Community Participation in Decision-Making
Evidence from an experiment in providing safe drinking water in Bangladesh

Malgosia Madajewicz[1], Anna Tompsett[2] and Ahasan Habib[3]

[1]NASA Goddard Institute for Space Studies, Columbia University, New York
[2]School of International and Public Affairs, Columbia University, New York
[3]NGO Forum for Public Health, Dhaka, Bangladesh

March 31, 2014

Abstract

The hypothesis that participation in decision-making by intended beneficiaries of social programs improves the outcomes of those programs has long been influential in the academic literature and in policy. This paper presents the first experimental evidence on whether participation in project decision-making affects the outcomes of a social program. We randomly assigned participatory and non-participatory decision-making structures to communities who received an otherwise identical intervention, a package of technical advices and subsidies to provide safe drinking water sources. Participation in decision-making resulted in larger reported increases in access to safe drinking water, but only when we imposed rules on the decision-making process that were designed to limit the appropriation of project benefits by elite or influential groups or individuals. Villages in which communities participated in decision-making under rules designed to prevent appropriation reported a significantly greater increase in access to safe drinking water (an increase of 21%) relative to villages in which project staff took decisions (13%). No statistically significant increase was reported in villages in which the communities participated in decision-making without imposed rules (15%). We conclude that the rules we applied to limit appropriation – minimum representation requirements and decision by unanimous consensus – were effective in accomplishing their objective.

JEL Classification Numbers: O20 General Development Planning and Policy; H41 Public Goods; D72 Political Processes

1 Introduction

The hypothesis that participation in decision-making by intended beneficiaries of social programs improves the outcomes of those programs has been influential in the academic literature and in policy for some time (e.g. Stiglitz (2002), World Bank (2004)). Advocates of the policy argue that involving communities in project decision-making has multiple benefits: improving project targeting, by drawing on information available to the community but not to outsiders; increasing 'buy–in' and generating a 'sense of ownership' of the project, therefore leading to better long-term management and maintenance of program assets; and improving transparency and accountability in project delivery. However, programs in which communities participate in decision-making may be more susceptible to the 'capture' of project benefits by elite or influential community members[1].

Much of the early evidence in support of this hypothesis was based on cross-sectional analyses [2] case studies [3], or was simply anecdotal. Since the choice of a decision-making structure is likely to be correlated with project, community and implementing agency characteristics, identification of causal effects is difficult and sensitive to critical assumptions. This paper presents the first experimental evidence on whether and how involving intended beneficiaries in program decision-making affects program outcomes.[4]

We randomly assigned different decision-making structures to communities who received an otherwise identical intervention, a package of subsidies and technical advice to provide up to three sources of arsenic-safe drinking water. Many rural Bangladeshi communities currently use sources of water that are susceptible to arsenic or, less commonly, bacterial contamination. Arsenic-safe drinking water sources are relatively expensive and the vast majority of households cannot afford to obtain them for themselves. As a result, the sources must generally be provided at a community level. The random assignment ensured that the communities in which we implemented the project under different decision-making structures were comparable in terms of all other characteristics, allowing us to draw causal inferences about the impacts of the decision-making structures on project outcomes.

[1] See Mansuri and Rao (2013) for a comprehensive review.

[2] Examples include: Isham, Narayan, and Pritchett (1995), Sara and Katz (1997), Khwaja (2004), Fritzen (2007)

[3] Examples include: Kleemeier (2000), Rao and Ibáñez (2005)

[4] Other recent experimental studies have provided evidence for how other changes to the decision-making procedure affect project outcomes e.g. Olken (2010) and Beath, Christia, and Enikolopov (2013) compare participatory decisions taken at representative-based or community-wide meetings to those taken by anonymous referendum.

The decision-making structures assigned included a non-participatory decision-making structure in which project staff took all decisions, based on information provided by the community, and two participatory decision-making structures. In the first participatory decision-making structure, the community took all decisions using their own internal decision-making processes. In the second, we imposed rules on the decision-making process. Under this structure, the community took all decisions by unanimous consensus at a meeting organized by project staff with a requirement for representation of women and the poor.

Under all decision-making models, we retained an important participatory component. After decisions were taken, all treatment villages were required to contribute between 10 and 20% of the total cost of water source installation. The communities then had to decide whether or not they would contribute, and how this contribution would be raised. We therefore identify the effects of participation and decision-making over and above the effects generated by any financial contribution.

Overall, the intervention led to an increase in reported access to safe drinking water of 16.3% relative to a control group. The average treatment effect rises to 18.3%, compared to a matched control group, when we exclude a subset of villages in which the only feasible technology for providing arsenic-safe drinking water year-round was an arsenic iron removal plant (AIRP). This technology has experienced issues with reliability and effectiveness in the past (Hossain et al., 2005) and our experience suggests that communities strongly prefer tubewells to AIRPs. The treatment effect in the villages in which AIRPs are the only technically feasible option is not statistically different from zero, compared to a matched control group.

The increase in access to safe drinking water was higher in villages in which the community took decisions and in which decision-making rules were imposed (21% in all villages; 24% if we exclude the AIRP villages) compared to the villages in which project staff took decisions (13%; 14% if we exclude the AIRP villages). However, no significant increases were realized when the community took decisions without the imposition of decision-making rules (15%; 16% if we exclude the AIRP villages). The difference between the change in reported access to safe drinking water in villages in which the community took decisions under imposed rules and the remainder of the treated villages is significant when we remove the villages in which AIRPs were the only option from the analysis. Since the treatment effect is zero in these villages regardless of the structure under which decisions were taken, including them in the analysis is not informative with regards to a comparison between decision-making structures.

We installed an average of 2.1 arsenic safe water sources in each of 127 treatment villages. We installed a slightly larger number of wells in villages in which the community was involved in decision-making (2.2 across both participatory decision-making structures) compared to those in which project staff took decisions (2.0). However, the differences are not statistically significant. Under the non-participatory structure, project staff were instructed to propose locations for water sources in public spaces wherever feasible in order to facilitate access to the sources. Under the participatory structures, communities were more likely to locate the water sources on private land. We installed 1.9 source per village on public land when project staff made decisions, and 1.3 when communities took decisions. A significantly smaller number of individuals contributed money towards the water sources in the communities which took decisions without any imposed rules (5 individuals per village), when compared to the other two models (9 individuals).

The results suggest that involving communities in decision-making can lead to greater project impacts in terms of number of projects successfully completed and changes in reported access to safe drinking water. However, the results also suggest that devolving decision-making authority to the community without measures to avoid co-option of the decision-making process by influential groups or individuals can lead to an increased incidence of elite capture. In our case, the number of safe water sources constructed increases without any reported increase in access to safe drinking water.

The paper is structured as follows. Section 2 describes the setting, the experimental design and the data; section 3 describes the results, and section 4 concludes.

2 Setting, Experimental Design and Data

2.1 Arsenic Pollution Problem in Bangladesh

The context for this study is the arsenic contamination problem in rural Bangladesh, where communities rely heavily on groundwater drawn from aquifers for drinking and cooking. In the 1970s and early 1980s, many international agencies promoted the use of groundwater as a safer alternative to surface water, which is often contaminated by pathogens. At the time, noone had realized that some aquifers in the region have high concentrations of naturally occurring arsenic. Arsenic contamination is not readily detectable in water, and symptoms of arsenic poisoning only appear after years of exposure and accumulation in the body. Information about high concentrations of arsenic in tubewells emerged only in the mid-1990s. The resulting epidemic of diseases associated with arsenic exposure has been described as the largest poisoning of a population in history (Smith, Lingas, & Rahman, 2000). In 2008, when this project began, UNICEF estimated that 20 million people were still using water from wells with arsenic concentrations above the Bangladeshi standard, which is itself five times higher than the WHO standard (UNICEF, 2008).

Creating access to safe drinking water in the presence of arsenic contamination presents a problem of providing a local public good. The great majority of tubewells in Bangladesh are privately owned, including almost all tubewells that have high concentrations of arsenic. Sources of water that have low concentrations of arsenic are considerably more expensive, and only the richest households can afford to purchase these sources themselves. For most households, they must be provided at the community level.

We conducted the study in communities located in two *upazilas* (subdistricts): Gopalganj, about 60 miles southwest of Dhaka, and Matlab, about 30 miles southeast of Dhaka. We focused on these sites because of the severity of the arsenic contamination problem in the area more than 80% of pre-existing tubewells were arsenic contaminated and because the sites had not yet received other interventions to address the problem. We studied 250 villages, equally split between the two upazilas, and ranging in size from a minimum of 7 households to a maximum of 1103, with the median size 170 households. [5]

The most commonly used sources of arsenic safe water are deep tubewells, which draw water from deep aquifers (approximately 700-800 feet below ground level) that have low concentrations of arsenic. Standard deep tubewells are relatively expensive to install, but easy to use and maintain and parts are readily available. In some areas, arsenic safe water is available at lesser depths of approximately 300-400 feet, although water drawn from aquifers at this depth may have other contaminants including manganese. In these areas, shallow tubewells can be constructed, which cost less to install than deep tubewells but are otherwise very similar in terms of functionality,

[5] Data on arsenic contamination of pre-existing tubewells and village size was drawn from the Bangladesh Arsenic Mitigation Water Supply Project.

maintenance requirements and ease of repair.[6] If there is considerable seasonal variation in water pressure in the aquifer, standard deep tubewells may not provide year-round access to safe drinking water. An alternative design – the deep-set tubewell – is available and can provide year-round access to safe drinking water in this context. The deep-set tubewell is more expensive and more difficult to repair in case of failure than the standard deep tubewell, but is as convenient and easy to use.

In some villages, there is no accessible arsenic-safe aquifer – for example, where an intermediate layer of rock cannot be penetrated using local drilling techniques – and we could not install tubewells. In these villages, we offered communities the opportunity to install an arsenic iron removal plant (AIRP). AIRPs remove arsenic from shallow groundwater by oxidation and filtration. They are more expensive, larger and significantly more difficult to operate and maintain than tubewells, and our experience suggested that communities strongly preferred tubewells. [7] As a result, we will report treatment effects by the type of feasible technology – AIRPs or tubewells – as well as the overall treatment effect.

Before installing a safe drinking water source, we required the community to contribute between 10% and 20% of its cost, depending on the technology installed. Table 1 shows the cost of installing each of these technologies and the community contribution that we required. The difference in required community contributions reflects the difference in cost of the selected technology. We also scaled the community contribution so that the subsidy could be either concentrated on one water source or spread between up to three water sources. The price per water source therefore increased as more water sources were installed in the village. Budget constraints meant that when the best feasible technology was one of the more expensive alternatives, we were only able to offer up to two water sources.

We carried out the interventions between 2008 and 2011, in partnership with a Bangladeshi NGO, NGO Forum for Public Health. NGO Forum for Public Health is a well-established actor in the water and sanitation sector with more than 30 years experience in the field.

[6] During the study implementation period, information emerged about a problem of manganese contamination in shallow tubewells. As a result, we replaced those shallow tubewells we had already installed which tested positive for manganese with deep-set tubewells, free of charge.

[7] Where tubewells were not feasible, we also offered communities the opportunity to install rainwater harvesting systems or a pond sand filter, but since no community selected either of these options, we do not describe them further in the paper. Both technologies have limitations with respect to tubewells or AIRPs

2.2 Experimental Design

The project intervention consisted of a package of technical advice and subsidies for the provision of up to three safe drinking water sources per community. Before interventions began, we carried out an information campaign about the arsenic problem, to ensure that all villages were initially equally well informed about the arsenic problem.

Of the 250 villages studied, we assigned 100 to a control group who did not receive the intervention. 126 villages received the intervention. We initially assigned a further 24 villages to receive the intervention who eventually did not receive the intervention, due to changes in the costs of providing safe water sources over the course of the project. We originally assigned one other village to treatment, but project staff determined before the project began that there were no feasible available technologies to provide safe drinking water in the community, because no arsenic safe aquifer was accessible, and arsenic concentrations in the shallow groundwater were too high for removal with an AIRP. There was one other village in which we determined after we began the intervention that there were no feasible available technologies to provide safe drinking water.

The original protocol for selection of treated villages was random, which should have resulted in treatment and control groups which were comparable at baseline. However, we later established that the project director at the time, who was later removed from the project for unrelated reasons, did not follow the original protocol when he implemented the division of the villages into control and study villages, and he included all villages in the southern area of Matlab in the treatment group. Villages in South Matlab have much lower access to safe drinking water than the average village in the sample, meaning that overall the treated group had significantly lower access to safe drinking water at baseline than the control group.

Table 2 confirms that this resulted in statistically significant differences between control and treatment groups. Treated villages had reported lower access to safe drinking water, and were less likely to have changed their source of drinking water because of the arsenic contamination problem in the last five years. In Table 2, we show baseline summary statistics and randomization checks for villages by treatment status. The table shows the mean and standard errors for a selection of baseline variables which measure baseline access to safe drinking water, factors that might influence the ease of providing safe drinking water, and community-level variables that might influence the likelihood of a successful collective action. In column 2), we test whether the difference in means between treated and control villages is statistically significant. The p-values are derived from Ordinary Least Squares (OLS) regressions with the following structure:

$$\bar{Y_v} = \alpha + \beta I_{t,v} + \epsilon_v \tag{1}$$

where v is a village, $\bar{Y_v}$ is the mean value of a variable in village v and $I_{t,v}$ is an indicator which is one if village v was treated and zero if village v was not treated. If the treatment was randomly assigned, the coefficient β should be zero as assignment to treatment should not be correlated with any baseline characteristics of the village. The p-values test whether the coefficient is equal to zero.

Since treatment was assigned at the village level, but we collected data at the household level, it is important to account for within-village correlation in variables. Within-village correlation implies that it is more likely that differences between mean outcomes in treated and control villages arise due to chance, than if we had been able to assign treatment at the household level. In order to ensure that the statistical analyses we carry out make the correct inference about whether or not a result is likely to be due to chance or not, we follow Angrist and Pischke (2009) and use one of two approaches. Here, we collapse the data to village-level means before carrying out the regression analysis, and use robust standard errors to allow for heteroskedasticity. Where we use the household data directly, we will cluster standard errors at the village level.

Columns 3) to 5) of Table 2 show that we can remove the bias induced by the failure of randomization by three methods. First, we can drop South Matlab from the sample. Second, we can create a synthetic treatment variable generated at random in South Matlab, and equal to the treatment variable elsewhere [8]. Third, we can use this synthetic treatment variable to instrument for treatment. In columns 3) to 5), we report the difference in means between treated and control villages under these three approaches, after accounting for the different proportions of treated villages in Gopalganj and Matlab, because differences in treatment and control groups would otherwise reflect differences between these areas. . We estimate the difference in means using an equation similar to Equation 1, including indicators for Gopalganj and South Matlab. In each

[8] Ideally, we would have used the original random assignment to treatment rather than this synthetic alternative but we have not been able to recover the initial, randomly assigned treatment lists.

case we show that no significant differences remain between treatment and control populations.

Since the non-random selection of treatment villages in South Matlab may have introduced bias into our estimates of treatment effects, we will therefore report both OLS results and results where treatment is instrumented using the synthetic treatment variable (the IV results).

Project staff implemented the intervention under one of three decision-making structures. The necessary decisions included if, how and where to install; and how to manage, each safe drinking water source. In all cases, project staff ensured that all decisions made were technically appropriate. Table 3 summarizes the main features of the different decision-making structures. We describe the three models in more detail in the following paragraphs.

The decision-making structures included one non-participatory structure, the Top-Down model (TD). Under this model, project staff took all project decisions, after an extended (typically 2-day) period of information gathering. The information gathering process consisted of participatory mapping of the village with members of the community, focusing on the locations of households and sources of drinking water. The information was cross-checked with other community members. Project staff then proposed sites for safe drinking water sources, prioritizing locations where the density of households not already served by safe drinking water sources was highest, choosing public locations wherever possible, and convenient locations where no suitable public land was available. Staff then organized and publicized a community meeting at which they presented the proposed locations. This model was designed to approximate the ‘’traditional’ approach to decision-making about local public goods in which decisions are taken by a centralized authority.

The decision-making structures also included two participatory structures, in which decision-making authority was devolved to the community. Under the pure Community Participation (CP) model, project staff visited the community to arrange a meeting at a site and time of the community's choosing. At the meeting, project staff explained the project rules and announced that they would return to the village after a few days to find out whether they wanted to participate in the project, and if so, which sites they had chosen. Sites that were not technically appropriate were rejected, but otherwise the community's decisions were final, conditional on raising the community contribution. We did not directly observe the decision-making process used, but communities reported to us that they took these decisions in a variety of ways including open meetings (sometimes but not always including women), meetings at a mosque, or closed-door meetings of village elites. This model was designed to approximate the way in which some organizations implement community participation in practice, avoiding interference with a community's internal hierarchies and decision-making processes .

Under the second participatory decision-making structure, the NGO-Facilitated Community Participation model (NGO), we imposed rules about how decisions should be taken. Project staff initially organized a series of separate small group meetings with men and women who the community identified as poor and non-poor. At these meetings, project staff explained the project rules and emphasised the right all individuals would have to participate in the decision-making process and benefit from the interventions. These meetings were followed by a community meeting at which project staff re-iterated the project rules. The meeting had to be attended by both men and women and poor and non-poor. The community proposed and selected water source locations by consensus at the meeting in the presence of project staff. If the community could not reach a consensus at the first meeting, a second and in some cases subsequent meetings were organized. This model was designed to approximate the way in which other organizations implement community participation, with project staff playing a strong facilitatory role, and rules imposed to avoid the co-option of the decision-making process by influential groups or individuals.

After the initial decision-making process, project staff gave the communities up to twelve weeks to raise the funds for the community contribution. Construction of the safe drinking water sources began as soon as the community had raised their contribution. If after twelve weeks the community had not raised their contribution, construction of the safe drinking water sources did not go ahead. We initially intended the decision-making structures to apply to decisions about who contributed to the community contribution, but this proved impossible to enforce. However, project staff did propose a list of contributors at the Top Down model meetings, and communities did agree a list of contributors at the NGO-Facilitated Community Participation meeting.

We randomly assigned the decision-making structures to the communities who received the intervention. Of the 126 treated villages, we assigned 42 to each decision-making model. We had initially assigned the village in which we determined before beginning the project that there was no feasible safe drinking water technology was assigned to the Top-Down model. We replaced this with another village, randomly drawn from the villages which we had initially assigned to treatment but in which we had not carried out the intervention due to budget constraints. As a result 43 villages were initially assigned to the Top-Down model.

Table 4 shows that the villages assigned to each decision-making model are comparable to the villages assigned to the other decision-making models. We test whether the difference in variable means between villages in which the project was implemented under a given decision-making structure and the remainder of the treated villages is statistically different from zero. The p-values in the table are therefore derived from OLS regressions similar in structure to Equation 1 but the indicator $I_{m,v}$ is one if village v received treatment under decision-making structure m, and zero otherwise:

$$\bar{Y_v} = \alpha + \beta I_{m,v} + \epsilon_v \tag{2}$$

Only the treated villages are included in the regressions in columns 3) to 5). We do not use the control group for comparison in this case because the results in column 2) already confirm that the treated villages are not directly comparable to the control villages.

We compare 16 variables across the 3 decision-making structures, resulting in a total of 48 tests. In 46 of these tests we fail to reject at the 10% level the null hypothesis that there is no difference in means between groups treated under one decision-making structure and the other treated villages. In 2 tests we find statistically significant differences between the mean of a variable in villages treated under one decision-making structure and in the remaining treated villages. This is approximately the same as the number (approximately 5) that we would expect to fail at this level due to chance. From these checks we conclude that there is no evidence to suspect that assignment to model, conditional on treatment, was not random, as required by the project protocol.

The same project staff – one team in Gopalganj and one team in Matlab – implemented the project under all three decision-making structures. We implemented the intervention in cycles during which project staff would complete the entire process from meeting organization to water source installation for a group of villages, where the villages were grouped geographically for ease of logistics. The project was initially implemented in 114 villages in 6 cycles across both upazilas. We later added an additional 12 villages in Gopalganj when funds became available, in a 7th cycle.

Government policy had changed by the time we carried out the 7th cycle, and community expectations that the government would provide free tubewells may have increased. We installed fewer safe water sources under the 7th cycle, but the number installed is not significantly less than under the first 6 cycles in Gopalganj, once we account for the feasible technology.

2.3 Data Description

We carried out a baseline survey in 2007 in 40 households in each of the 250 villages, sampled randomly from census lists . We surveyed a total of 9797 households, as in some very small villages there were fewer than 40 households. The baseline questionnaire included standard components of a household survey with a special focus on social networks and social capital, and full details on water use behavior. We also collected village-level information from focus groups.

We encountered significant problems with the data entry process after the baseline survey. First, some of the individuals employed to enter the data in spreadsheets copied and pasted entire villages of data, changing names and other identifiers to conceal what they had done. Data checking revealed this problem by chance several months after data collection and entry had been carried out. When we discovered this problem, we checked extensively for additional incidences and had the missing data re-entered. Second, by the time we discovered this problem, termites had unfortunately attacked the stored questionnaires, and destroyed a small percentage of the questionnaires. As result, we are missing baseline data from 140 households from control and treated villages, since enumerators did not initially enter the data correctly and termites then destroyed the hard copy of the questionnaires. We do not however have any reason to think that there was any systematic pattern to either the false data entry or the losses to termites, so the remaining baseline data should still represent a randomly selected sample of the baseline population.

We carried out follow-up surveys in control and treated villages in 2010 and 2011 after we carried out the safe water intervention, interviewing the same households that we interviewed for the baseline survey. We did not carry out follow-up surveys in the 24 villages which were initially assigned to treatment but in which we did not carry out the intervention. We therefore attempted to resurvey 8,890 households from the original panel, of which we successfully re-surveyed 8630 households, representing an average attrition rate of 2.9%. The attrition rates broken down by treatment group are as follows: 2.7% in control villages; 3.1% in treated villages; 2.6% in NGO-Facilitated Community Participation villages; 3.2% in Community Participation villages; 3.4% in

Top-Down villages. The attrition rates in treated groups and sub-groups were not statistically different from the control group, or from each other.

We also carried out follow-up surveys in 1424 additional households in treated villages, to bring the minimum survey coverage up to 15% of households in all treated villages (based on census data). The additional households were randomly selected from the remaining households on the census lists who had not been surveyed at baseline. Extending the survey coverage in this way was intended to ensure that the survey accurately captured the effects of the intervention in larger villages, where the three safe drinking water sources constructed were unlikely to serve the entire community. In these additional households, we asked questions intended to help us recover information about their circumstances at baseline, as well as the questions asked in all other households at followup.

We also collated data on the numbers and types of safe drinking water sources installed, and details of the implementation process, including the number of contributors in each community and the time taken to raise the community contribution. We also carried out focus group discussions in treatment villages to obtain qualitative information about why the project was successful in some communities and not in others.

3 Results

We will first describe the project outcomes in the study villages and compare the numbers of safe water sources installed, the numbers of safe water sources installed in public places and the number of individuals contributing money towards the cost of the water source. We describe in the text these outcomes in the villages in which AIRPs were the only feasible technology, and report the results both for all treated villages, and for villages in which tubewells were feasible only.

Since the project outcomes vary strongly by whether tubewells were feasible or whether AIRPs were the only feasible technology, we use geographical variation in geology and propensity score matching to create a matched control group for villages in Gopalganj in which tubewells were feasible, and villages in which only AIRPs were feasible. We then report the average treatment effect in terms of changes in reported access to safe drinking water for all villages, villages in which tubewells were feasible and villages in which only AIRPs were feasible, using the matched or full control group where appropriate.

We then report how the treatment effect varied by model for all villages, and for villages in which tubewells were feasible.

3.1 Project Outcomes

Table 5 shows how the decision-making model influenced project outcomes in the treated villages. On average, we installed 2.14 safe water sources in the treated villages. If we had installed all technically feasible water sources given our project rules, we would have installed an average of 2.75 safe water sources.

We offered communities the choice between all locally technically appropriate technologies to provide safe drinking water. In Gopalganj, we carried out the intervention in 70 villages. In 16 villages, AIRPs were the only feasible technology. In two villages, no treatment was feasible, as there was a layer of impenetrable rock, and shallow groundwater was too strongly contaminated with arsenic and iron for removal with an AIRP. In Matlab, tubewells were feasible in all villages.

A clear preference gradient between the available technologies emerged. Shallow tubewells are the most preferred option, followed by standard deep tubewells and deepset tubewells, which are all similar in terms of use and maintenance but increasing in cost. AIRPs were the least preferred option by a significant margin. There were 16 villages where the only type of water source that could be installed was an AIRP, meaning that we could have installed a total of 32 AIRPs. We were only successful in installing 5 AIRPs during the course of the project, a success rate of approximately 16%. In comparison, in the villages in Gopalganj in which tubewells were feasible, we installed 79% of the maximum number of wells we could have installed under our project rules. The reasons given by the communities for rejection of the AIRPs were that they took up too much space, required too much work to operate and maintain, and were not perceived to be reliable or trustworthy. When we consider only the villages in which tubewells were feasible, the average number of water sources constructed rises to 2.45 out of a maximum possible 2.85.

The rejection of AIRPs did not seem to be a direct function of the price of the technology, although we do not know what would have happened if we had offered AIRPs at a lower price. However, in Matlab, in the 10 villages where only deep-set tubewells could be installed (for which we required the same level of community contribution as for the AIRPs), we installed on average 90% of the maximum feasible number of water sources, compared to an average of 89% in all other villages in Matlab (where either deep tubewells or shallow tubewells were feasible).

We installed more water sources in the villages in which communities participated in decision-making than in the villages in which project staff took decisions, as shown in column 1). Installing more water sources is one measure of success of the project, but it may not translate into increased access to safe drinking water if the sources are not fully accessible to the community. However, the differences are not statistically significant. In Table 5, we assess whether differences in project outcomes across models are statistically significant using OLS regression for the following equation:

$$Y_v = \beta_{NGO} I_{NGO,v} + \beta_{CP} I_{CP,v} + \beta_{TD} I_{TD,v} + \gamma' Z_v + \epsilon_v \tag{3}$$

where Z_v is a vector of village control variables including an indicator for the upazila, upazila-specific flexible controls for village size and controls for the best available technology. We then test pairwise equality of the coefficients β_{NGO}, β_{CP} and β_{TD}. The differences between the number of water sources installed under the different decision-making models are attenuated further when we consider only the villages in which tubewells were feasible.

We installed more water sources in public spaces, as recorded by our project staff, under the non-participatory Top-Down model. Public spaces included communal land, open spaces, areas beside roads, and institutions such as mosques or schools, as opposed to privately owned land. Under this model, project staff had a specific mandate to install water sources in public places. The differences are strongly significant with respect to both the participatory decision-making models. Water sources installed in public places may be accessible to a larger number of people. However, space that is not vulnerable to flooding is quite strongly constrained in the study villages. Safe water sources cannot be installed on land that is vulnerable to flooding because of potential contamination. The most convenient location for a water source may not necessarily be located on public land.

Overall, the number of contributors was relatively low in all cases, considering that the median village size was 170 households. In villages where we successfully installed at least one safe water source, the mean number of contributors per water source installed was 5.1 in NGO villages, 2.3 in CP villages and 4.0 in TD villages. There was only one contributor per safe water source installed in 34% of the NGO villages, 56% of the CP villages and 45 % of the TD villages.

Fewer people contributed to raising the community contribution in the unregulated Community Participatory model than under the other two models. The difference is significant with respect to both the other two decision-making models. A small number of contributors may be efficient, as some community members will have a greater ability to contribute than others. However, it may also be indicative of a high degree of influence over the decision-making procedure, which may not be efficient if used to co-opt project benefits for private use.

3.2 Reported Project Impact

We measure access to safe drinking water based on an outcome variable which measures whether or not the household reports using safe drinking water. The indicator is based on the source of water that the household identifies as being its most important source of water for drinking and cooking. The indicator for reporting use of safe drinking water is constructed as being equal to one where the household reports using a source of drinking water that is considered safe from both bacterial and arsenic contamination, and zero when they report using a source that is either considered unsafe, if they report that it is unsafe, or if they report that they don't know its safety status. Further details regarding the construction of this variable is included in Appendix A.

We will focus on differences between baseline and follow-up in our estimates of impact on reported access to safe drinking water. At baseline, there is substantial variation across villages in terms of reported access to safe drinking water. The magnitude of differences between treatment groups is quite large with respect to the treatment effects we estimate. In particular, 36% of households in NGO Facilitated Community Participation villages report having access to safe drinking water at baseline in comparison to 41% in Community Participation villages and 44% in Top Down villages.

We report the average overall treatment, relative to the full control group, but we also break down the treatment effect by whether tubewells or only AIRPs were feasible. In these cases, we use a matched control group, because we do not observe which technologies are feasible in the control villages. There is strong spatial correlation between locations where only AIRPs are feasible, reflecting the extent of the rock layer overlaying the deep aquifer. Since other village level characteristics are also spatially correlated, there are as a result significant differences on some baseline characteristics between villages in which tubewells were feasible and villages in which only AIRPs were feasible in Gopalganj.

We exploit this spatial correlation to create a matched control group to the villages in which AIRPs were feasible, and a matched control group to the villages in which tubewells are feasible. Details of the construction of the matched control group are given in Appendix A and Appendix Table B1 shows that there are no statistically significant differences between treated villages in which tubewells are feasible and their matched control villages, and that the same is true for villages in which only AIRPs were feasible.

Average treatment effect

In Table 6 we show the average treatment effect across all villages in columns 1) and 2); in all villages in which tubewells were feasible in columns 3) and 4); and in those villages where only AIRPs were feasible in column 5). The OLS and IV results are almost identical, suggesting that although assignment to treatment was not random in all areas, it was not correlated with trends in reported access to safe drinking water. The estimated average treatment effect is 16.3% overall, and 18.3% in villages in which tubewells were feasible. There was no significant treatment effect in the AIRP villages.

To estimate the treatment effects, we use data from all households for which we have baseline and follow-up data and estimate a first difference equation as follows:

$$\Delta Y_i = Y_{if} - Y_{ib} = \alpha + \beta I_{t,v} + \epsilon_i \tag{4}$$

where i is a household and ΔY_i is the change in access to safe drinking water between baseline and followup. With two time periods, the first difference analysis is directly equivalent to including household fixed effects. As before, we cluster standard errors at the village level to account for within-village correlation in outcomes.

For these results and the remainder of the results in this section, we use survey weights which ensure that each village counts equally in the analysis. Where part of the data for a village was lost through the baseline data entry problems, the baseline weights compensate for these losses, as there is no reason to think that the lost data introduces any bias to the estimates of a variable in the village. We do not introduce compensatory weights for migration, but attrition rates were low overall.

This analysis does not make full use of the data that we collected at followup, because we collected additional data from households in large villages at followup. However, we have detected statistically significant differences in outcomes in the households we added at follow-up to those which were included in the baseline survey. Taking into consideration the fact that villages in which additional households were added were different to other villages, the difference between households added to the survey at followup and households from the original baseline survey is 9.3%, with a standard error of 3.9%, a difference which is statistically significant from zero (p-value 0.019). We also find that this difference is greatest and only statistically significant in NGO model villages where the difference between additional households and panel households is 23.0% (standard error 8.0%) as compared to 1.4% (standard error 5.3%) in CP model villages and 7.3% (standard error 6.2%) in TD model villages. At present we have yet to establish whether these differences are the result of a problem in the way additional households were added to the survey at followup, or a result of changes in households' survey response behaviour contingent on whether or not they were surveyed at baseline (as in Zwane et al. (2011)). Note that these households were surveyed at the same time as the main followup survey i.e. before differences between the models emerged.

In Appendix Table B2 we show the results from analyses which use this additional data first by treating the household observations as a repeated cross section with village fixed effects, and then by using all the available data to estimate village-level means at baseline and followup before estimating a first difference equation. Including the additional households increases the estimated average treatment effect by approximately 1.5%. However, as a result of this uncertainty, we focus in this paper on the first difference results using only the households surveyed at baseline and followup, as the more conservative specification.

Treatment effect by decision-making model

In Table 4 we showed that conditional on treatment, villages assigned to different decision-making models were comparable on baseline statistics. In Table 7 we show in more detail that the groups treated under each of the decision-making models are different from the control group at baseline in terms of reported access to safe drinking water, but are not statistically distinguishable from each other. The results use equations with the following structure:

$$Y_{b,i} = \alpha + \beta_{NGO} I_{NGO,v} + \beta_{CP} I_{CP,v} + \beta_{TD} I_{TD,v} + \epsilon_i \quad (5)$$

The sample consists of all households for which we have baseline data available.

In column 2), we instrument for treatment with the synthetic treatment variable using interactions between the model indicators and the synthetic treatment variable. This removes statistically significant differences between the individual treatment groups and the control group, but under the instrumented results, there are now statistically significant differences at baseline between the TD and NGO model villages. This suggest that the IV specification may give us a less biased overall estimate of the treatment effect under each model, since it compensates for the non-random assignment to treatment, but that the OLS specification may give us a more reliable estimate of the differences between models.

In columns 3) and 4) we repeat the analysis for the villages in which tubewells were feasible. These results show that this does not alter the baseline comparisons between decision-making models set out in columns 1) and 2). Table 4 showed that there were no statistically significant differences between the number of AIRP villages assigned to each model.

We find that the reported increase in safe drinking water is greatest in NGO model villages, and the reported increase in safe drinking water is almost exactly equivalent in TD and CP model villages. In Table 8, we show the results from a first difference regression of the change in reported access to safe drinking water using the following equation:

$$\Delta Y_i = Y_{if} - Y_{ib} = \alpha + \beta_{NGO} I_{NGO,v} + \beta_{CP} I_{CP,v} + \beta_{TD} I_{TD,v} + \epsilon_i \quad (6)$$

In columns 1) and 2) we show the results from the full sample of treated and control villages. In columns 3) and 4) we focus on the villages in which tubewells were feasible. The motivation for excluding the villages in which either only AIRPs were feasible, or no feasible technology was available, is that the treatment effect is zero in all these villages. Including these villages therefore reduces all the estimated model-specific treatment effects, making it more difficult to distinguish between them, and introduces noise that is not informative with respect to a comparison between the decision-making models. In columns 2) and 4) we instrument for the model assigned with the interaction between a model dummy and the synthetic treatment variable. The difference in coefficients between columns 1) and 3) and columns 2) and 4) is small. There is a slight increase in the standard errors associated with the IV estimates resulting from the two stage estimating procedure. The difference between the NGO model villages and the other treated villages is consistent across all specifications and is statistically significant in column 3), when we focus on villages in which tubewells were feasible.

Including data from the additional households surveyed at followup increases the magnitude and the statistical significance of the difference between the NGO model villages and the other treated villages, but we do not report these results in the main body of the paper as we have yet to determine whether the data from the additional households can be used with confidence. In Appendix Tables B3 and B4 we show a full set of results which includes similar analyses using the data from the additional households.

4 Conclusions

This study has demonstrated that delegating decision-making authorities to communities in projects to provide safe drinking water has the potential to improve projects in terms of outcomes and reported impact. In villages where we implemented the project under a participatory decision-making structure (the NGO Facilitated Community Participation model), we installed a slightly larger number of safe drinking water sources (0.2 more sources) but obtained an 8% higher increase in access to safe drinking water, than under a non-participatory decision-making structure (the Top-Down model). These results are broadly consistent with evidence accumulated in the past through practitioner's experience and cross-sectional analysis, but this is the first time that experimental evidence has been available to test the hypothesis that participation in decision-making has a positive impact on the result of social programs.

However, the study also provides the first experimental evidence to indicate that these benefits may not be realised if protective measures are not put in place to prevent the decision-making process from being co-opted by influential groups or individuals. Under the 'pure' Community Participation model, under which communities took decisions without imposed rules, we installed the same number of sources as under the NGO-Facilitated Community Participation model, but we obtained a 6% smaller increase in access to safe drinking water.

Since we did not test alternative strategies for preventing the decision-making process from co-option, we cannot comment as to whether the method used here (imposing the requirement that decisions be taken by unanimous consensus at a community meeting where all groups were represented and conducting small group meetings beforehand to raise awareness about the project objectives and the rights of all individuals to participate) was the most effective possible in the context. We also did not delegate technical decision-making authority to the community (our project staff determined the feasibility of any given technology and location) and therefore cannot determine whether the results would be the same or different if decision-making authority is delegated to the community over other types of decisions.

A potential weakness of our results is that we rely on reported data, and it is possible that participation in project decision-making may influence the way in which intended beneficiaries report project outcomes. We have also collected data on actual use of the installed water sources by monitoring their use directly using enumerator observations. This data is currently being analysed.

The role of the community contribution appears key in determining outcomes. The number of contributors is low over all. Those that can contribute towards the cost of the water source may have significant influence over the decision-making process. The number of contributors is lowest in the pure Community Participation villages, where we find suggestive evidence of a higher degree of elite capture. Anecdotally, project staff reported to us that in some Top-Down model villages where community groups failed to raise the community contribution, individuals volunteered to pay the community contribution, but only if the water source was installed on their private land.

The result of delegating decision-making authority to the community may vary substantially depending on the local context, for example depending on existing inequalities within the community or on the size and homogeneity of the group to which authority is delegated. We cannot determine whether the results of this study would be applicable in other contexts. The study would benefit from replication in different social and cultural contexts.

Bearing these caveats in mind, the results provide important experimental evidence regarding an influential policy recommendation, and suggest that careful consideration should be given to the structure of a participatory decision-making process, if the potential benefits are to be realized.

References

Angrist, J. D. & Pischke, J. (2009). *Mostly harmless econometrics*. Princeton University Press.

Beath, A., Christia, F., & Enikolopov, R. (2013, January). *Direct democracy and resource allocation: Experimental evidence from Afghanistan*. Centre for Economic and Financial Research at New Economic School Working Paper No 192.

Fritzen, S. (2007). Can the design of community-driven development reduce the risk of elite capture? Evidence from Indonesia. *World Development*, *35*(8), 1359–1375.

Hossain, M. A., Sengupta, M. K., Ahamed, S., Rahman, M. M., Mondal, D., Lodh, D., ... Chakraborti, D. (2005). Ineffectiveness and poor reliability of arsenic removal plants in West Bengal, India. *Environmnetal Science Technology*, *39*(11), 4300–4306.

Isham, J., Narayan, D., & Pritchett, L. (1995). Does participation improve performance? Establishing causality with subjective data. *World Bank Economic Review*, *9*(2), 175–2000.

Khwaja, A. (2004). Is increasing community participation always a good thing? *Journal of the European Economic Association: Papers and Proceedings of the Eighteenth Annual Congress of the European Economic Association*, *2*(2/3), 427–436.

Kleemeier, E. (2000). The impact of participation on sustainability: An analysis of the Malawi rural piped scheme program. *World Development*, *28*(5), 929–944.

Mansuri, G. & Rao, V. (2013). *Localizing development: Does participation work?* Washington, DC: World Bank.

Olken, B. A. (2010). Direct democracy and local public goods: Evidence from a field experiment in Indonesia. *American Political Science Review*, *104*(2).

Rao, V. & Ibáñez, A. M. (2005). The social impact of social funds in Jamaica: A 'participatory econometric' analysis of targeting, collective action, and participation in community-driven development. *Journal of Development Studies*, *41*(5), 788–838.

Sara, J. & Katz, T. (1997). *Making rural water sustainable: Report on the impact of project rules.* UNDP – World Bank Water and Sanitation Program.

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.

Stiglitz, J. (2002). Participation and development: Perspectives from the comprehensive development paradigm. *Review of Development Economics*, *6*(2), 163–182.

UNICEF. (2008, October). *Arsenic mitigation in bangladesh*.

World Bank. (2004). *World Development Report 2004: Making services work for poor people*. Washington, DC: The World Bank.

Zwane, A. P., Zinman, J., Dusen, E. V., Pariente, W., Null, C., Miguel, E., ... Banerjee, A. (2011). Being surveyed can change later behavior and related parameter estimates. *PNAS*, *108*(5), 1821–1826.

Tables and Figures

Table 1: Technologies to provide arsenic-safe drinking water

Technology	Cost	Required community contribution per safe water source installed		
		1	2	3
Deep tubewell	50000	4500	6000	7500
Shallow tubewell	20000	3000	3500	4000
Arsenic-Iron Removal Plant	60000	6000	7500	N/A
Deep-set tubewell	60000	6000	7500	N/A

Note: All prices in Bangladeshi Taka. 1 US$≈ 80BDT.

Table 2: Baseline Summary Statistics and Randomization Checks

		Control (1)	Treated (2)	Treatment - Control (3)	(4)	(5)
Gopalganj	Mean	0.51	0.55			
	s.e.	0.05	0.04			
	p-value		0.492			
South Matlab	Mean	0.0	0.23***			
	s.e.	0.0	0.04			
	p-value		0.000			
No of households in village	Mean	245	221	-26	-18	-20
	s.e.	21	17	29	26	30
	p-value		0.382	0.372	0.512	0.502
% of water sources arsenic contaminated	Mean	0.96	0.95	0.01	0.0	0.0
	s.e.	0.01	0.01	0.01	0.01	0.01
	p-value		0.768	0.531	0.669	0.666
Reports using arsenic safe water	Mean	0.55	0.40***	-0.01	0.0	0.0
	s.e.	0.04	0.03	0.04	0.03	0.04
	p-value		0.003	0.787	0.930	0.929
Changed source of drinking water due to arsenic in last 5 years?	Mean	0.49	0.35***	0.0	0.01	0.01
	s.e.	0.04	0.03	0.03	0.03	0.04
	p-value		0.003	0.956	0.761	0.759
Anyone in household has symptoms of arsenic poisoning?	Mean	0.009	0.009	-0.001	-0.001	-0.001
	s.e.	0.002	0.001	0.003	0.002	0.003
	p-value		0.998	0.750	0.599	0.596
Total value of household assets	Mean	572053	541059	-14356	-38550	-44372
	s.e.	30542	21633	41382	37588	43045
	p-value		0.408	0.729	0.306	0.303
Access to electricity?	Mean	0.46	0.39	-0.05	-0.07	-0.08
	s.e.	0.03	0.03	0.05	0.04	0.05
	p-value		0.117	0.326	0.119	0.116
Household head literate	Mean	0.608	0.599	0.007	-0.003	-0.004
	s.e.	0.02	0.02	0.03	0.03	0.03
	p-value		0.706	0.830	0.905	0.904
Household head Muslim	Mean	0.70	0.70	0.04	0.01	0.01
	s.e.	0.04	0.04	0.05	0.05	0.06
	p-value		0.956	0.416	0.888	0.887
Household head farmer	Mean	0.42	0.45	0.03	0.02	0.02
	s.e.	0.02	0.01	0.02	0.02	0.02
	p-value		0.152	0.189	0.346	0.340
Number of associations in community	Mean	6.24	6.29	-0.19	-0.17	-0.20
	s.e.	0.14	0.15	0.22	0.19	0.22
	p-value		0.822	0.380	0.372	0.367
Number of collective actions in community	Mean	0.89	0.96	0.05	0.04	0.05
	s.e.	0.08	0.09	0.05	0.06	0.06
	p-value		0.574	0.279	0.470	0.465
	N	99	127	197	226	226

Note: P-values test significance of difference in village-level means between treated and control villages, controlling for the different treatment proportions in Gopalganj and Matlab, in columns 3-5). Data in rows 1) and 2) come from project records. Data in rows 3) and 4) comes from data from the Bangladesh Arsenic Mitigation Water Supply Project. All other data is from household surveys. N is the number of observations in the respective group in columns 1) and 2) and the number of observations used to estimate the difference in columns 3) - 5). Two villages are missing all baseline data as a result of the data entry and termite losses. Standard errors are robust.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Decision-making structures

Non-participatory	Top Down (TD)	Project staff took all project decisions, after an extended (typically 2-day) period of information gathering, using the following criteria to decide water source location: public/convenient location, population density, existing safe water options.
Participatory	Community Participation (CP)	The community took all project decisions using their own (unobserved) decision-making structures, following a community-wide information meeting led by project staff.
	NGO-Facilitated Community Participation (CP)	The community took all project decisions at a community-wide meeting, following smaller information meetings for different groups. We imposed two decision-making rules. If decisions made did not satisfy these rules, project staff did not implement the decisions: • Attendance at the community meeting had to include: at least 10 men, of which 5 had to qualify as poor; and at least 10 women, of which 5 had to qualify as poor. • Decisions had to be unanimous.

Table 4: Baseline Summary Statistics and Randomization Checks

		NGO (1)	CP (2)	TD (3)
Gopalganj	Mean	0.55	0.56	0.56
	s.e.	0.08	0.08	0.08
	p-value	0.900	0.933	0.967
South Matlab	Mean	0.24	0.22	0.23
	s.e.	0.07	0.06	0.06
	p-value	0.882	0.843	0.964
No of households in village	Mean	213	210	236
	s.e.	33	24	32
	p-value	0.789	0.665	0.518
% of water sources arsenic contaminated	Mean	0.96	0.95	0.95
	s.e.	0.01	0.01	0.01
	p-value	0.461	0.700	0.733
AIRP	Mean	0.10	0.15	0.14
	s.e.	0.05	0.06	0.05
	p-value	0.425	0.663	0.767
Reports using arsenic safe water	Mean	0.36	0.41	0.44
	s.e.	0.05	0.05	0.05
	p-value	0.255	0.821	0.366
Changed source of drinking water due to arsenic in last 5 years?	Mean	0.32	0.35	0.37
	s.e.	0.05	0.05	0.05
	p-value	0.499	0.929	0.561
Anyone in household has symptoms of arsenic poisoning?	Mean	0.004**	0.009	0.012*
	s.e.	0.002	0.003	0.003
	p-value	0.012	0.803	0.088
Total value of household assets	Mean	544360	547704	531500
	s.e.	39772	41943	30342
	p-value	0.917	0.840	0.730
Access to electricity?	Mean	0.37	0.39	0.42
	s.e.	0.05	0.05	0.05
	p-value	0.590	0.936	0.537
Household head literate	Mean	0.60	0.58	0.62
	s.e.	0.03	0.03	0.02
	p-value	0.894	0.412	0.258
Household head Muslim	Mean	0.69	0.70	0.73
	s.e.	0.07	0.06	0.06
	p-value	0.735	0.921	0.656
Household head farmer	Mean	0.44	0.46	0.44
	s.e.	0.03	0.02	0.02
	p-value	0.897	0.491	0.616
Number of associations in community	Mean	6.35	6.04	6.45
	s.e.	0.25	0.19	0.31
	p-value	0.755	0.189	0.471
Number of collective actions in community	Mean	0.91	1.00	0.97
	s.e.	0.14	0.16	0.15
	p-value	0.650	0.760	0.903
	N	42	41	43

Note: P-values test significance of the difference between model and other treated villages. Data from household surveys except rows 1), 2) and 5) which come from project records and rows 3) and 4) which come from the Bangladesh Arsenic Mitigation Water Supply Project. Baseline data for one CP village is missing as a result of the data entry and termite losses. Standard errors robust. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Project Outcomes

			Outcome Variable		
			Water sources installed	Water sources installed in public places	Number of contributors
			(1)	(2)	(3)
A: All villages	All treated	Mean	2.14	1.38	7.81
		s.e.	0.10	0.10	0.83
	NGO	Mean	2.21	1.19	9.37
		s.e.	0.16	0.14	1.62
	CP	Mean	2.21	1.00	5.40
		s.e.	0.17	0.14	0.93
	TD	Mean	2.00	1.93	8.67
		s.e.	0.18	0.18	1.63
	NGO = CP	p-value	0.712	0.568	0.048**
	CP = TD	p-value	0.205	0.000***	0.024**
	TD = NGO	p-value	0.344	0.000***	0.675
		N	127	127	126
B: Villages where tubewells feasible	All treated	Mean	2.45	1.58	9.02
		s.e.	0.08	0.10	0.92
	NGO	Mean	2.46	1.30	10.61
		s.e.	0.13	0.14	1.74
	CP	Mean	2.53	1.14	6.11
		s.e.	0.14	0.15	1.03
	TD	Mean	2.36	2.31	10.33
		s.e.	0.16	0.15	1.82
	NGO = CP	p-value	0.682	0.755	0.068*
	CP = TD	p-value	0.292	0.000***	0.025**
	TD = NGO	p-value	0.458	0.000***	0.521
		N	109	109	108

Note: P-values test pairwise significance of the difference between the means across two models. They are derived from a regression of the outcome variable on indicators for the three types of treatment (with no constant) and controls (indicators for small and large villages, a Gopalganj dummy, and interactions between the two; and indicators for the best available technology). Standard errors are robust. In villages where no water sources were installed, the number of contributors is coded as zero. In one village, the number of contributors was not recorded.
*** p<0.01, ** p<0.05, * p<0.1.

Table 6: Change in Reported Access to Safe Drinking Water: Treatment versus Control

		Dependent Variable: Change in reported use of safe drinking water				
		OLS	IV	OLS	IV	OLS
		All villages	All villages	Tubewell villages	Tubewell villages	AIRP Villages
		(1)	(2)	(3)	(4)	(5)
Treated	Coefficient	0.164***	0.163***	0.182***	0.183***	0.004
	s.e.	0.03	0.04	0.03	0.04	0.09
	p-value	0.000	0.000	0.000	0.000	0.962
Constant	Coefficient	-0.01	-0.02	0.00	-0.02	-0.04
	s.e.	0.02	0.02	0.02	0.02	0.05
Controls	Gopalganj?		Yes		Yes	
	South Matlab?		Yes		Yes	
Control villages		All	All	Matched	Matched	Matched
First stage F-test			1278		860	
N		8427	8427	7154	7154	1200

Note: Treatment is instrumented using synthetic assignment to treatment in South Matlab in columns 2) and 4). The change is estimated using first differences at the household level. Survey weights are applied so that each village counts equally in the analysis. Standard errors are robust and clustered at the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Baseline Reported Access to Safe Drinking Water: Decision-Making Models

		Dependent Variable: Reported use of safe drinking water at baseline			
		OLS	IV	OLS	IV
		All villages	All villages	Tubewell villages	No Tubewell villages
		(1)	(2)	(3)	(4)
NGO	Coefficient	-0.19***	-0.06	-0.24***	-0.08
	s.e.	0.06	0.05	0.07	0.06
	p-value	0.003	0.222	0.001	0.154
CP	Coefficient	-0.13**	0.01	-0.17**	0.00
	s.e.	0.07	0.06	0.07	0.06
	p-value	0.046	0.830	0.013	0.954
TD	Coefficient	-0.10	0.04	-0.13*	0.06
	s.e.	0.06	0.05	0.07	0.06
	p-value	0.106	0.400	0.067	0.316
Constant	Coefficient	0.55	0.83	0.61	0.83
	s.e.	0.04	0.03	0.04	0.03
NGO = CP	p-value	0.437	0.227	0.395	0.196
CP = TD	p-value	0.708	0.631	0.563	0.428
TD = NGO	p-value	0.239	0.063	0.151	0.027
NGO = pooled	p-value	0.258	0.076	0.185	0.042
CP = pooled	p-value	0.819	0.679	0.878	0.783
TD = pooled	p-value	0.372	0.180	0.246	0.087
	N	8695	8695	7375	7375
Controls	Gopalganj	No	Yes	No	Yes
	South Matlab	No	Yes	No	Yes
Control villages		All	All	Matched	Matched

Note: Model is instrumented using model interacted with synthetic assignment to treatment in South Matlab in columns 2) and 4). In columns 3) and 4) the control group is matched to the subset of treated villages using baseline propensity score matching. Standard errors are robust and clustered at the village level. The sample consists of all households for which we have baseline data with baseline survey weights.
*** p<0.01, ** p<0.05, * p<0.1.

Table 8: Change in Reported Access to Safe Drinking Water: Decision-Making Models

		Dependent Variable: Change in reported use of safe drinking water			
		OLS	IV	OLS	IV
		All villages	All villages	Tubewell villages	Tubewell villages
		(1)	(2)	(3)	(4)
NGO	Coefficient	0.22***	0.21***	0.25***	0.24***
	s.e.	0.05	0.06	0.05	0.06
	p-value	0.000	0.000	0.000	0.000
CP	Coefficient	0.14***	0.15***	0.15***	0.16***
	s.e.	0.04	0.05	0.04	0.05
	p-value	0.002	0.001	0.000	0.000
TD	Coefficient	0.13***	0.13***	0.15***	0.14***
	s.e.	0.04	0.05	0.04	0.05
	p-value	0.002	0.010	0.001	0.008
Constant	Coefficient	-0.01	-0.02	0.00	-0.02
	s.e.	0.02	0.02	0.02	0.02
NGO = CP	p-value	0.178	0.296	0.116	0.205
CP = TD	p-value	0.946	0.693	0.954	0.662
TD = NGO	p-value	0.151	0.168	0.109	0.117
NGO = pooled	p-value	0.119	0.174	0.078	0.113
CP = pooled	p-value	0.429	0.672	0.338	0.582
TD = pooled	p-value	0.356	0.286	0.298	0.225
	N	8427	8427	7154	7154
Controls	Gopalganj	No	Yes	No	Yes
	South Matlab	No	Yes	No	Yes
Control villages		All	All	Matched	Matched

Note: Model is instrumented using model interacted with synthetic assignment to treatment in South Matlab in columns 2) and 4). The treatment effect under each model is estimated by regressing the change in reported access to safe drinking water on indicators for treatment under each of the three decision-making models. The sample consists of all households for which we have baseline and followup data with baseline survey weights. Standard errors are robust and clustered at the village level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendices

Appendix A: Data

Variable Construction

We asked the households to list all the sources of water they used for drinking and cooking. In the analysis, we focus on the most important source of water for drinking and cooking, which we asked households to list first. We also asked households to report the percentage of water for drinking and cooking that they obtained from each source, but results based on the source from which households report drawing the largest percentage of water are unstable between baseline and followup, whereas the results based on the first-listed water source are more consistent. This may be attributable to slight differences in the way in which the question was asked as to whether the question referred to water used for drinking only or drinking and cooking.

Reports using safe drinking water If the household reports using a tubewell, we code the household as reporting using safe water if they report that the source is arsenic-safe, and reporting unsafe water if it is unsafe or if they dont know the source's safety. If the household reports using an unsafe source with respect to bacterial contamination (i.e. a dug well or surface water), we code the household as reporting using unsafe water. Some sources can be presumed to be safe from both bacterial and arsenic contamination (e.g. AIRPs, PSF, rainwater, deep-set tubewells). In these cases, we code the household as reporting using safe water unless they specifically report that the water is unsafe. The numbers of households using these sources is small. If they report using any other source, we code the household as reporting using safe water if they report that the source is safe, and reporting unsafe water if it is unsafe or if they dont know the source's safety status.

Construction of matched control groups in Gopalganj

In Gopalganj, there are 18 unions (the smallest rural administrative and local government units in Bangladesh). In three of these unions, only AIRPs were feasible in all treated villages. We assign the five control villages in these unions to the AIRP-matched control group. In five unions, tubewells were feasible in all treated villages. We assign the 21 control villages in these unions to the tubewell-matched control group. For the remainder of the unions, we construct a propensity score index by running a logit regression of an indicator for whether tubewells (or only AIRPs) were feasible on a set of baseline characteristics which we observe for both treated and control groups and which predict feasible technology in the treatment group. We then assign the remaining villages to the relevant matched control group where their propensity score is greater than 0.5 (although in reality the propensity scores are strongly clustered around 0 and 1).

Appendix Table B1 shows a comparison between villages in which the only feasible technology was the arsenic-iron removal plant (AIRP) or in which there was no feasible safe water technology, and villages in which deep tubewells were feasible. The comparison is limited to Gopalganj, as tubewells were technically feasible in all villages in Matlab. Columns 1) and 2) confirms that overall in Gopalganj there is no evidence to suggest that assignment to treatment was not random, as originally intended. Only one of the 12 tests comparing treated to control villages in Gopalganj shows statistically significant differences at the 10% level, which is approximately what we would expect to see due to chance. The villages in which only AIRPs were feasible, and villages in which tubewells were feasible, have statistically significant baseline differences to the control groups in some respects. We do not report the p-values in the table, but we observe significant differences relative to the control group in household assets, access to electricity and having a farmer as the household head. Columns 3) - 6) show that there are no statistically significant differences between the treated villages in which a particular technology was feasible and their matched control villages.

Details of analysis using additional households

The analysis in columns 1) and 2) of Table B2 does not make full use of the data that we collected at followup, because we collected additional data from households in large villages at followup. We do not have baseline data for these households. In columns 3) and 4) we treat the data as a repeated cross section and estimate the following model.

$$Y_{it} = \alpha + \theta_v + \gamma_f + \beta I_{tf} + \epsilon_i \tag{7}$$

where: i is a household and t is a time period, either b, baseline, or f, followup; θ_v is a village-level fixed effect; γ_f is a wave fixed effect to capture overall differences between baseline and followup; and I_{tf} is an indicator for whether the village was treated, interacted with an indicator for whether the observation was taken at followup. Including the data from the additional households in the followup villages slightly increases the estimated treatment effect, although the estimates in columns 3) and 4) are well within the confidence intervals from the regressions in columns 1) and 2). Note that the change in the coefficient comes almost exclusively from including the data from the additional households, rather than from changing the regression strategy; running a repeated cross-section from the same sample as 1) and 2) yields an estimated average treatment effect of 16.6%.

In columns 5) and 6) we collapse the household data at baseline and followup to village level means before running a first-difference regression with the following structure. We use all available data including the additional households surveyed at followup:

$$\Delta Y_v = Y_{vf} - Y_{vb} = \alpha + \beta I_t + \epsilon_v \tag{8}$$

The results are consistent with the results from column 3) and 4).

Appendix B: Figures and Tables

Table B1: Baseline Summary Statistics for AIRP and non-AIRP villages in Gopalganj

		Control	Treated	AIRP-matched control	AIRP	Tubewell-matched control	Tubewell
		(1)	(2)	(3)	(4)	(5)	(6)
No of households in village	Mean	257	262	268	338	255	242
	s.e.	33	26	44	71	45	27
	p-value		0.904		0.405		0.814
% of water sources arsenic contaminated	Mean	0.96	0.97	0.97	0.97	0.96	0.96
	s.e.	0.01	0.01	0.01	0.01	0.01	0.01
	p-value		0.889		0.720		0.762
Reports using arsenic safe water	Mean	0.26	0.26	0.25	0.23	0.27	0.26
	s.e.	0.04	0.03	0.07	0.07	0.04	0.04
	p-value		0.951		0.804		0.919
Changed source of drinking water due to arsenic in last 5 years?	Mean	0.20	0.21	0.21	0.22	0.18	0.19
	s.e.	0.03	0.03	0.07	0.06	0.04	0.03
	p-value		0.748		0.945		0.809
Anyone in household has symptoms of arsenic poisoning?	Mean	0.009	0.006	0.008	0.005	0.003	0.006
	s.e.	0.003	0.002	0.004	0.002	0.002	0.002
	p-value		0.341		0.489		0.378
Total value of household assets	Mean	511810	470994	532043	571037	492694	438088
	s.e.	40838	21346	82707	47413	47070	23037
	p-value		0.378		0.685		0.300
Access to electricity?	Mean	0.45	0.37	0.68	0.60	0.35	0.31
	s.e.	0.04	0.04	0.03	0.07	0.05	0.04
	p-value		0.214		0.327		0.493
Household head literate	Mean	0.56	0.58	0.49	0.54	0.62	0.59
	s.e.	0.04	0.03	0.07	0.06	0.04	0.03
	p-value		0.648		0.550		0.569
Household head Muslim	Mean	0.48	0.56	0.46	0.66	0.46	0.52
	s.e.	0.06	0.05	0.10	0.11	0.08	0.06
	p-value		0.364		0.167		0.503
Household head farmer	Mean	0.46	0.50	0.39	0.45	0.49	0.51
	s.e.	0.02	0.02	0.03	0.04	0.03	0.02
	p-value		0.101		0.180		0.565
Number of associations in community	Mean	6.86	6.74	7.17	6.65	6.79	6.75
	s.e.	0.22	0.24	0.50	0.36	0.23	0.29
	p-value		0.717		0.402		0.903
Number of collective actions in community	Mean	0.14	0.23*	0.12	0.26	0.15	0.22
	s.e.	0.02	0.04	0.04	0.08	0.03	0.05
	p-value		0.073		0.126		0.245
	N	50	70	16	16	33	52

Note: P-values test significance of the difference in means between treated and control villages for all villages in Gopalganj in column 2), for AIRP villages and matched control villages in column 4) and for tubewell villages and matched control villages in column 6. Data in rows 1) and 2) comes from data from the Bangladesh Arsenic Mitigation Water Supply Project. The remaining data comes from household surveys. Standard errors are robust.
*** p<0.01, ** p<0.05, * p<0.1.

Table B2: Change in Reported Access to Safe Drinking Water: Treatment versus Control
All Villages

		Dependent Variable: Change in reported use of safe drinking water					
		OLS	OLS	OLS	IV	IV	IV
		Panel	Panel	Repeated Cross Section	Repeated Cross Section	Village Means	Village Means
		(1)	(2)	(3)	(4)	(5)	(6)
Treated	Coefficient	0.164***	0.163***	0.178***	0.179***	0.178***	0.178***
	s.e.	0.03	0.04	0.03	0.04	0.03	0.04
	p-value	0.000	0.000	0.000	0.000	0.000	0.000
Constant	Coefficient	-0.01	-0.02	0.47	0.81	-0.02	-0.02
	s.e.	0.02	0.02	0.01	0.01	0.02	0.02
Controls	Gopalganj?		Yes		Yes		Yes
	South Matlab?		Yes		Yes		Yes
First stage F-test			1278		900		1263
N		8427	8427	18686	18686	224	224

Note: Treatment is instrumented using synthetic assignment to treatment in South Matlab in columns 2), 4) and 6). The change is estimated using first differences in columns 1) and 2) and 5) and 6) and with village-level fixed effects. In columns 1) to 4) survey weights are applied so that each village counts equally in the analysis. Standard errors are robust and clustered at the village level in columns 1) to 4).
*** p<0.01, ** p<0.05, * p<0.1.

Table B3: Change in Reported Access to Safe Drinking Water: Decision-Making Models

		Dependent Variable: Change in reported use of safe drinking water					
		OLS	IV	OLS	IV	OLS	IV
		First Difference	First Difference	Repeated Cross Section	Repeated Cross Section	First Difference	First Difference
		Households	Households	Households	Households	Villages	Villages
		(1)	(2)	(3)	(4)	(5)	(6)
NGO	Coefficient	0.22***	0.21***	0.25***	0.23***	0.25***	0.25***
	s.e.	0.05	0.06	0.05	0.05	0.05	0.06
	p-value	0.000	0.000	0.000	0.000	0.000	0.000
CP	Coefficient	0.14***	0.15***	0.14***	0.14***	0.14***	0.15***
	s.e.	0.04	0.05	0.04	0.04	0.04	0.05
	p-value	0.002	0.001	0.001	0.002	0.001	0.001
TD	Coefficient	0.13***	0.13***	0.14***	0.12**	0.14***	0.13***
	s.e.	0.04	0.05	0.04	0.05	0.04	0.05
	p-value	0.002	0.010	0.001	0.012	0.001	0.008
Constant	Coefficient	-0.01	-0.02	0.47	0.83	-0.02	-0.02
	s.e.	0.02	0.02	0.01	0.02	0.02	0.02
NGO = CP	p-value	0.178	0.296	0.061	0.160	0.061	0.114
CP = TD	p-value	0.946	0.693	1.000	0.786	0.996	0.722
TD = NGO	p-value	0.151	0.168	0.062	0.112	0.061	0.067
NGO = pooled	p-value	0.119	0.174	0.037	0.093	0.037	0.057
CP = pooled	p-value	0.429	0.672	0.245	0.476	0.242	0.428
TD = pooled	p-value	0.356	0.286	0.246	0.262	0.240	0.182
	N	8427	8427	18686	18686	224	224
Controls	Gopalganj	No	Yes	No	Yes	No	Yes
	S. Matlab	No	Yes	No	Yes	No	Yes

Note: Model is instrumented using model interacted with synthetic assignment to treatment in South Matlab in columns 2), 4) and 6). Standard errors are robust and clustered at the village level in columns 1), 3) and 5).
*** p<0.01, ** p<0.05, * p<0.1.

Table B4: Change in Reported Access to Safe Drinking Water: Decision-Making Models
No AIRP villages

		Dependent Variable: Change in reported use of safe drinking water					
		OLS	IV	OLS	IV	OLS	IV
		First Difference	First Difference	Repeated Cross Section	Repeated Cross Section	First Difference	First Difference
		Households	Households	Households	Households	Villages	Villages
		(1)	(2)	(3)	(4)	(5)	(6)
NGO	Coefficient	0.26***	0.26***	0.29***	0.28***	0.29***	0.29***
	s.e.	0.05	0.06	0.05	0.06	0.05	0.06
	p-value	0.000	0.000	0.000	0.000	0.000	0.000
CP	Coefficient	0.16***	0.18***	0.16***	0.17***	0.16***	0.18***
	s.e.	0.04	0.04	0.04	0.04	0.04	0.04
	p-value	0.000	0.000	0.000	0.000	0.000	0.000
TD	Coefficient	0.16***	0.15***	0.16***	0.16***	0.16***	0.16***
	s.e.	0.04	0.05	0.04	0.05	0.04	0.05
	p-value	0.000	0.003	0.000	0.002	0.000	0.002
Constant	Coefficient	-0.01	-0.03	0.49	0.82	-0.02	-0.03
	s.e.	0.02	0.02	0.01	0.02	0.02	0.02
NGO = CP	p-value	0.116	0.206	0.038	0.089	0.038	0.075
CP = TD	p-value	0.954	0.661	0.961	0.743	0.969	0.715
TD = NGO	p-value	0.109	0.118	0.050	0.068	0.049	0.053
NGO = pooled	p-value	0.078	0.115	0.026	0.050	0.026	0.040
CP = pooled	p-value	0.337	0.583	0.176	0.377	0.175	0.347
TD = pooled	p-value	0.298	0.227	0.224	0.191	0.216	0.156
	N	7784	7784	16902	16902	207	207
Controls	Gopalganj	No	Yes	No	Yes	No	Yes
	South Matlab	No	Yes	No	Yes	No	Yes

Note: Model is instrumented using model interacted with synthetic assignment to treatment in South Matlab in columns 2), 4) and 6). Standard errors are robust and clustered at the village level in columns 1), 3) and 5).
*** p<0.01, ** p<0.05, * p<0.1.