Working paper

International
Growth Centre

Group size and collective action

Evidence from Bangladesh

Serena Cocciolo
Ahsan Habib
Anna Tompsett

January 2019

When citing this paper, please use the title and the following reference number:
S-31402-BGD-1

DIRECTED BY

FUNDED BY

Group Size and Collective Action
Evidence from Bangladesh

Serena Cocciolo Ahasan Habib Anna Tompsett*

January 24, 2019

Abstract

The effect of group size on the ability to take collective action is theoretically ambiguous and empirically unresolved. This paper provides the first causal empirical evidence from a real-world setting of the effect of group size on collective action. Exogenous variation in group size arises from an application of Maimonides' rule, combined with a randomized controlled experiment. We find that when communities are faced with a collective action problem—to cooperate on a program of safe drinking water provision—in larger groups, per capita effort falls. Larger groups are nonetheless somewhat more successful overall in installing safe wells. However, despite the additional wells installed, larger groups achieve smaller increases in use of safe drinking water, possibly because reduced participation weakens constraints on elite capture.

---

[*] Corresponding author: anna.tompsett@ne.su.se. We thank the International Growth Centre, the Swedish Research Council and the International Initiative for Impact Evaluation (3ie) for financial support and advice. We thank NGO Forum for Public Health for their collaboration on this project, particularly Executive Director S.M.A. Rashid. We thank Sabbir Ahmed, Cristina Altomare, Tillmann von Carnap, Merve Demirel, Md. Shahadat Hossain, Jahirul Islam, Md. Tariqul Islam, Md. Rezwanul Karim, Rezaul Karim, Ranjan Kumar Mohonto, Mir Abu Raihan, Lorenzo Schirato, and Tahmid Sharif for exemplary research assistance and support.

1 Introduction

The effect of group size on the ability to take collective action is theoretically ambiguous (Banerjee, Iyer, & Somanathan, 2007). In larger groups, people may be more inclined to "free ride" on others' efforts, allowing others to do the work but hoping to share in the resulting benefits, whereas in smaller groups it may be easier to develop, monitor and enforce agreements that restrict free riding. On the other hand, since there are more individuals who might potentially contribute, larger groups might still, on average, be more successful in collective action. Previous empirical evidence on this question is limited by the fact that groups of different size may also differ with respect to other characteristics that also affect collective action. These differences potentially bias simple comparisons between collective action in groups of different sizes. This paper combines a randomized controlled experiment with a rule that generates exogenous variation in group size to provide the first causal empirical evidence from a real-world setting of the effect of group size on collective action.

The context is a program to provide safe drinking water in rural Bangladesh. We implement the program in geographically contiguous communities of between 50 and 250 households, which we refer to interchangeably as "treatment units", or simply, communities. Communities participating in the program must take several forms of collective action: agreeing on locations for new safe water sources; working together to coordinate community contributions in cash or labour; and collectively maintaining water sources after installation. If group size affects collective action, then the number of households in a community will affect project impact.

To measure the causal effect of group size on collective action, we exploit a rule we use to define "treatment unit" size. Our approach is inspired by Maimonides' Rule.[1] In the region we are working in, the lowest-level local administrative unit varies in size from fewer than 50 households to nearly 1000 households. To define treatment units of a manageable size, we treat administrative units smaller than 250 households as one treatment unit, and administrative units larger than 250 households as two treatment units, and so on and so forth at other thresholds which are multiples of 250. The resultant assignment rule creates sharply non-monotonic variation in treatment unit size as a function of administrative size, which we exploit in an instrumental variables framework.

Intuitively, our empirical approach compares communities which lie just below each threshold to communities which lie just above each threshold. Communities either side of

---

[1] As exploited by Angrist and Lavy (1999) to estimate the causal effect of class size on student outcomes.

each threshold are very similar to communities just above the threshold in all other respects. These thresholds also serve no other administrative purpose, meaning that the distribution of administrative units is smooth across these boundaries. However, communities below the first threshold have the project implemented in treatment units of 250 households, while communities just above the first threshold have the project implemented in treatment units of on average 125 households. As far as possible, we hold all project resources constant in per capita terms. The key research question is, therefore: holding the ratio between population and project resources constant, do larger groups collaborate more successfully than smaller groups?

We evaluate the effect of group size on measures of collective action and public good provision. Throughout, we draw on rich data describing program implementation as well as comprehensive objective data on program impact from a large program of household and water source quality testing. We combine the strategy described above with experimental assignment to the safe drinking water program to evaluate project impact relative to a control group who receives no intervention. We find that participation in collective action is lower in larger groups: fewer households are represented at community meetings, with representation by female household members falling particularly sharply; communities spend less time deliberating over source location; and fewer households contribute towards the costs of each installed safe drinking water source. However, larger groups are still weakly more successful at solving collective action problems on aggregate, leading to slightly greater installation rates of safe water sources. Despite these positive effects on installation, the average impact of the program declines with group size, possibly because reduced participation weakens constraints on elite capture. However, we note that our estimated effects on impact are imprecisely measured, because average impact is attenuated by low take-up in one arm of the study and by well installation failures, which occur for hydrogeological reasons that are exogenous to community actions. In future revisions, we may be able to improve precision by accounting for these attenuating effects.

This paper contributes to the broad literature on the determinants of successful collective action.[2] In particular, we contribute to the specific literature on group size and collective action. Group size is of particular interest to policymakers because—in contrast to other fixed characteristics of a community, such as ethnic heterogeneity—group size can be at least partially controlled by organizations designing development programs. The effect of group size on collective action is however, ambiguous: a rich theoretical literature[3] establishes the "group size paradox" (Esteban & Ray, 2001): smaller groups may be less susceptible to the

---

[2] See e.g. Chattopadhyay and Duflo (2004), Banerjee and Iyer (2005), and Miguel and Gugerty (2005).

[3] Dating back to Pareto (1927), Olson (1971) and Chamberlin (1974).

free rider problem, but larger groups are more likely to contain at least one individual willing to provide a public good at private expense.

Prior empirical evidence on behaviour in the field is limited because groups typically form endogenously, and most studies focus on comparisons across communities of different sizes,[4] meaning that we cannot distinguish the effects of group size from the effects of other characteristics that correlate with group size.[5] Laboratory experiments, typically on student volunteers interacting anonymously via computer programs, have provided important evidence on this question [6], helping to distinguish between pure effects of group size and diminishing marginal returns to contributions in the presence of congestion effects. However, the results from laboratory studies may not generalize to real world settings, where subjects have pre-existing social ties and can development and implement enforceable agreements. The primary contribution of this paper is to provide evidence of how real-world behaviour changes in response to exogenous changes in group size. Importantly, we expect that the empirical strategy we propose can be applied in other contexts to further develop our understanding of the relationship between group size and collective action.

Our results have direct implications for policy-makers designing development interventions that depend on collective action. In particular, they suggest that participation by community members in participatory processes is likely to fall in larger groups. Policy-makers taking decisions about group size or scale should consider the potential consequences of increasing group size on participation and impact as well as the costs of implementation.

The rest of this report proceeds as follows. Section 2 describes the context for this study, the arsenic contamination problem in rural Bangladesh, and the intervention we evaluate. Section 3 describes the study design and Section 4 describes the data. Section 5 describes the empirical approach. Section 6 describes the results and Section 7 concludes.

---

[4] e.g. Bardhan (2000) finds a negative correlation between group size and successful management of irrigation systems; Poteete and Ostrom (2002) describe an inverse U-shaped relation between group size and success in forest management, but note that group size may itself be an outcome of institutional success; Brunner and Sonstelie (2003) find that voluntary contributions from parents to schools declines with school size, but much more slowly than theory predicts.

[5] Bandiera, Barankay, and Rasul (2005) show that fruit pickers cooperate more effectively in small groups, conditional on individual and field average productivity. However, the number of pickers assigned to a field is determined by expected productivity, which varies over time. Bandiera et al. (2005) show that similar patterns do not arise in the absence of incentives to cooperate, but a spurious correlation between group size and productivity cannot be conclusively ruled out. Sawada, Kasahara, Aoyagi, Shoji, and Ueyama (2013) report effects of group size on a collective action from a context where some farmers are allocated to communities by government lottery, but the lottery only determines allocation for 30% of farmers.

[6] e.g. Isaac and Walker (1988).

2 Context

2.1 Arsenic contamination in rural Bangladesh

The context for this study is the arsenic contamination problem in rural Bangladesh, where arsenic contamination occurs naturally in shallow groundwater. Arsenic concentrations vary across space, and arsenic is not detectable in drinking water except via a chemical test. Most households in rural Bangladesh depend on tubewells for drinking water, the majority of which are relatively shallow and therefore at risk of arsenic contamination. As a result, many households are inadvertently exposed to arsenic contamination, with severe consequences for health: daily use of water contaminated with arsenic at the Bangladeshi standard (50 parts per billion or ppb) is associated with an additional 1 in 100 lifetime risk of cancer, rising to more than 1 in 10 at higher contamination levels (Smith, Lingas, & Rahman, 2000). The WHO recommends that arsenic concentrations in drinking water are limited to 10ppb. In Bangladesh as a whole, around 39 million people still drink water that is contaminated with arsenic at the WHO standard (BBS & UNICEF, 2015).

Our study area is located near the city of Bogra in Northern Bangladesh, as shown in Figure 1. Arsenic contamination is concentrated in relatively small geographical areas in this region, which, as the map shows, is relatively distant from the epicentre of the arsenic problem. As a result, this area has received little in the way of external interventions to provide arsenic safe drinking water, motivating the selection of this area for this study.

Within the study area, we targeted communities with high levels of arsenic contamination. At baseline, 62% of households in the study population had drinking water that tested positive for arsenic contamination at the WHO standard, and 24% of households had drinking water that tested positive at the less conservative Bangladeshi standard.

2.2 The intervention

The intervention we evaluate is a program of subsidies and technical advice to provide arsenic-safe sources of drinking water, implemented by a Bangladeshi NGO, "NGO Forum for Public Health". The new safe sources of water are deep tubewells. These wells draw water from aquifers that are sufficiently deep to be free from arsenic contamination, which for hydro-geochemical reasons only affects shallow groundwater. These wells are much more expensive than the much more common shallow tubewells, costing on average US$700 in the study area. Programme subsidies range in value from 90 to 100% of well installation costs. Where subsidies are less than 100%, a community contribution is required, either in cash or labour.[7] The technical advice consists of guidance in safe site selection,[8] coordination with the contractors who install the safe wells, and quality control to ensure that the installed wells are safe from arsenic contamination.[9]

Participation in the program requires several types of collective action. First, households must participate in a community meeting at which they decide how many, if any, of the offered water sources to install; where to construct the source(s); how to divide any required contributions between the households; and which households will take responsibility for the management and maintenance of each new water source. Communities are required to take these decisions by unanimous consensus in the presence of project staff at the community meeting, at which women and the poor must be represented.[10] Second, they must secure access to the chosen site(s), requiring the landowner to agree to construction of a well on their land and to maintaining open access to the site during the lifetime of the source. Third, they must collectively raise the cash contribution or coordinate the labour contributions, if these are required. Finally, they must maintain, and if necessary repair, the water source(s) over time.

3 Study design

This study measures the effect of group size by exploiting an assignment rule which generates exogenous variation in the size of communities invited to participate in the safe drinking water program, conditional on a smooth function of the variable used to assign treatment unit size. Additionally, we exploit random assignment to treatment and control groups to evaluate how program impact varies with group size. In this section, we describe relevant features of the study design.

Recruitment We recruited communities to the study who faced a substantial problem with arsenic contamination at baseline. Recruitment continued until we had reached our

---

[7] Communities are randomly assigned to the requirement to contribute in either cash or labour. We are evaluating how this requirement affects program impact in related ongoing work (Cocciolo, Ghisolfi, Habib, & Tompsett, 2018a).

[8] For example, safe sites for tubewell installation must be a safe distance from any nearby latrines, must have adequate drainage, and must be at low risk of flooding.

[9] Tubewell installation is risk-free for communities: cash contributions are refunded and labour contributions compensated at prevailing wages if installation of a safe well is unsuccessful.

[10] If communities did not agree at the first meeting, we organized a second meeting, up to a maximum of three meetings. In practice, no community held more than two meetings. Only one community which organized a meeting failed to reach an agreement. In this community, they declined to hold further meetings after a second meeting was unsuccessful in reaching agreement. One community declined to organize a meeting.

target recruitment goals. All recruited communities had arsenic contamination in at least 25% of water sources, or a substantial cluster of arsenic-contaminated sources.[11] No eligible community declined to participate in the study.

Our intervention only makes sense in communities facing an arsenic contamination problem, as otherwise there is no incentive to install costly deep wells. However, arsenic contamination is widespread in Bangladesh. Previous research also shows that arsenic contamination is largely uncorrelated with other characteristics which might affect the success of collective action (Madajewicz et al., 2007; Field, Glennerster, & Hussam, 2011). As a result, we do not expect the focus on communities with high levels of arsenic contamination to materially affect the external validity of the results.

Treatment unit definition We implement the program in treatment units of between 50 and 250 households. Previous experience (Madajewicz, Tompsett, & Habib, 2018) suggested that it would be both logistically difficult to implement the program and statistically difficult to detect effects in communities of larger than 250 households. However, the smallest local administrative units in the study area range in size from below 50 households to almost 1000 households. These administrative units typically correspond to villages or small groups of villages. To implement the program, we therefore divided the 117 administrative units represented in the study into 171 treatment units in the target range of between 50 and 250 households.

To define these treatment units, we first obtained the most up-to-date available lists of resident households from administrative sources. We note that these administrative lists were neither held in a centralized repository nor systematically digitized before our program activities, so that they are not used for assignment of any other programs. The lists typically recorded names of household heads by the *para*, or cluster, of residence. Field staff visited each community and produced sketch maps of the geography of the constituent clusters. Using the lists and sketch maps, we obtained administrative unit sizes, excluded administrative units with less than 50 households, and divided larger administrative units into smaller treatment units along natural boundaries.

Figure 2 illustrates the process we followed. The top panel shows a typical administrative unit (yellow boundary) which lies just below the threshold of 250 households. Accordingly, we treat this administrative unit as a single treatment unit (orange boundary). The bottom panel shows an administrative unit which lies just above the threshold of 250 households (yellow boundary). Accordingly, we divide the administrative unit into two separate treatment units (orange boundaries). In this case, the administrative unit divides naturally into two separate

---

[11] Appendix A1 provides further details on the recruitment procedure.

clusters, so we simply treat each cluster as a separate treatment unit. In this case, the assignment rule results in two treatment units, one with 105 households and the second with 149 households. The rule for treatment unit definition predicts a treatment unit size of 127 households for both of these two treatment units.

Figure 3 plots treatment unit size as a function of administrative unit size for all treatment units in our dataset. The blue line shows the expected predicted treatment unit size, given the project rules. Each dark grey dot represents a single treatment unit. For administrative units with less than 250 households, administrative unit size maps almost exactly to treatment unit size. For administrative units of more than 250 households, we divide the administrative unit along natural boundaries, for logistical reasons. As a result, individual treatment units may be larger or smaller than the expected treatment unit size, but the mean treatment unit size across treatment units within an administrative unit corresponds closely to the assignment rule.[12]

Assignment to treatment and control We randomly assigned treatment units to treatment and control at public lottery meetings, held sequentially in each union or local area. Approximately 75% of study treatment units were assigned to receive the safe drinking water program, under one of the three randomly-assigned contribution requirements: a requirement to contribute approximately 10% of the project costs in either cash or labour, or a waiver of the requirement to contribute. The remaining 25% of treatment units were assigned to the control arm and did not receive any intervention under this program, although we did not prevent control units from receiving any other interventions.

We used public lottery meetings to assign treatment to minimize potential discontent, which might have otherwise arisen because of the different contribution rules applied in neighbouring treatment units. All public lottery meetings were conducted in the presence of members of the research team or research assistants from both Sweden and Bangladesh. Public lotteries took place after collection of baseline data.

Take-up Take-up of the program was moderately high. Communities successfully completed all the collective action tasks required for us to attempt installation for approximately

---

[12] We occasionally had to reevaluate the total number of households in an administrative unit after discovering errors in the administrative lists provided to us. In our analysis, we use the administrative size as observed at the time of treatment unit definition to predict treatment unit size according to project rules and the corrected data to measure final treatment unit size. Additionally, we were in some cases unable to follow the treatment rule exactly; for example, because the geometry of the community meant that we could not divide it in the target number of treatment units while still respecting the rule that individual treatment units should be between 50 and 250 households in size. We provide more detail on these cases in Appendix A2. In all cases, the instrument is the *predicted* treatment unit size, not the treatment unit size finally implemented, meaning that these deviations do not alter the validity of the exclusion restriction.

2/3 of the water sources we offered. However, take-up was heterogeneous across the contribution requirements: communities only completed all the collective action tasks successfully under the cash contribution requirement for 22% of offered water sources, while they did so for 86% and 88% of the offered sources under the labour and waiver treatment arms respectively.

Unfortunately, we were unable to install water sources in about 10% of attempted installations, despite communities having successfully completed all the required collective action tasks. In these cases, we could not complete installation because of hydrogeological conditions which prevented successful installation.[13] Unsurprisingly, since these hydrogeological conditions are unpredictable and unobservable from the surface—and therefore very unlikely to be correlated with any other local characteristics—these failures occurred at the same rate in attempted installations across all treatment arms.

We pool all treated communities into one group, and we do not account either for the lower take-up rate or the installation failures when we estimate average program impacts. These effects likely attenuate the estimates on impact, as we discuss in Section 5. In future revisions, we will present results that account for these attenuating factors.

Timing We collected baseline data in late 2015 and early 2016. We carried out implementation of the safe drinking water program between spring 2016 and the end of 2017.[14] We completed follow-up data collection in mid-2018.

4 Data

We draw on three main sources of data. First, we use data from a household survey, conducted in a random sample of 40 households in each treatment unit,[15] covering household demographics, health, wealth, social networks, and behaviour related to water collection and use. Our data also matches households with all the water sources they use: each household identified the water source(s) used to obtain water for drinking or cooking, selecting water sources from a list of all available water sources established during a baseline water source census.[16]

---

[13] In particular, either an impenetrable rocky layer obstructed drilling or a sandy layer caused the excavation to collapse before the well pipe could be installed.

[14] We added 16 treatment units to the study in 2017 when additional funding became available. We collected baseline data for these treatment units in early 2017.

[15] We randomly sampled households from the administrative lists using pseudo-random number generation. In some cases, we surveyed more than 40 households, usually because we needed to redefine treatment unit boundaries after discovering problems with the administrative lists.

[16] We used local descriptors to list community water sources, and verified that the correct water source was recorded using photographs of the water sources. At both baseline and follow-up, enumerators could

Second, we use data from an arsenic testing program to evaluate the effects of the program on water quality.[17] At baseline and follow-up, we tested water quality in both household drinking water and all water sources the households report using.[18] To test for arsenic contamination, we use the EZ Arsenic High Range Test Kit (or Hach Kit). We focus on the presence of arsenic contamination at or above the WHO threshold of 10 parts per billion (10 ppb).[19] Results of the water quality tests are available after twenty minutes and were reported to survey respondents in the field at the time of the survey.[20]

To test household drinking water quality, we requested that households obtain a sample of drinking water the same way they would if someone in the household wished to drink. To test source water quality, enumerators collected water directly from the source. Since households obtain drinking water from multiple sources,[21] we aggregate the information to obtain a single value for each household, weighting test results according to the fraction of water obtained from each source.[22]

The two measures of arsenic contamination—in household drinking water and at source—have different advantages and disadvantages. The ability of the Hach Kit to detect the presence of arsenic decreases the longer water is stored. Arsenic begins to oxidize once the water is stored in a container that is open to the air, and the test does not detect oxidized arsenic. In contrast, the test is more likely to detect all arsenic present in water drawn directly from a tubewell. As a result, our measure of arsenic in household drinking water may underestimate the true extent of arsenic contamination, while variation in storage behaviour may introduce additional noise into our estimates.

On the other hand, a concern about arsenic contamination data from water sources is the potential presence of social desirability and experimenter demand effects (see e.g. Ahuja,

---

choose to add a new source if the source did not appear in the water source census data.

[17] As part of a larger study, we also collected data on fecal contamination. We report these results in Cocciolo, Ghisolfi, Habib, Rashid, and Tompsett (2018). We focus on arsenic contamination here because the effect of new water source construction on fecal contamination in household drinking water is ambiguous, because improvements in source water quality may be offset by increased recontamination during transport and storage (see e.g. Kremer, Leino, Miguel, & Zwane, 2011).

[18] At baseline, we conducted water quality tests in all community water sources during the baseline water source census. At follow-up, we conducted water quality tests in all water sources that at least one household reported using.

[19] Arsenic concentration is highly skewed, which is why we focus on an indicator for arsenic contamination at a given threshold instead of arsenic concentration levels. The WHO standard is more conservative than the Bangladeshi standard, but displays greater variation, making it easier to detect effects. See Cocciolo, Ghisolfi, Habib, Rashid, and Tompsett (2018).

[20] Enumerators also provided all households with contaminated sources with information about ways to reduce or avoid exposure to contamination in drinking water, regardless of treatment status.

[21] Households reported using on average 1.03 sources at baseline. At follow-up, households in the treated and control group reported using 1.24 and 1.12 sources, respectively.

[22] For example, a household obtaining half their water from an arsenic-contaminated source and half from an arsenic-safe source would have a weighted average value of source arsenic contamination of 0.5

Kremer, & Zwane, 2010). These effects might lead households, especially those who have been exposed to interventions designed to increase use of safe drinking water, to overreport use of safe sources. However, note that in this context, we find that source water contamination strongly predicts household water contamination, and that there is no change in the relationship between source and household contamination as a result of treatment (Cocciolo, Ghisolfi, Habib, Rashid, & Tompsett, 2018). As a result, social desirability and experimenter demand effects may not be a significant concern.

We also use detailed information about project implementation recorded by research assistants throughout the study time period.

5 Empirical approach

5.1 Main specification

The relationship of interest is between group size and either measures of collective action observed during the project implementation or measures of success in public good provision. The main structural equation is therefore:

$$Y_i = \beta_0 T_j + \beta_1 T_j \times SIZE_j + Z_i \gamma + \epsilon_i (1)$$

where Y_i is an outcome in household i, located in treatment unit j, in administrative unit k; T_j is an indicator for assignment to treatment; and $SIZE_j$ is the number of households in the community. Together, β_0 and β_1 describe the treatment effect of the program and how it varies with community size.[23]

Since T_j is randomly assigned to communities, meaning that it is uncorrelated with any other determinants of Y_i, the estimated coefficient β_0 has a causal interpretation when we estimate Equation 1 by OLS. The estimated coefficient β_1 also has an immediate causal interpretation in the sense that it describes how the treatment effect varies with respect to community size. However, community size is not randomly assigned. If community size is correlated with any other community characteristics which are in turn associated with heterogeneous effects of the safe drinking water program, then β_1 will also capture heterogeneity associated with these other characteristics.

To eliminate this risk of bias, we exploit exogenous variation in $SIZE_j$ generated by the rule we applied to define treatment units as a function of administrative unit size. We use this rule to predict treatment unit size, $SIZE_{pred,k}$, as a function of administrative unit

---

[23] In some cases, we are interested in outcomes which we only observe in treated communities. In these cases, we adjust Equation 1.

size, $AUSIZE_k$. As seen in Figure 3, the assignment rule results in a highly non-linear and non-monotonic function of administrative unit size. In contrast, all other community-level characteristics are likely to vary smoothly with administrative unit size, if they vary at all, given that administrative unit size is somewhat arbitrarily defined. Conditional on a smooth function of administrative unit size, $SIZE_{pred,k}$ predicts $SIZE_j$ but is uncorrelated by construction with any other community characteristics. Our empirical strategy therefore uses $SIZE_{pred,k}$ interacted with T_j as an instrument for $SIZE_j$ interacted with T_j in Equation 1, in a 2SLS estimating framework, controlling for a smooth function of $AUSIZE_k$ interacted with T_j. In the following subsections, we demonstrate that the conditions for instrument validity are met, using falsification tests in pre-intervention data to support the exclusion restriction.

The control vector Z_i includes predicted treatment unit size (without its interaction with treatment status) whenever we include the control communities in the analysis, and a quadratic of administrative unit size, interacted with treatment status.[24] We measure administrative unit size relative to its median value in the data, so that the coefficient β_0 estimates the treatment effect at median administrative unit size in the data. The control vector also includes either a constant or a full set of lottery fixed effects.[25]

The empirical strategy means that observations close to the thresholds have potentially high leverage and the sample is relatively small. To minimize the risk of drawing incorrect inference, we bootstrap standard errors in all regressions by resampling 2000 times at the administrative unit level, respecting stratification at the lottery level.[26]

We construct all outcome measures to be scale free, to avoid mechanical correlations with treatment unit size. For ease of interpretation, we also rescale treatment unit size to be measured in hundreds of households. Thus, the coefficients in all tables show the effects of increase treatment unit size by one hundred households.

5.2 First stage

Table 1 shows the first stage regressions. Columns 1 and 2 show the first stage regressions for regressions in which each observation is a treatment unit; columns 3 and 4 show the

---

[24] Preliminary robustness checks suggest that the results are insensitive to using alternative smooth functions of adminstrative unit size.

[25] When we analyze the treated communities alone, we omit the constant, so the coefficient β_0 captures the intercept in the treated group. When we analyze the treated communities and include lottery fixed effects, we omit the treated dummy as it is colinear with the lottery fixed effects in the sample of treated communities only, and we thus estimate only β_1 in these regressions.

[26] Clustered standard errors at the administrative level do not yield the correct inference in this context, with a larger fraction of baseline balance tests rejected. Results available on request. See Young (2017) for further discussion of the advantages of the bootstrap in conducting inference in 2SLS.

same regressions for data at the household level. Columns 2 and 4 include a full set lottery fixed effects in place of the constant in columns 1 and 3. All regressions include the controls specified in the previous section.

All four columns show that the instrument, predicted treatment unit size interacted with treatment status, is a very strong predictor of the endogenous variable, implemented treatment unit size interacted with treatment status. First stage χ^2 statistics confirm these strong relationships. The estimated coefficients are slightly smaller than one because it was more common for us to have to divide an administrative unit into more treatment units than our rule prescribed than vice versa.

It is however worth noting that the partial R-squared, which measures the percentage of the variance in the endogenous variable explained by the excluded instrument, ranges between 9 and 11%. Since instrument power is a function of the partial R-squared, these relatively low levels suggest some concern regarding power to detect effects, even before we account for heterogeneity in take-up or installation failures.

5.3 Instrument validity

The exclusion restriction requires that, conditional on the control variables including a smooth function of administrative unit size, predicted treatment unit size is uncorrelated with any other determinants of the outcome variables through any other channel than its effect on treatment unit size. Our study design aims for this requirement to be met by construction. Table 2 shows the results of tests which evaluate whether predicted unit size is correlated with program characteristics or community characteristics that could predict success in collective action. We estimate a reduced form version of Equation 1, substituting predicted treatment unit size for implemented treatment size, and using only baseline data.

One key way in which the exclusion restriction could be violated is if the amount of resources dedicated to the project varied with group size in per capita terms. Deep tubewells can be thought of as local public goods, providing a source of safe drinking water that is essentially non-rival in consumption,[27] although subject to mild congestion effects (time waiting at well versus other sources), but to which households must walk to obtain water, imposing use costs which increase with the distance between a household and the source. As a result, the empirical comparison of interest is one which holds constant the ratio of households to water sources. Within the constraints imposed by the allocation of lumpy

---

[27] In practice, of course, the water from the well is rival in consumption. However, as long as extraction volumes are modest—as is the case when water is only extracted for drinking and cooking, rather than for high volume uses such as irrigation—then one household's extraction does not diminish the ability of other households to extract safe water.

investments—we could not assign half a tubewell—we designed our assignment rules to hold this ratio constant.[28] Column 1 of Table 2 shows that we were successful in holding this ratio constant: the instrument does not predict the ratio of households to tubewells.

Another potential concern is our use of natural boundaries in defining the treatment units. As a result, one might worry that treatment unit size is correlated with something that is also affected by natural divisions within a community, such as fragmentation. However, recall that although implemented treatment unit size (the endogenous variable) is determined by natural boundaries, predicted treatment unit size (the instrument) is not. As a result, there is no correlation between the instrument and a measure of geographical fragmentation, the number of clusters per household in the treatment unit (Column 2 of Table 2).

Columns 3 to 7 of Table 2 confirm that predicted treatment unit size, interacted with treatment, does not predict any baseline characteristics. We focus on measures of arsenic contamination, since that is one of our main outcome variables, and on social characteristics that may also predict success in collective action. Out of five tests, only one shows a statistically significant correlation with the instrument, and only at the ten percent level. This difference likely arises due to chance.

Note that arsenic contamination *is* correlated with treatment status (columns 3 and 4). In Cocciolo, Ghisolfi, Habib, and Tompsett (2018b), we show a larger number of randomization tests, along with joint tests of the null that treatment status is uncorrelated with a full set of baseline variables. These tests confirm that these differences arise due to chance. Nonetheless, we always estimate effects on arsenic contamination in first differences, meaning that we remove any baseline differences from the comparison between treatment and control groups.

We repeat the balance checks with lottery fixed effects (Appendix Table C1). Lottery fixed effects absorb mean differences in outcomes between treatment units assigned at different lotteries. However, including lottery fixed effects does not improve the baseline balance, so our main specifications exclude lottery fixed effects. We show results that include lottery fixed effects as robustness checks.

Figure 4 shows the distribution of administrative units by administrative unit size, illustrating that there is no “bunching” around the thresholds we use to define treatment unit size, implying that there are no incentives for administrative units to manipulate their size with respect to these thresholds. Appendix Figures C1a to C1c illustrate the results of McCrary Density tests that more formally confirm that there are no discontinuities across any of the thresholds.

The final requirement for instrument validity, monotonicity, is met by construction.

---

[28] Appendix A3 describes the rules we implemented to achieve this.

6 Results

We show results in Tables 3 to 7. Each table shows both the reduced form results and the 2SLS results for each outcome variable. Corresponding Appendix Tables show the results with lottery fixed effects. Section 6.1 reports results on measures of collective action while Section 6.2 reports results on public good provision.

6.1 Collective action

Meeting attendance The first stage of collective action required by the program is attendance at a community meeting and participation in collective decision-making. Columns 1 and 3 of Table 3 shows that representation at the meeting falls with increasing treatment unit size. While 53% of households are represented at a community meeting in a treatment unit with 125 households, this falls to 47% in a treatment unit with 250 households. However, the difference is not statistically significant at the 10% level. Also, there is one community meeting in each treatment unit,[29] so the average distance households have to walk to attend the meeting rises mechanically with treatment unit size. Representation falls with increasing walking distance. The relationship we observe between treatment unit size and representation almost completely attenuates when we account for distance between households and the meeting sites (columns 2 and 4).

The results on female representation are stronger.[30] While 32% of households are represented by at least one female member, on average, in treatment units of 125 households, this falls to 23% in treatment units of 250 households, and the differences are statistically significant (columns 5 and 7). Unlike the effects on any representation, these effects are not explained by accounting for distance between households and meeting sites: female representation is significantly lower, even accounting for distance (columns 6 and 8). These results suggest that women are dissuaded from attending meetings in larger groups for other reasons than simply distance.

Deliberation Column 1 and 2 of Table 4 show that households spend less time deliberating per offered source in larger treatment units. The outcome variable is time taken for the meeting divided by the number of sources offered to the community. While treatment units with 125 households spend approximately 79 minutes in the meeting per offered source, this falls to 29 minutes per offered source in treatment units with 250 households. One

---

[29] This is the exception to our rule of maintaining resources constant in per capita terms across treatment units

[30] Note, however, that we do not have power to detect whether effects on female representation differ from effects on representation by any household member.

interpretation of this result is that less deliberation goes into each decision in larger treatment units; a second interpretation is that decision-making is more efficient in larger treatment units.

The results in columns 3 and 4 show how many possible locations are discussed in the meeting for each offered well. While 1.7 locations are discussed on average for each offered well in communities of 125 households, this falls to 1.1 locations for communities of 250 households. Together with the results in columns 1 and 2, this suggests that the correct interpretation is that less deliberation about each site takes place in larger communities.[31]

Contributions The final stage of collective action required prior to installation is the collection of community contributions in cash or the coordination of community contributions in labour. Columns 1 and 3 of Table 5 show the aggregate results on rates on fraction of households contributing.[32] A weakly smaller fraction of households in our sample contribute in larger treatment units: about 4.4% of households contribute in a community of 125 households, falling to 1.9% of households in a community of 250 households.[33]

The total fraction of households contributing nests two factors: first, whether the group successfully met the contribution requirement, and second, how many households contributed, conditional on successfully meeting the contribution requirement. Columns 2 and 4 show instead the fraction of households who contribute, conditional on successfully raising contributions for at least one source in the community. These columns show that in larger groups, a smaller number of households contribute to each source for which contributions are successfully raised, even though, as we show in Section 6.2, larger communities are weakly more likely to successfully raise contributions. About 8.2% of households contribute in treatment units of 125 households which successfully raise contributions, falling to 1.8% in treatment units with 250 households.

Columns 5 to 8 show an alternative measure of contributions: the number of contributing households per tubewell. In columns 5 and 7, we report effects on the number of contributing households per offered tubewell, coding the number of contributors as zero where the community failed to meet the contribution requirement. In columns 6 and 8, we report effects on the number of contributing households only for tubewells where the community successfully met

---

[31] We will be able to shed more light on this in future revisions of this paper, as we have data tracking the use of time in the meeting. We can thus distinguish between time spent discussing general issues (the arsenic contamination problem, project rules) and time spent discussing site location.

[32] A minor change in the way we estimate the effects on contributions is that we are unable to fully stratify the bootstrap when we focus only on wells for which communities successfully raised contributions, because in some cases there is only one administrative unit from a given lottery in this sample. To generate the results in this table, we pool three smaller lotteries into a single strata, and use this when we stratify the bootstrap.

[33] Although note that these differences are hard to detect in our random sample of households: these figures correspond to on average only 1 or 2 households in the sample in each treatment unit.

the contribution requirement. The results are similar to those in columns 1 to 4. The total number of contributors per offered well falls from approximately 5.1 to 2.7 as treatment unit size rises from 125 to 250 households. Conditional on successfully meeting the contribution requirements, the total number of contributors falls from 9.2 to 3.2 households over the same range of treatment unit sizes.

6.2 Public good provision

Installations Table 6 shows results which measure how successfully the communities resolved the collective action problem, leading to an attempt at installation (indicating that the community successfully held a meeting, chose a location, secured agreement from a landowner, and if necessary met the contribution requirement) and, hydrogeological conditions permitting, successful installation. Figure 5 also visualizes these results. The results in columns 1 and 2 indicate that the average rate at which we attempted installations was around 65% of offered tubewells in treatment units of 125 households, rising to 70% of offered tubewells in treatment units of 250 households. The differences in installation rates across the full range of treatment unit sizes are small, and the confidence intervals wide.

Columns 3 and 4 of Table 6 and panel b) of Figure 5 repeat the analysis for the fraction of offered tubewells successfully installed. The entire graph shifts downwards, reflecting the fact that about 10% of attempted installations were unsuccessful. The same pattern of results remains: if anything, installation rates increase with treatment unit size, but the differences are small and imprecisely estimated.

Impacts on access to safe drinking water Table 7 shows the estimated effects of the program on access to safe water. We report three measures of program impact: first, the share of household water for drinking and cooking obtained from a project source (columns 1 and 2); second, the change in arsenic contamination at the household level (columns 3 and 4); and third, the weighted mean change in arsenic contamination at water sources. Figure 6 visualizes the same results.

Across all three measures of program impact, the program effect declines with treatment unit size. While about 8.0% of community water used for drinking and cooking is provided from a new source in treatment units of 125 households, this falls to 3.4% in treatment units of 250 households; a 5.0% reduction in household arsenic contamination falls to 1.6% over the same range; and a 12% reduction in water source contamination reverses to a 4% increase, also over the same range. Note, however, that although these effects are substantial in magnitude, the confidence intervals for all the estimated effects of treatment unit size do not exclude zero.

The effects are imprecisely estimated, especially the effects on arsenic contamination in household drinking water,[34] for several reasons. The main reason for the imprecision is attenuation as a result of low take-up in the cash treatment arm, and further attenuation due to the installation failures. Since we estimate an average treatment effect of the program, our effects are attenuated by both these effects. The effects on household contamination are likely further attenuated by the reduced ability of the Hach test to detect arsenic in stored water: more than 60% of households recover their household drinking water sample from storage rather than direct from the source.

These results are potentially most sensitive to any deviation from our study design in terms of our ability to hold constant the ratio between households and tubewells. Appendix Table C6 shows results controlling fairly conservatively for a quadratic in tubewells per household.[35] There is some attenuation in the estimated effects of treatment unit size on program impact in terms of water use from project sources and arsenic contamination in household drinking water, but the estimated effects on water source contamination are larger. These results suggest that our main estimating strategy is successful in holding constant the ration between households and tubewells.

7 Discussion

In the absence of an effective state, groups of people must work together—or take collective action—to provide public goods and services. Many development interventions also depend on some type of collective action, such as collective decision-making, maintenance of a communal asset, or participatory monitoring and management. In practice, successful collective action is elusive. Theorists have posited many different explanations for why collective action often fails, and under which circumstances it might succeed. However, we have remarkably little robust empirical evidence to evaluate these theories.

This study has made progress in addressing this knowledge gap. We combine a randomized controlled experiment with an assignment rule that generates exogenous variation in treatment unit size to evaluate how treatment unit size affects collective action and public good provision in the context of a program to provide safe drinking water in rural Bangladesh. We hold the ratio of resources to households constant across treatment units and evaluate

---

[34] Indeed, there is no statistically significant average effect of the program on arsenic contamination in household drinking water (Cocciolo, Ghisolfi, Habib, Rashid, & Tompsett, 2018).

[35] Our study design held constant the ratio of households to tubewells, which is most closely correlated with treatment unit size. However, this is not the same as holding constant the ratio of tubewells to households. Larger treatment units have slightly smaller ratios of tubewells to households, and more variance in the ratio of tubewells to households.

all outcomes using scale-free measures.

We find that participation in collective action is lower in larger treatment units, across a range of outcomes: attendance of community decision-making meetings, especially by women; less time taken to deliberate over sites; fewer potential sites discussed for each location to be decided; and fewer households contributing to each successfully installed tubewell. The larger groups are weakly (insignificantly) more successful at installing more wells, but the program has less impact in terms of providing safe drinking water across three measures of impact in these larger groups. However, some attenuating effects mean that these heterogeneous effects are imprecisely measured and none of the estimated effects on how group size affects measures of impact are individually statistically significant.

The question of at which scale to implement a program is one that all organizations must answer. In the context of this study, for example, organizations which wish to implement safe drinking water programs must decide whether to implement the program at district level, local administrative unit level, or at some smaller level that more closely corresponds to communities of individuals who interact frequently with one another. Our results confirm that there are drawbacks to implementing at larger scales, particularly with respect to participation in collective action, and quantify these drawbacks. However, there are also potential economies of scale in implementing programs at larger scales. We have detailed cost data which we can leverage to understand, in this context, the trade-offs between economies of scale and group size effects. We will incorporate this information into future revisions of this paper.

Two further issues are worth highlighting here, which we will expand upon in future revisions. First, much of the theory describing the relationship between group size and collective action concerns theoretical goods that are pure public goods within the group. In our context, as in many real-world contexts, the public goods at stake are local public goods, meaning that smaller sub-groups (in our case, clusters) within groups (in our case, treatment units) may compete over the available resources. The response of effort in a sub-group to total group size is a different theoretical question to the response of effort in a sub-group to sub-group size. In future revisions, we will provide a theoretical formalization of this idea and apply it to our context by evaluating heterogeneous effects on participation by cluster size.

Second, our study relates to the literature on decentralization or federalism (see e.g. Oates, 1972; Galasso & Ravallion, 2005). We hold the ratio between population and resources constant across treatment units, but arsenic contamination varies across and within treatment units. Larger treatment units therefore have greater ability to target wells to areas with high arsenic contamination, another potential advantage of treatment unit size in this context, which makes the negative effects on impact still more striking. In future revisions, we will provide additional analysis on the locations chosen for proposed and installed wells, and how these locations correspond to the areas of highest arsenic contamination.

In summary, this paper provides new causal evidence on the effect of group size on collective action in a real-world context. We show that in larger groups, households participate less in collective action and programs to provide safe drinking water are less effective, although the estimated effects on program impact are imprecisely measured. Our results suggest that policy-makers should carefully consider the choice of scale in program design in order to maximize both cost-effectiveness and equity in program impact.

References

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

Angrist, J. D. & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, *114*(2), 533–575.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Cooperation in collective action. *Economics of Transition*, *13*(3).

Banerjee, A. & Iyer, L. (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in India. *American Economic Review*, *95*(4).

Banerjee, A., Iyer, L., & Somanathan, R. (2007). Public action for public goods. *Handbook of Development Economics*, *4*(39).

Bardhan, P. K. (2000). Irrigation and cooperation: An empirical analysis of 48 irrigation communities in South India. *Economic Development and Cultural Change*, *48*(4).

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

Brunner, E. & Sonstelie, J. (2003). School finance reform and voluntary fiscal federalism. *Journal of Public Economics*, *87*(9/10).

Chamberlin, J. (1974). Provision of collective goods as a function of group size. *American Political Science Review*, *68*(2).

Chattopadhyay, R. & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, *72*(5).

Cocciolo, S., Ghisolfi, S., Habib, A., Rashid, S., & Tompsett, A. (2018). *Access to safe drinking water: Experimental evidence from new water sources in bangladesh*. Final report to 3ie.

Cocciolo, S., Ghisolfi, S., Habib, A., & Tompsett, A. (2018a). *Community contributions, participatory decision-making and local public goods: A field experiment in Bangladesh*. Work in progress.

Cocciolo, S., Ghisolfi, S., Habib, A., & Tompsett, A. (2018b). *Community contributions, participatory decision-making and local public goods: A field experiment in Bangladesh*. Work in progress.

Esteban, J. & Ray, D. (2001). Collective action and the group size paradox. *American Political Science Review*, *95*(3).

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.

Galasso, E. & Ravallion, M. (2005). Decentralized targeting of an antipoverty program. *Journal of Public Economics*, *89*(4).

Isaac, R. M. & Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, *103*(1).

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

Madajewicz, M., Pfaff, A., van Geen, A., Graziano, J., Hussein, I., Momotaj, H., … Ahsan, H. (2007). Can information alone change behavior? Response to arsenic contamination of groundwater in Bangladesh. *Journal of Development Economics*, *84*.

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

Miguel, E. & Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, *89*.

Oates, W. E. (1972). *Fiscal federalism.* NY: Harcourt Brace Jovanovich.

Olson, M. (1971). *The logic of collective action.* Harvard University Press.

Pareto, V. (1927). *Manual of political economy.* New York: A.M. Kelley.

Poteete, A. R. & Ostrom, E. (2002). Heterogeneity, group size and collective action: The role of institutions in forest management. *Development and Change*, *35*(3).

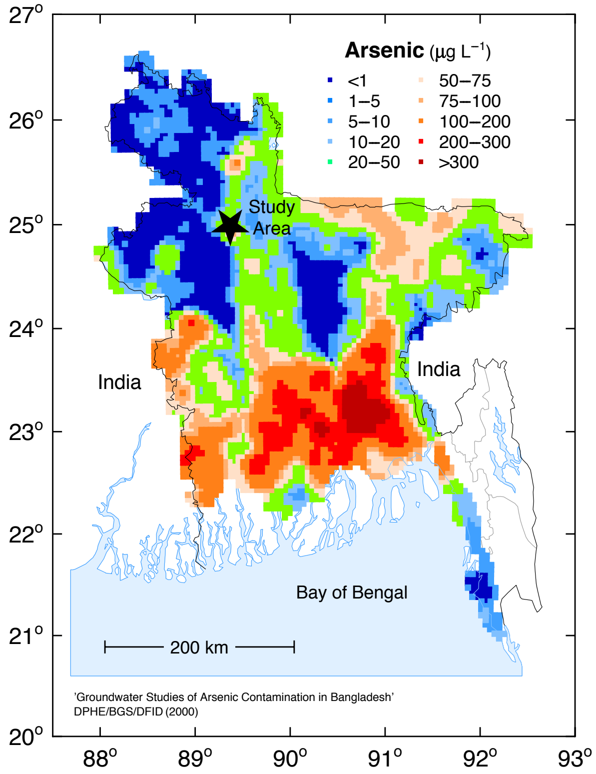
Sawada, Y., Kasahara, R., Aoyagi, K., Shoji, M., & Ueyama, M. (2013). Modes of collective action in village economies: Evidence from natural and artefactual field experiments in a developing country. *Asian Development Review*, *30*(1).

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.

Young, A. (2017). *Consistency without inference: Instrumental variables in practical application.* Working Paper.

Figures and Tables

Figure 1: Arsenic contamination in Bangladesh and study site location

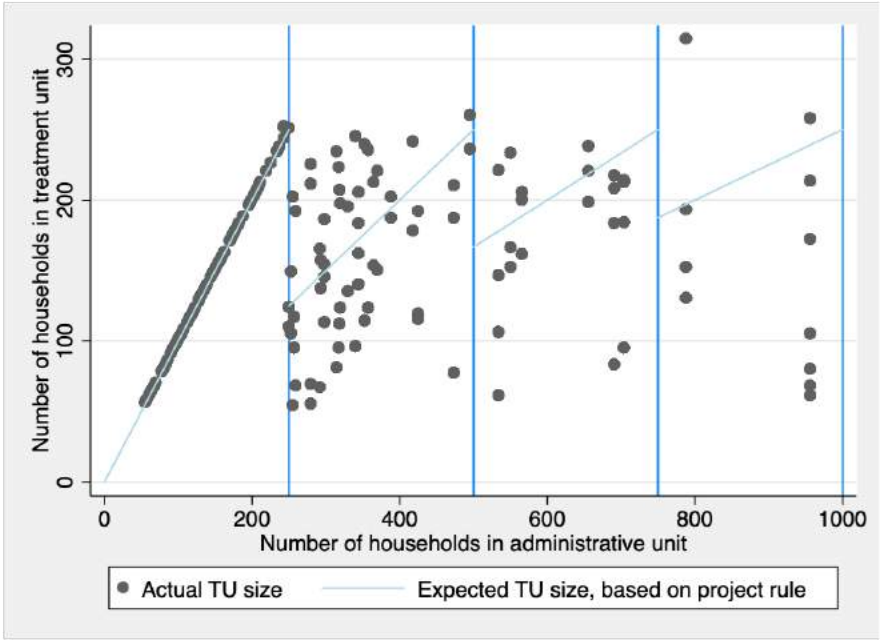


'Groundwater Studies of Arsenic Contamination in Bangladesh'
DPHE/BGS/DFID (2000)

Figure 2: Examples of treatment unit definition

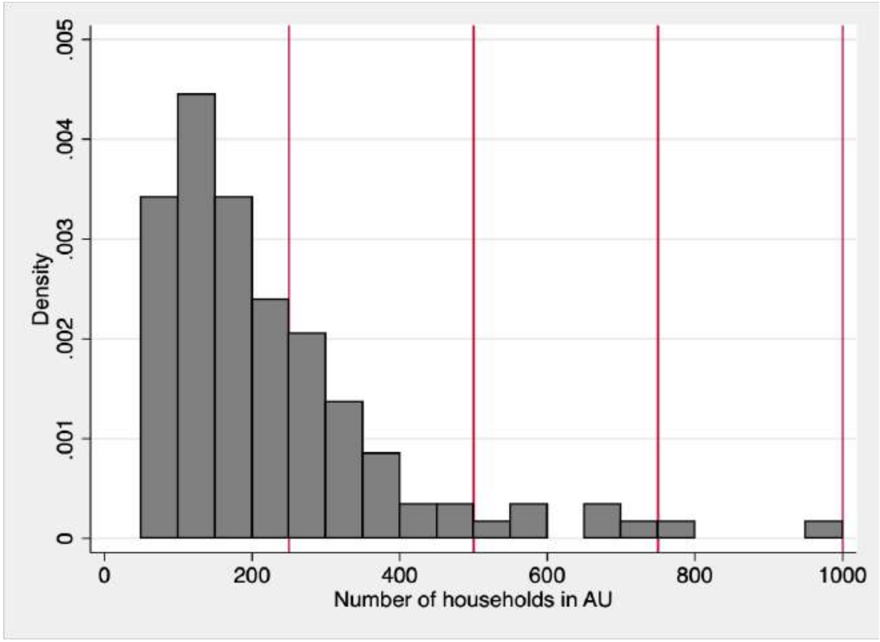
Notes Figures illustrate treatment unit definition. Top panel shows an administrative unit with 243 households. Since it is below the threshold of 250 households, the administrative unit is treated as a single treatment unit. Bottom panel shows an administrative unit of 254 households. Since it lies above the threshold of 250 households, the administrative unit is divided in two along natural geographical boundaries.

Figure 3: Treatment unit definition rule



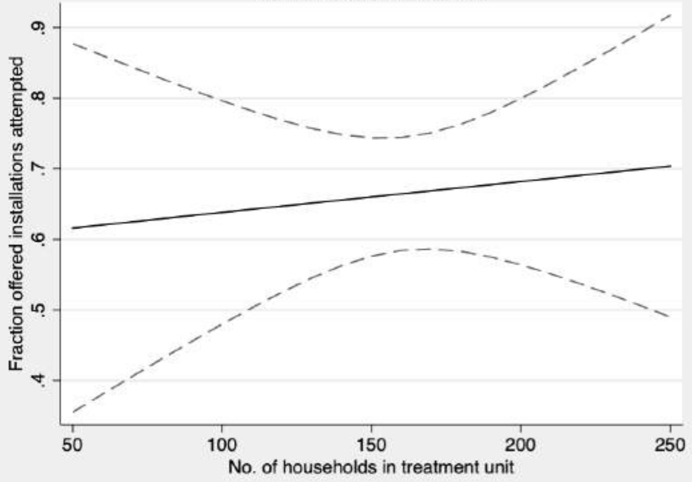
Notes Graph illustrates treatment unit definition rule. Blue lines illustrate project rule for treatment unit definition i.e. target treatment unit size. Dots represent actual treatment units.

Figure 4: Density of administrative unit distribution

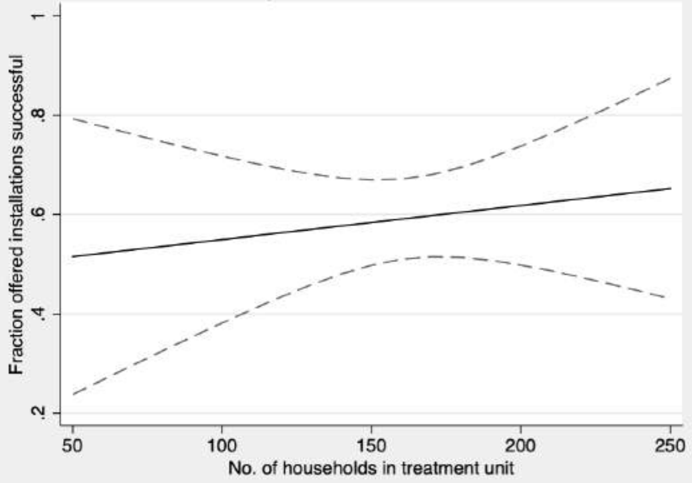


Notes Histogram shows distribution of administrative unit size. Red lines show the discontinuities in the treatment unit definition rule. The distribution of administrative units is smooth across all boundaries.

Figure 5: Estimated effects on installation rates



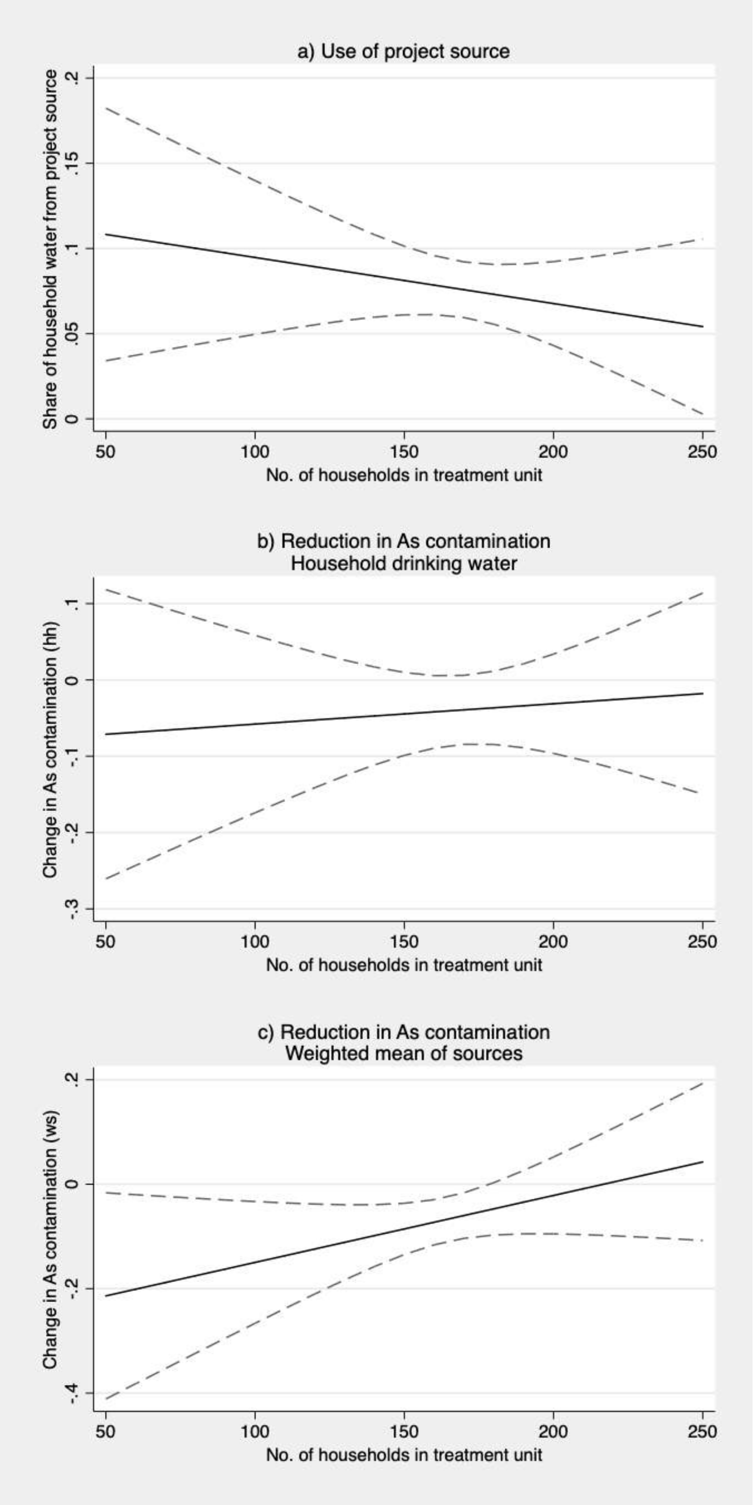
a) Installations attempted



b) Installations successful

Notes Figure shows variation in fraction of offered sources resulting in attempted (panel a) and successful (panel b) installations with treatment unit size. Estimates obtained by 2SLS regression, using predicted treatment unit size as an instrument for observed treatment unit size, controlling for a smooth function of administrative size. 90% confidence interval shown as dashed line.

Figure 6: Estimated effects on access to safe water



Notes Figure shows variation in use of project-installed sources (panel a), estimated reductions in arsenic contamination in household drinking water (panel b) and reductions in arsenic contamination in source drinking water (panel c). Estimates obtained by 2SLS regression, using predicted treatment unit size as an instrument for observed treatment unit size, controlling for a smooth function of administrative size interacted with treatment status. 90% confidence interval shown as dashed line.

Table 1: First stage regressions

	TU level		HH level	
Dependent variable: TU size x treated	(1)	(2)	(3)	(4)
Predicted TU size x treated	0.892***	0.894***	0.837***	0.827***
	(0.080)	(0.098)	(0.092)	(0.117)
Union FE	No	Yes	No	Yes
First stage χ^2	125.47	83.33	83.13	49.94
Partial R-squared	0.11	0.10	0.10	0.09
N	171	171	6529	6529

Notes: Table shows results of first stage regression, controlling for a quadratic function of administrative unit size, a treatment dummy, and their interactions, along with predicted treatment unit size. Bootstrapped standard errors in parentheses (2000 replications). lottery fixed effects where specified.

Table 2: Baseline balance check

	Ratio 100 hhs: TWs	Clusters per 100 hhs	As HH (WHO)	As WS (WHO)	Network size	High trust	Knows association
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted TU size x	0.029	-0.094	-0.184	-0.219*	-0.087	0.005	-0.088
treated	(0.060)	(0.302)	(0.115)	(0.117)	(0.256)	(0.086)	(0.075)
Treated	1.111***	0.274	0.381*	0.419**	0.226	-0.009	0.140
	(0.104)	(0.524)	(0.207)	(0.211)	(0.467)	(0.146)	(0.120)
Union FE	No	No	No	No	No	No	No
Regression level	TU	TU	HH	HH	HH	HH	HH
N	171	171	6526	6513	6529	6523	6452

Notes: Table shows results of OLS regression on predicted treatment unit size, a treatment dummy, and their interactions, controlling for a quadratic function of administrative unit size, a treatment dummy and their interactions. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

Table 3: Participation in community meeting

	Household member attended meeting				Female household member attended meeting			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TU size			-0.063	-0.018			-0.088**	-0.076*
			(0.041)	(0.051)			(0.040)	(0.043)
Predicted TU size	-0.053	-0.014			-0.074**	-0.061*		
	(0.035)	(0.040)			(0.034)	(0.035)		
Treated	0.576***	0.599***	0.593***	0.603***	0.375***	0.411***	0.398***	0.431***
	(0.062)	(0.071)	(0.071)	(0.084)	(0.064)	(0.066)	(0.072)	(0.075)
Estimation	RF	RF	IV	IV	RF	RF	IV	IV
Treated only	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Union FE	No	No	No	No	No	No	No	No
Conditional on distance	No	Yes	No	Yes	No	Yes	No	Yes
Regression level	HH	HH	HH	HH	HH	HH	HH	HH
N	4866	4428	4866	4428	4866	4428	4866	4428

Notes: Table shows results of OLS regression on predicted treatment unit size, (columns 1, 2, 5, and 6) or 2SLS regressions using predicted treatment unit size as an instrument for treatment unit size (columns 3, 4, 7 and 8). All regressions control for a quadratic function of administrative unit size. Columns 2, 4, 6, and 8 additionally control for distance between the household and the meeting site. Data from project records. Treated units only. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

Table 4: Deliberation

	Meeting duration		Discussed sites	
	(1)	(2)	(3)	(4)
TU size		-39.855***		-0.531***
		(6.248)		(0.196)
Predicted TU size	-35.447***		-0.473***	
	(5.548)		(0.171)	
Treated	122.676***	128.929***	2.302***	2.385***
	(10.211)	(11.117)	(0.315)	(0.350)
Estimation	RF	IV	RF	IV
Union FE	No	No	No	No
Regression level	TU	TU	TU	TU
N	127	127	128	128

Notes: Table shows results of OLS regression on predicted treatment unit size, (columns 1 and 3) or 2SLS regressions using predicted treatment unit size as an instrument for treatment unit size (columns 2 and 4). All regressions control for a quadratic function of administrative unit size. All outcome variables normalized by number of offered sites. Treated units only. Bootstrapped standard errors in parentheses (2000 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Contributions

	Household contributed				Contributing households per TW			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TU size			-0.020	-0.051			-1.928	-4.846**
			(0.025)	(0.041)			(1.364)	(2.113)
Predicted TU size	-0.018	-0.042			-1.740	-4.145**		
	(0.023)	(0.035)			(1.259)	(1.655)		
Treated	0.066	0.131*	0.070	0.145*	7.273***	14.057***	7.546***	15.278***
	(0.044)	(0.070)	(0.047)	(0.078)	(2.424)	(3.327)	(2.563)	(4.108)
Estimation	RF	RF	IV	IV	RF	RF	IV	IV
Contribution requirement only	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Successful installation	No	Yes	No	Yes	No	Yes	No	Yes
Union FE	No	No	No	No	No	No	No	No
Regression level	HH	HH	HH	HH	TW	TW	TW	TW
N	3266	1873	3266	1873	111	67	111	67

Notes: Table shows results of OLS regression on predicted treatment unit size, (columns 1, 2, 5 and 6) or 2SLS regressions using predicted treatment unit size as an instrument for treatment unit size (columns 3, 4, 7, and 8). All regressions control for a quadratic function of administrative unit size. Treatment units assigned to a cash or labour contribution only; columns 2 and 4 results for treatment units where at least one water source installed only; columns 6 and 8 show results for installed tubewells only. In columns 5 and 7, the number of contributors is coded as zero where tubewells were not installed. Bootstrapped standard errors in parentheses (2000 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Safe water sources installed

	Attempted installations		Successful installations	
	(1)	(2)	(3)	(4)
TU size		0.044		0.068
		(0.135)		(0.145)
Predicted TU size	0.039		0.061	
	(0.119)		(0.126)	
Treated	0.601***	0.594***	0.492**	0.481**
	(0.200)	(0.224)	(0.213)	(0.241)
Estimation	RF	IV	RF	IV
Union FE	No	No	No	No
Regression level	TU	TU	TU	TU
N	129	129	129	129

Notes: Table shows results of OLS regression on predicted treatment unit size, controlling for a quadratic function of administrative unit size. Treated units only. Bootstrapped standard errors in parentheses (2000 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Access to safe water

	Share household water from project source		Change in As contamination (household)		Change in As contamination (source)	
	(1)	(2)	(3)	(4)	(5)	(6)
TU size x treated		-0.036		0.026		0.126
		(0.033)		(0.075)		(0.084)
Predicted TU size x treated	-0.030 (0.028)		0.021 (0.062)		0.106 (0.069)	
Treated	0.115**	0.124**	-0.075	-0.082	-0.240**	-0.273**
	(0.050)	(0.059)	(0.113)	(0.133)	(0.113)	(0.138)
Estimation	RF	IV	RF	IV	RF	IV
Treated only	Yes	Yes	No	No	No	No
Union FE	No	No	No	No	No	No
Regression level	HH	HH	HH	HH	HH	HH
N	4886	4886	6431	6431	6439	6439

Notes: Table shows results of OLS regression on predicted treatment unit size, a treatment dummy, and their interactions, controlling for a quadratic function of administrative unit size, a treatment dummy and their interactions. Bootstrapped standard errors in parentheses (2000 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendices

The following Appendices provide supporting material for the analyses discussed in the main paper. Appendix A provides additional details about the intervention and describes supplementary analyses. Appendix B contains additional tables and figures referenced in the main text. Tables and figures discussed in Appendices Abut not mentioned in the main text, are embedded in the text where they are discussed.

A Supplementary description and analyses

A1 Recruitment

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

For these candidate administrative units, we provisionally defined treatment units of between between 50 and 250 households, as described in Section 3. We identified a total of 192 candidate treatment units in 103 administrative units, 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. We continued to recruit new unions and communities to the study and implement the same recruitment policy until we achieved our target recruitment levels.

A2 Implementation of assignment rule

In this section, we provide an account of the practicalities of implementing our treatment unit definition rule. We also provide a comprehensive account of the ways in which we had to deviate from our treatment unit definition rule. Note that since we use the predicted treatment unit size in our instrument construction, rather than the implemented treatment unit size, these deviations do not affect the exclusion restriction. See Section 5.3 for more information.

Administrative units vs geographical units Administrative units are defined rather arbitrarily in rural Bangladesh. In some cases, this results in an awkward local geography, meaning that clusters cannot be practically included in the same treatment unit.

In 12 cases, parts of the same administrative unit are geographically very distant from one another, separated by a major road, or separated by large areas with no arsenic contamination. In these cases, we split the administrative unit into its geographically contiguous components, and treated each of these "geographic units" as a separate administrative unit. Note that consistent with the hypothesis that administrative unit size is arbitrarily determined, we rarely find that the controls for administrative unit size affect any of the outcome variables.

In 4 cases, clusters listed as being part of an administrative unit were actually geographically part of another administrative unit. We included those clusters as part of the geographically-contiguous unit and updated both administrative unit sizes accordingly. In 5 cases, par

Additionally, in two cases, we received additional administrative lists for other parts of an administrative unit after having begun implementation, meaning that treatment units were already defined for the other parts of the administrative unit included in the study. We therefore treat the two parts of the administrative unit as two separate administrative units.

Eligibility at the cluster level We used information about arsenic contamination at the sub-administrative unit level to define eligibility. In 6 cases, we excluded parts of an administrative unit, either because some clusters had no arsenic contamination, or some clusters could not be included in the project because they were too large or small according to project rules (less than 50 or more than 250 households) and could not be combined with any other cluster. In these cases, we used the total population of the clusters eligible for the study to define administrative unit size.

Departures from treatment assignment rule We had to depart from our treatment unit size definition rule in a number of cases. In some cases, the geography of the community meant that it was simply impossible to respect natural boundaries and divide the community into the target number of treatment units without including clusters of fewer than 50 or more than 250 households.

In 1 case, we created 3 treatment units when the rule required 2; in 3 cases we created 4 treatment units when the rule required 3; in 1 (unusual) case we formed 7 treatment units rather than 4. In one case, an administrative unit with 251 households could not be subdivided, but the NGO partner was keen not to exclude the administrative unit because of high contamination. We also included two treatment units of slightly above the design threshold of 250 households, because the geography of the community made it impossible to reduce this further. These cases weaken the first stage, but these changes do not affect the exclusion restriction, because only the implemented treatment unit size was changed, not the predicted treatment unit size.

Transcription errors In three administrative units, transcription errors meant that we accidentally used the wrong value for administrative unit size in the definition of the treatment units. The errors resulted in two cases from errors in the administrative lists, and in the second case because we determined that one cluster was ineligible after collecting baseline data, and we did not update our list of administrative unit size to reflect this change. In these cases, we use the administrative unit size that we used in determining treatment unit size in the construction of the instrument.

A3 Assignment of tubewells to treatment units

We implemented two rules for assigning numbers of tubewells to treatment units. The goal of both these rules was to keep the ratio of households to tubewells constant around the thresholds.

The first rule we implemented was to assign one tubewell to administrative units with less than 150 households and two tubewell to treatment units with more than 300 households, and so on. The goal was that the change in resource intensity should not coincide with the discontinuities in treatment unit size. However, in cases where there was a lot of variation in treatment unit size, this rule generated large inequalities in the ratio of households to tubewells. For example, if we divided an administrative unit with 260 households into one cluster of 210 and one cluster of 50 households, each would receive one tubewell. This policy was unpopular with our field staff, and also threatened to inadvertently create discontinuities in the average ratio between tubewells and households at higher thresholds.

The second rule we implemented was simpler and more popular with our field staff, and was simply designed to keep the ratio of households to tubewells as close to 125 as possible. In practice, this implied assigning one tubewell in treatment units with up to 175 households, and two tubewells for larger treatment units. We verified ex ante that this assignment rule yielded no discontinuity in the ratio of tubewells to households at the discontinuities in the treatment unit assignment rules.

B Additional Figures and Tables

Figure C1a: McCrary Density Test: Threshold value 250 households

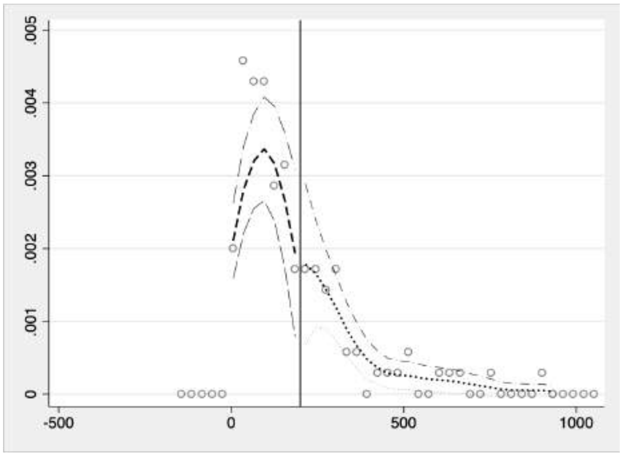


Figure C1b: McCrary Density Test: Threshold value 500 households

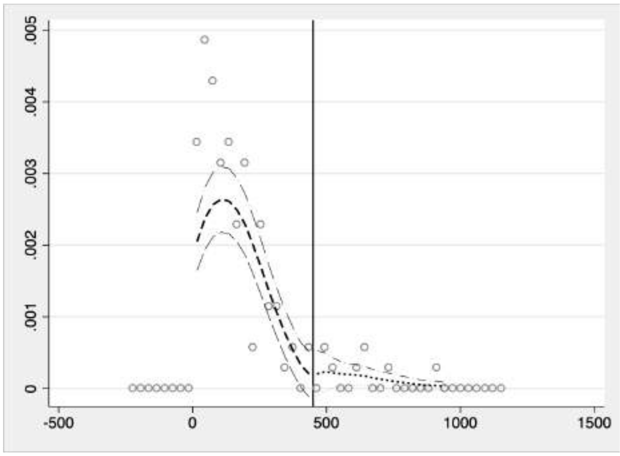
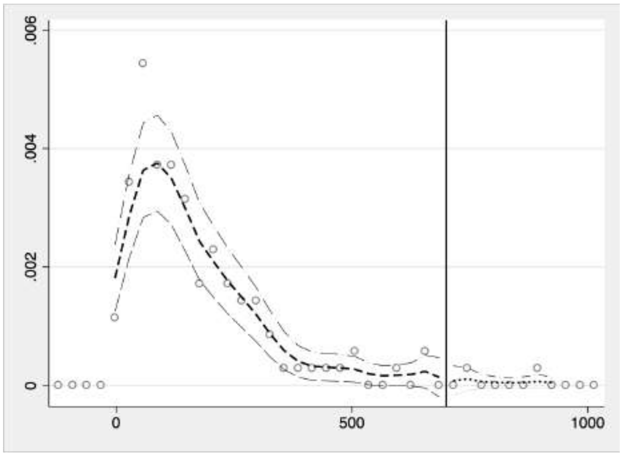


Figure C1c: McCrary Density Test: Threshold value 750 households



Notes Figures display McCrary density tests for bunching at threshold values. Note the x axes are displaced by 50 households, so that the fitted density function corresponds to the support of the data.

Table C1: Baseline balance check: with lottery fixed effects

	Ratio 100 hhs: TWs	Clusters per 100 hhs	As HH (WHO)	As WS (WHO)	Network size	High trust	Knows association
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted TU size x treated	-0.020	-0.232	-0.173*	-0.225**	-0.253	-0.051	-0.123*
	(0.083)	(0.312)	(0.101)	(0.108)	(0.221)	(0.064)	(0.065)
Treated	1.191***	0.490	0.373**	0.444**	0.490	0.079	0.197*
	(0.145)	(0.547)	(0.177)	(0.190)	(0.411)	(0.114)	(0.105)
Union FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regression level	TU	TU	HH	HH	HH	HH	HH
N	171	171	6526	6513	6529	6523	6452

Notes: Table shows results of OLS regression on predicted treatment unit size, a treatment dummy, and their interactions, controlling for a quadratic function of administrative unit size, a treatment dummy and their interactions, and lottery fixed effects. Bootstrapped standard errors are in parentheses (2000 replications). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table C2: Participation in community meeting: with lottery fixed effects

	HH attended				HH female attended			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TU size			-0.069	-0.010			-0.103**	-0.086*
			(0.045)	(0.053)			(0.046)	(0.049)
Predicted TU size	-0.058	-0.008			-0.087**	-0.069*		
	(0.038)	(0.042)			(0.037)	(0.038)		
Treated								
Estimation	RF	RF	IV	IV	RF	RF	IV	IV
Treated only	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Union FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conditional on distance	No	Yes	No	Yes	No	Yes	No	Yes
Regression level	HH	HH	HH	HH	HH	HH	HH	HH
N	4866	4428	4866	4428	4866	4428	4866	4428

Notes: Table shows results of OLS regression on predicted treatment unit size, (columns 1, 2, 5, and 6) or 2SLS regressions using predicted treatment unit size as an instrument for treatment unit size (columns 3, 4, 7 and 8). All regressions control for a quadratic function of administrative unit size and lottery fixed effects. Columns 2, 4, 6, and 8 additionally control for distance between the household and the meeting site. Data from project records. Treated units only. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

Table C3: Deliberation: with lottery fixed effects

	Meeting duration		Discussed sites	
	(1)	(2)	(3)	(4)
TU size		-40.531***		-0.482**
		(7.083)		(0.209)
Predicted TU size	-35.965***		-0.430**	
	(6.213)		(0.181)	
Estimation	RF	IV	RF	IV
Union FE	Yes	Yes	Yes	Yes
Regression level	TU	TU	TU	TU
N	127	127	128	128

Notes: Table shows results of OLS regression on predicted treatment unit size, (columns 1 and 3) or 2SLS regressions using predicted treatment unit size as an instrument for treatment unit size (columns 2 and 4). All regressions control for a quadratic function of administrative unit size and lottery fixed effects. All outcome variables normalized by number of offered sites. Treated units only. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

Table C4: Safe water sources installed: with lottery fixed effects

	Attempted installations		Successful installations	
	(1)	(2)	(3)	(4)
TU size		0.056		0.046
		(0.158)		(0.165)
Predicted TU size	0.050		0.041	
	(0.138)		(0.144)	
Estimation	RF	IV	RF	IV
Union FE	Yes	Yes	Yes	Yes
Regression level	TU	TU	TU	TU
N	129	129	129	129

Notes: Table shows results of OLS regression on predicted treatment unit size, controlling for a quadratic function of administrative unit size. Treated units only. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

Table C5: Access to safe water : with lottery fixed effects

	Share household water from project source		Change in As HH (WHO)		Change in As WS (WHO)	
	(1)	(2)	(3)	(4)	(5)	(6)
TU size x treated		-0.024		0.036		0.157*
		(0.033)		(0.083)		(0.081)
Predicted TU size x treated	-0.020 (0.027)		0.030 (0.066)		0.129** (0.062)	
Treated			-0.091	-0.101	-0.284***	-0.325**
			(0.117)	(0.143)	(0.105)	(0.135)
Estimation	RF	IV	RF	IV	RF	IV
Treated only	Yes	No	No	No	No	No
Union FE	Yes	Yes	Yes	Yes	Yes	Yes
Regression level	HH	HH	HH	H	HH	H
N	4886	4886	6431	6431	6439	6439

Notes: Table shows results of OLS regression on predicted treatment unit size, a treatment dummy, and their interactions, controlling for a quadratic function of administrative unit size, a treatment dummy and their interactions. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

Table C6: Access to safe water : with controls for tubewells per household

	Share household water from project source		Change in As contamination (household)		Change in As contamination (source)	
	(1)	(2)	(3)	(4)	(5)	(6)
TU size x treated		-0.022		0.007		0.166*
		(0.030)		(0.089)		(0.095)
Predicted TU size x treated	-0.017 (0.022)		0.005 (0.067)		0.127* (0.070)	
Treated	0.091**	0.099**	-0.049	-0.052	-0.285**	-0.347**
	(0.039)	(0.050)	(0.128)	(0.161)	(0.123)	(0.163)
Estimation	RF	IV	RF	IV	RF	IV
Treated only	Yes	Yes	No	No	No	No
Union FE	No	No	No	No	No	No
Regression level	HH	HH	HH	HH	HH	HH
Resource controls	Yes	Yes	Yes	Yes	Yes	Yes
N	4886	4886	6431	6431	6439	6439

Notes: Table shows results of OLS regression on predicted treatment unit size, a treatment dummy, and their interactions, controlling for a quadratic function of administrative unit size, a treatment dummy and their interactions. Bootstrapped standard errors in parentheses (2000 replications). *** p <0.01, ** p <0.05, * p <0.1.

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC
International
Growth Centre

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid
from the British people

Designed by soapbox.co.uk