indicator variable for whether or not household i stores drinking water (as opposed to collecting drinking water on demand). Where the household uses multiple water sources, the values of FC^w and $DIST^w$ are weighted averages across the sources the household reports using. 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 will estimate versions of Equation 3 with and without these community-level dummy variables, as there is no clear ex-ante reason to prefer one approach over the other.[4] When we include the community-level dummy variables, Equation 3 only exploits within-community variation in changes in the right-hand side variables to estimate causal effects.

The difference-in-difference yields causal estimates under the assumption that changes in the right hand side variables are uncorrelated with other changes in household drinking water contamination e.g. through changes in household hygiene practices. Such an assumption is not unreasonable. However, although assignment to the safe drinking water program is random, selection of locations for water source installation is determined, by consensus, at a community meeting. As a result, it remains possible that changes in distance to collect drinking water, or source water contamination, may be correlated with other changes that also affect household drinking water contamination, through other channels. These confounding factors might in principle bias the above analysis. To address this concern, we carry out a second, instrumental variables analysis which exploits the experimental assignment of the safe drinking water program.

The instrumental variables approach uses baseline data to predict where in a village a community will decide to install a water source. We predict location of constructed water sources using baseline data using two methods. First, members of the research team in Stockholm inspect the map of water sources and select a location or locations based on population density and existing source quality. The research assistants who carry out this task do not have any information on final chosen locations, and they follow the same procedure in treatment and control villages.

Second, we will also program an algorithm to similarly predict locations in treatment and control villages. We do not fully pre-specify the details of this algorithm as we are testing several variants, and do not rule out the possibility that we will be able to improve the predictive power of the algorithm. The class of algorithms we will test are deterministic functions of baseline characteristics, augmented by our knowledge of how communities take decisions. For example, candidate algorithms might minimize the distance between all households in a treatment unit and the new location(s), or the distance between households using arsenic-contaminated sources at baseline and the new location(s). Alternatively, the algorithm might minimize the distance between households and their nearest safe source. In all cases, we constrain the set of candidate locations to feasible locations i.e. those that are on habitable (low flood risk land), defined by proximity to households and pre-existing water sources.

We do prespecify the approach we will use to select between alternative algorithms, which is to

[4] Either approach may increase precision, depending on the exact structure of ϵ_i.

inspect the partial R-squared from a regression at household level of distance to the nearest finally selected location on distance to the nearest predicted location. We recorded the final selected location in almost all treatment units (with the exception of one treatment unit which declined to participate; and one treatment unit which did not reach a decision on a location), not only in those treatment units where we were successful in installing a water source. We will identify improvements to the algorithm or select between alternatives based on those algorithms which deliver the highest partial R-squared. Importantly, this regression has no direct role in our final estimating strategy, meaning that we lessen the risk of unintentionally introducing bias into our estimates by selecting an instrument that happens to be correlated with unobservable variables that predict treatment variables. For the same reason, we do not intend to use other approaches, such as machine learning approaches to predict locations, in our main analysis. We will report specifications using alternative algorithms in the case that we have several alternatives that perform similarly well. We will also verify that distance to the nearest predicted location does not differ between treatment and control villages,[5] and does not correlate with baseline characteristics differently in treatment and control villages.[6]

We then use the predicted source location to construct the following instruments. The first instrument is predicted change in source fecal contamination, constructed as follows:

$$PRED\left(FC_{if}^{w} - FC_{ib}^{w}\right) = TAKEUP\left(As_{ib}^{w}, DIST_{i}^{*}\right) \cdot \left(0 - FC_{jb}\right) \cdot T_{c} \tag{4}$$

where $TAKEUP$ is the average take-up rate for households with baseline source arsenic contamination As^{w} and distance from nearest predicted location $DIST^{*}$. We calculate average take-up rates using the full dataset for treatment units where we successfully installed tubewells, calculating take-up rates for households categorized according to baseline source arsenic contamination[7] and distance to the nearest installed source.[8] We allow take-up rates to vary across these groups because households with higher arsenic contamination at baseline report willingness to walk much further to adopt a new source than households with no or low arsenic contamination. $0 - FC_{jb}$ is a measure of the change in source fecal contamination if a household adopted the new source; and T_c is a dummy variable that is 1 if a household belongs to a treated community, and 0 otherwise.

The second instrument is the predicted change in distance to drinking water between baseline and follow-up, constructed as follows:

$$PRED\left(DIST_{if}^{w} - DIST_{ib}^{w}\right) = TAKEUP\left(As_{jb}, DIST_{i}^{*}\right) \cdot \left(DIST *_{i} - DIST_{ib}^{w}\right) \cdot T_{c} \tag{5}$$

where $DIST *_{i} - DIST_{ib}^{w}$ is the change in distance that household i would experience, if the household adopted the new source, and if it were built at the optimal location.

We can then estimate the difference-in-difference equation (Equation 3) using predicted changes as instruments for observed changes in source fecal contamination and distance to drinking water, augmenting the difference-in-difference equation to include controls for the endogenous components

[5] We will estimate Equation 1 with distance to the nearest predicted location as an outcome variable.

[6] We will estimate Equation 1 augmented with distance to the nearest predicted source and its interaction with the treatment dummy, for the same set of baseline characteristics as in the main balance checks, and confirm that the coefficients on the interaction terms are consistent with the same prediction procedure in both treatment and control villages.

[7] Categories are: no contamination, low arsenic contamination (above WHO threshold but below Bangladeshi threshold) or high arsenic contamination (above Bangladeshi threshold).

[8] Estimated for intervals of approximately 1 minute walking time.

of the instruments, as follows:

$$
\begin{aligned}
FC^h_{if} - FC^h_{ib} = b_0 &+ b_1(FC^w_{if} - FC^w_{ib}) + b_2(DIST^w_{if} - DIST^w_{ib}) \\
&+ b_3(STORAGE_{if} - STORAGE_{ib}) \\
&\quad + b_4 TAKEUP\,(As^w_{ib}, DIST^*_i) \cdot (0 - FC_{jb}) \\
&\qquad + b_5 TAKEUP\,(As_{jb}, DIST^*_i) \cdot (DIST *_i - DIST^w_{ib}) \\
&\qquad\quad + \eta_c + \epsilon_i
\end{aligned} \tag{6}
$$

Conditional on these controls, instrument exogeneity follows from the inclusion of the treated indicator, which is randomly assigned. Identification then follows from comparing changes in households with similar baseline characteristics in treated and control groups. Throughout, we cluster standard errors at the treatment unit level to account for spatial correlation in outcome variables. As with the difference in difference analysis, we will report results both with and without community dummies. Additionally, we will report results from estimating Equation 6 with an additional control for whether or not the treatment unit received treatment or not. Including this treatment dummy should not alter the results under the assumption that treatment only affects fecal contamination in household drinking water through changes in behaviour with respect to use of water sources. If the results change, we will evaluate potential channels through which this assumption could be violated, for example via reporting bias or other changes in hygiene behaviour.

In our simulated power calculations, we obtain Sanderson-Windmeijer first stage F-statistics of more than 10 for both instruments in about 85% of simulations. However, the IV approach sacrifices considerable power: the IV approach has minimum detectable effects that are approximately ten times larger than the difference in difference approach. We will interpret the results from the IV analysis with these limitations in mind.

3.4 Heterogeneous effects

A key source of heterogeneity in the effects of new water sources on household bacterial contamination comes from heterogeneity in the effect of new water sources on water transport distance. Households that previously walked a long way to collect safe water may walk a shorter distance as a result of a new, nearby safe source. Households that previously collected water at a nearby unsafe source may walk further to collect water, if a new safe source becomes available. This motivates two heterogeneity analyses for the reduced form analyses (Key Research Questions 1 and 2).

For all heterogeneity analyses, we will re-estimate Equation 2 at the household level, with weights constructed so that each treatment unit counts equally in the analysis, and model heterogeneous effects as follows:

By use of safe/unsafe sources at baseline Those already using safe sources are more likely to reduce their distance to collect water, as they will only adopt a new source if it is closer to them or otherwise more convenient to use. Those using unsafe sources are more likely to increase their distance to collect water, because they may be willing to walk further to access safe water than they are willing to walk to access unsafe water. To analyze heterogeneous effects by safety of baseline source, we will construct categorical variables describing arsenic contamination levels at baseline: no contamination, low contamination (10-50ppb), high contamination (50-100 ppb), and very high contamination (>100 ppb). We will then include dummy variables for these categories of arsenic contamination at baseline, and their interactions with the treatment dummy.

By distance to the new source The effect on transport distance is likely to vary with distance from the source. Since the source location is chosen by the community, we will use distance to the predicted location as a proxy for distance to the constructed source. This analysis has important equity implications for the planning of safe drinking water access projects. To analyze heterogeneous effects by distance to the predicted location, we will categorize households into 5 quantiles by distance to the predicted location. We will then include dummy variables for these categories of distance to the predicted location, and their interactions with the treatment dummy. We will also report results for a similar analysis including distance to the predicted location and its interaction with the treatment dummy.

We will report results with and without community-level dummies. The two sets of results have slightly different interpretations. Without community-level dummies, the comparisons include heterogeneous effects that are correlated with the community-average characteristics (for example, a potentially greater likelihood of successful well installation in more highly-arsenic contamination villages). When we include community-level dummies, the comparisons describe within-village heterogeneity only.

We will also evaluate whether the estimated heterogeneous effects on these groups for Key Research Questions 1 and 2 are consistent with the expected effect sizes implied by the results of the analysis of Key Research Question 3.

Additionally, we will report heterogeneous effects for the poor (28% self-report status as poor at baseline), as these results are of particular interest to policy makers. Similarly to the above, we will analyze heterogeneity by including a dummy for self-reported status as poor at baseline, and its interaction with the treatment dummy.

3.5 Standard Error Adjustments

Our main analyses will report cluster-robust standard errors, treating outcome variables as correlated within treatment units. Additionally, we will report p values derived from randomization-based inference, by randomly reshuffling treatment status 500 times and comparing the estimated coefficients to the distribution of coefficients obtained under the randomization-based inference.

For each of the key research questions, we will report "naïve," per comparison, p values (Kling, Lieberman, & Katz, 2007). Since we test multiple hypotheses, there is therefore a possibility that some of the comparisons which appear statistically significant occur due to chance. As a result, we will also follow Westfall and Young (1992) and calculate adjusted p values which correctly control the family wise error rate in the presence of correlated effects within each group of research questions (1, 2 and 3).

References

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

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

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

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Jr, J. M. C. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and meta-analysis. *The Lancet Infectious Disease*, *5*.

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

Human Rights Watch. (2016). *Nepotism and neglect: The failing response to arsenic in the drinking water of Bangladesh's rural poor.* Human Rights Watch.

Kling, J. R., Lieberman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1).

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

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

United Nations. (2016). *Sustainable development goals. goal 6: Facts and figures.*

Westfall, P. H. & Young, S. S. (1992). *Resampling-based multiple testing: Examples and methods for p-value adjustment.* John Wiley & Sons.

Wright, J., Gundry, S., & Conroy, R. (2004). Household drinking water in developing countries: A systematic review of microbiological contamination between source and point-of-use. *Tropical Medicine and International Health*, *9*(1).

Wu, J., van Geen, A., Ahmed, K. M., Alam, Y. A. J., Culligan, P. J., Escamilla, V., ... Emch, M. (2011). Increase in diarrheal disease associated with arsenic mitigation in Bangladesh. *PLOS One*, *6*(12).